Spatial and Spatio-temporal Epidemiology 21 (2017) 67–75

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Original Research

Hot spots, cluster detection and spatial outlier analysis of teen birth rates in the U.S., 2003–2012

Diba Khan[a,*], Lauren M. Rossen[b], Brady E. Hamilton[b], Yulei He[a], Rong Wei[a], Erin Dienes[c]

[a] Division of Research and Methodology/CDC/NCHS, Hyattsville, MD 20782, USA
[b] Division of Vital Statistics/CDC/NCHS, Hyattsville, MD 20782, USA
[c] RADARS SYSTEM/Rocky Mountain Poison & Drug Center (RMPDC)/A Division of Denver Health, 990 Bannock Street M/C 0180 | Denver, CO 80204, USA

ARTICLE INFO

Article history:
Received 7 July 2016
Revised 5 January 2017
Accepted 16 March 2017
Available online 27 March 2017

Keywords:
Bayesian
Hotspots
Spatial outliers

ABSTRACT

Teen birth rates have evidenced a significant decline in the United States over the past few decades. Most of the states in the US have mirrored this national decline, though some reports have illustrated substantial variation in the magnitude of these decreases across the U.S. Importantly, geographic variation at the county level has largely not been explored. We used National Vital Statistics Births data and Hierarchical Bayesian space-time interaction models to produce smoothed estimates of teen birth rates at the county level from 2003–2012. Results indicate that teen birth rates show evidence of clustering, where hot and cold spots occur, and identify spatial outliers. Findings from this analysis may help inform efforts targeting the prevention efforts by illustrating how geographic patterns of teen birth rates have changed over the past decade and where clusters of high or low teen birth rates are evident.

1. Introduction

Teen childbearing is associated with negative health, social, and economic outcomes for both the mothers and infants (Ventura et al., 2014). Infants born to teen mothers are more likely to be born preterm, and to die within the first year of life compared to infants born to older mothers (Ventura et al., 2014). Moreover, the economic burden of teen childbearing to the public has been estimated to total 9.4 billion dollars annually in the U.S. (The National Campaign to Prevent Teen and Unplanned Pregnancy, 2013). At a national level, teen birth rates have evidenced a significant decline over the past few decades. In 2014, there were 24.2 births for every 1,000 adolescent females (15–19 years of age), a decline of nearly 60 percent over the past 25 years (Hamilton et al., 2014; Martin et al., 2013). Most of the states in the U.S. have mirrored this national decline, though some reports have illustrated substantial variation in the magnitude of these decreases across the U.S. with declines ranging from 52% to 71% (Ventura et al., 2014). In addition, states in New England have historically had the lowest teen birth rates in the U.S., while several southern states (e.g., Arkansas, Mississippi, Oklahoma) have teen birth rates higher than the national average (The National Campaign to Prevent Teen and Unplanned Pregnancy, 2013; National Center for Health Statistics, 2014; Ventura et al., 2014).

Although many studies and reports have reported large-scale geographic patterns, spatial patterns at the county-level have not been much explored. Describing how counties with high or low teen birth rates cluster geographically may aid efforts to further reduce teen birth rates in specific

* Corresponding author at: Diba Khan, National Center for Health Statistics, Hyattsville, MD-20782.
E-mail address: ild1@cdc.gov (D. Khan).

http://dx.doi.org/10.1016/j.sste.2017.03.002
1877-5845/Published by Elsevier Ltd.

areas of the U.S. Moreover, identifying counties that represent spatial outliers can inform future research seeking to better understand what factors might be driving greater success in reducing teen birth rates in some areas as compared to others. Many factors have been implicated in explaining the declines in teen birth rates such as increased contraceptive use and shifts toward more effective methods of contraception, delay in sexual initiation, the effects of the economic downturn, and the impact of various federal, state, and local public health programs and interventions to prevent teen pregnancy (Romero et al., 2015; Ventura et al., 2012). The objectives of this analysis were: to examine whether teen birth rates show evidence of clustering, to identify where hot and cold spots occur (groups of counties with extremely high or low teen birth rates), and to identify spatial outliers (counties with high or low teen birth rates surrounded by counties with dissimilar values), over the study period.

