



Original Research

Comparing multilevel and multiscale convolution models for small area aggregated health data



Mehreteab Aregay^{a,*}, Andrew B Lawson^a, Christel Faes^b, Russell S Kirby^c, Rachel Carroll^a, Kevin Watjou^b

^a Department of Public Health, Medical University of South Carolina, Charleston, SC, USA

^b Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium

^c Department of Community and Family Health, University of South Florida, Tampa, FL, USA

ARTICLE INFO

Article history:

Received 15 November 2016

Revised 6 June 2017

Accepted 6 June 2017

Available online 5 July 2017

Keywords:

Multilevel model

Multiscale model

Convolution model

Shared random effects

Scaling effects

Contextual effects

ABSTRACT

In spatial epidemiology, data are often arrayed hierarchically. The classification of individuals into smaller units, which in turn are grouped into larger units, can induce contextual effects. On the other hand, a scaling effect can occur due to the aggregation of data from smaller units into larger units. In this paper, we propose a shared multilevel model to address the contextual effects. In addition, we consider a shared multiscale model to adjust for both scale and contextual effects simultaneously. We also study convolution and independent multiscale models, which are special cases of shared multilevel and shared multiscale models, respectively. We compare the performance of the models by applying them to real and simulated data sets. We found that the shared multiscale model was the best model across a range of simulated and real scenarios as measured by the deviance information criterion (DIC) and the Watanabe Akaike information criterion (WAIC).

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

In spatial epidemiology, data are often arrayed hierarchically, i.e., individual level data are aggregated into areal units (e.g. counties) that are clustered to form larger areal units (e.g. states). The clustering of individuals into areal units, which in turn are grouped into larger areal units, can induce contextual effects (Lawson, 2013, 2016); meaning that individuals within areal units have similar characteristics. In general, contextual effects arise from the underlying spatial distribution of the individual level outcomes. Appropriate model parameters should be used to adjust for the contextual effects and when such parameters are not included then bias would be induced in the estimated relative risk, which is the parameter of interest in the disease

mapping. Hence, researchers considered multilevel modeling of hierarchically available individual level data to encompass contextual effects (Bobashev and Anthony, 1998; Goldstein et al., 2002; Leyland and Goldstein, 2001; Merlo et al., 2004; Preisser et al., 2003). However, this approach only handles the correlation between the outcomes within a single areal unit; it ignores spatial correlation among neighboring areal units.

Multilevel models often assume that all spatial correlation can be reduced to within area correlation (Chaix et al., 2005); thus, there is no spatial random effect component that handles the correlation between the neighboring areas. The random effects in multilevel models only account for the correlation between the individual level outcomes within a given spatial unit. Therefore, it provides partial information on the geographical variation of health outcomes in measuring the correlation within a spatial unit but not the correlation between neighboring regions. Researchers extend multilevel models to incorporate

* Corresponding author.

E-mail address: mihrish00m@gmail.com (M. Aregay).

spatial interaction effects in different fields such as geography and spatial econometrics (Browne et al., 2001; Chaix et al., 2005; Dong et al., 2015; Langford et al., 1999; Tranmer et al., 2014). In this paper, we (i) extend the multilevel models to account for spatial correlations between adjacent areas using a convolution model developed by Besag et al., (1991); and (ii) develop models for aggregated unit level data by adjusting for contextual effects. The proposed multilevel model focuses on the risk variation at the fine level areal units by incorporating the contextual effects in the model. In addition, the estimation of the risk variation at larger areal units (coarse level) is possible by aggregating (e.g. averaging) the fine level estimates within the coarse level, while using only the data at the fine level.

In practice, data could be available at different geographically aligned levels. For example, the outcome of interest could be available in the form of aggregation at the census block, block group, and census tract: the responses at the census block could be summed up to obtain the responses at the block group, which in turn could be summed up to obtain the responses at the census tract. This kind of data aggregation results in losing information at the coarse level (e.g. block group and census tract). This is known as a scaling effect in geography (Wong, 2009). Scaling effects arise when data are aggregated from a lower (e.g. census tract) into a higher geographical level (e.g. county). In the literature, multiscale models have been used in different fields to solve scaling problems at multiple scale levels (Basseville et al., 1992; Berliner et al., 1999; Calder et al., 2009; Chou et al., 1994; Craigmile and Guttorp, 2011; Delouille et al., 2006; Huang and Cressie, 2000; Huang et al., 2002; Johannesson et al., 2007; Kolaczyk and Huang, 2001; Nychka et al., 2002; Vidakovic, 1999; Wikle et al., 2001; Zhu and Yue, 2004).

In spatial epidemiology, researchers have implemented multiscale models to account for scaling effect due to the aggregation of data (Banerjee et al., 2004; Cressie, 1996; Wong, 2009) by factorizing the likelihood at the coarse (high) level into the fine (low) level (Louie and Kolaczyk, 2004, 2006a, 2006b). Alternatively, we (Aregay et al., 2015a, 2015b, 2016a, 2016b) developed a shared random effect multiscale model that accommodates the aggregation (scale) effect by inheriting the coarse level effect into the fine level. However, it could be argued that the latter approach uses the data twice as the data at the coarse level are an aggregation of the data at the fine level. The objective of this paper is to describe risk variations at fine and coarse levels simultaneously by accommodating scaling and contextual effects. To achieve this goal, we applied and compared different models. First, we compare the shared multiscale model with the shared multilevel model in real and simulated data sets. Second, we study the impact of ignoring the contextual effects on the estimation of the risk variations at both the fine and coarse levels by simulating data with strong contextual effects. Note that the focus of this paper is on studying contextual effects although we touch on scaling effects as well.

The structure of the paper is as follows. In Section 2, we present the data that motivated us to conduct this research. Section 3 describes the statistical methods as well as the design of the simulation study, while Section 4 ded-

icates to the results obtained from fitting the models to the real and simulated data sets. Finally, in Section 5, we present the discussion and concluding remarks.

2. Georgia oral cancer data

We are motivated by the county level data available in the state of Georgia via OASIS system (<http://oasis.state.ga.us>). We consider the number of persons discharged from non-federal acute-care inpatient facilities for oral cancer in 2008. The observed outcomes of the counties are aggregated (summed up) to the public health (PH) districts. These aggregations of data can induce a scaling effect. The state of Georgia consists of 159 counties (see left panel in Fig. 1) that are classified into 18 PH districts (see the right panel of Fig. 1). The grouping of the counties into PH districts can induce a contextual effect. Each PH district consists of one or more counties. The PH districts are used for administration of public health resources. The Georgia Department of Public Health (DPH) funds and collaborates with the 18 PH districts. The goal of modeling the risk variation at both the county and PH district levels is that it can be used for allocating of health resources at both levels in a cost-effective manner. Hence, the DPH can use the risk mapping results to legislate regulations to protect the public health in each county as well as in each PH district.

The observed standardized morbidity ratio (SMR), which is the ratio of the outcome to the expected number of cases, at both the county and PH levels are displayed in Fig. 2. We can see that the scaling effect smooths out the county level risk variation when the data are aggregated into the PH districts. To address both the contextual and scale effects, we propose different models described in Section 3. It is worth mentioning contextual and scale effects have an inverse relationship. When we have strong presence of contextual effects, we will have weak scale effects because the risk variations will be similar at the fine (e.g. county) and coarse (e.g. PH districts) levels (see Fig. 3). The scales of the relative risk (RR) ranges from 0.61 to 1.68 on both scale levels. On the other hand, when there are weak contextual effects, the scale effects will be strong (see Fig. 4). The scale of the RR in the left panel is between 0.45 and 2.23, whereas in the right panel it is between 0.97 and 1.49 indicating that the presence of a relatively strong scale effect smoothed out the risk variation at the county level when it is aggregated into the PH level. We can also see this kind of behavior in Fig. 2. The application of the models to the Georgia oral cancer data is deferred to Section 4.2.

3. Models for aggregated small area data

In the next section, we present the models most relevant to small area aggregated data. To make it clear, we abbreviated the four models considered below as M1, M2, M3, and M4 and they represent Model 1, Model 2, Model 3, and Model 4, respectively. We define the models using the two scale levels Georgia oral cancer study. Assume that y_{ij} is the outcome of interest for the j^{th} county (fine scale) at the i^{th} public health district (coarse scale)

Download English Version:

<https://daneshyari.com/en/article/5118929>

Download Persian Version:

<https://daneshyari.com/article/5118929>

[Daneshyari.com](https://daneshyari.com)