2. Methods

2.1. Data and estimates

Data on the number of live births for women aged 15–19 were extracted from the National Vital Statistics Birth Data Files for the years 2003–2012, (National Center for Health Statistics, 2003–2012). These data were then aggregated to the county level to provide a count of births to women 15–19 years of age for each county and year. The denominators to calculate teen birth rates were obtained from intercensal and postcensal population estimates of the number of females aged 15–19 years residing within each county over the same time period. These population denominators were extracted from the files containing intercensal and postcensal bridged-race population estimates provided by the National Vital Statistics System (National Vital Statistics System, 2009). There were 929 counties where there were fewer than 20 births to women 15–19 years of age in 2003 and 1156 counties in 2012. Direct estimates are typically suppressed in counties with fewer than 20 cases in the numerator due to concerns about the stability and reliability of the estimates. To address these problems, we employed small area estimation methods to produce stable county-level estimates of teen birth rates from 2003–2012. These methods are described elsewhere (Khan et al., 2018) and are reviewed briefly here. More detail can be found in the supplemental appendix.

Hierarchical Bayesian space-time interaction models were employed to produce county-level estimates of the teen birth rates for each year (Lawson, 2013). These models accounted for spatial and temporal dependence along with space time interaction terms to generate county-level estimates of teen birth rates for each year, 2003–2012. These models included covariates related to county-level income and poverty, such as: per capita income, percent of county in poverty, unemployment rate, education, and racial/ethnic composition, proportion of foreign born residents, education level, measured at different time points. In addition to these covariates- which were significantly associated with teen birth rates at the county level, the number of family planning and Title X clinics by county, based on data provided by the Guttmacher Institute (Kost, 2014), were initially included in the models but subsequently removed due to the negligible association with teen birth rates at the county level. These models borrow strength across counties and time to produce stable estimates of teen birth rates at the county level, addressing problems due to data sparsity and allowing for further examination of geographic and temporal patterns.

2.2. Model

Let y_{it} = counts of teen births in county i and year t, and n_{it} = counts of teen population in county i and year t. Then,

$y_{it} \sim Binomial(n_{it},\ p_{it})$; $i = 1, \ldots, m$ counties and $t = 1, \ldots, T$ years,where p_{it} = probability of teen births in county i at time t (Lawson, 2013; Khan et al., 2018).

The convolution model is:

$logit(p_{it}) = \alpha_0 + a_{1i} * year_t + X_i{}'\boldsymbol{\gamma} + u_i + v_i + \psi_{it}$. The components in the convolution model correspond to:

1. logit link function $\log(p_{it}/(1 - p_{it}))$.
2. α_0, an intercept.
3. time trend term $a_{1i}{}^*\ year_t$.
4. $X_i{}'\boldsymbol{\gamma}$, where X_i : is the ith row of the covariates matrix and $\boldsymbol{\gamma}$ is a vector of regression parameters.
5. spatial random effects u_i by county to model strong spatial autocorrelation $i = 1, \ldots, m$ counties.
6. non-spatial random effects v_i by county to model residual spatial autocorrelation that were not dealt with by our spatial random effects, u_i, $i = 1, \ldots, m$ counties.
7. space-time interaction term ψ_{it}, a random effect where ψ_{it} is a function of its past values, $\psi_{i,t-1}$, plus an error term.

Parameters under (5) are modeled via normal conditional autoregressive priors (CAR) (Besag et.al, 1991). Parameters under (6) are modeled via normal conditional priors. Parameters under (7) are modeled via Type II random walk interaction (Knorr-Held and Rasser, 2000), which is included to account for any residual spatiotemporal dependence or variation that is not captured by the spatial or temporal main effects. The values for a given county in a given year depend upon the values observed for that county in the prior year plus a residual (Knorr-Held and Rasser, 2000; Lawson, 2013).

The predicted county-level teen birth rates obtained from the best model were merged with US Census Tiger/Line files and mapped using ArcGIS 10.1 (E.S.R. Institute, 2011).

3. Spatial statistical tools

Having obtained estimates from the models described above, several spatial statistical tools were implemented to examine spatial clustering and outliers.

3.1. Global index of spatial autocorrelation – Moran's I

Global indexes of spatial autocorrelation were used to assess the similarity, or spatial dependence, across counties