Spatial and Spatio-temporal Epidemiology 12 (2015) 39–52

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

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

Original Research

Spatial variation in attributable risks

Peter Congdon

School of Geography, QMUL, United Kingdom

ARTICLE INFO

Article history:
Received 27 June 2014
Revised 21 October 2014
Accepted 6 February 2015
Available online 24 February 2015

Keywords:
Attributable risk
Spatial
Diabetes
Obesity
Deprivation
Bayesian

ABSTRACT

The attributable risk (AR) measures the contribution of a particular risk factor to a disease, and allows estimation of disease rates specific to that risk. While previous studies consider variability in ARs over demographic categories, this paper considers the extent of spatial variability in ARs estimated from multilevel data with confounders both at individual and geographic levels. A case study considers the AR for diabetes in relation to elevated BMI, and area rates for diabetes attributable to excess weight. Contextual adjustment includes known area variables, and unobserved spatially clustered influences, while spatial heterogeneity (effect modification) is considered in terms of varying effects of elevated BMI by neighbourhood deprivation category. The application is to patient register data in London, with clear evidence of spatial variation in ARs, and in small area diabetes rates attributable to excess weight.

1. Introduction

The attributable risk (AR) seeks to quantify the proportion of disease due to a particular risk factor, which may be termed the focus risk factor (Uter and Pfahlberg, 2001; Benichou, 2001). Other terms include the attributable fraction, population attributable risk, and population attributable fraction. The AR measures impacts of risk factors on disease levels, taking into account both associations (i.e. relative risk) between disease and exposure, and the proportion of subjects exposed. Using attributable risks one may ascertain disease rates and burdens specific to a particular risk factor (Steenland and Armstrong, 2006; Ezzati et al., 2006; Gefeller, 1995). With a risk factor expressed in binary form, and P_E as the proportion of subjects exposed, a point estimator of the attributable risk is

$$[P_E(RR-1)]/[P_E(RR-1)+1], \tag{1a}$$

where RR is the relative risk for those exposed as compared to those unexposed. The latter should be adjusted for confounders (Darrow and Steenland, 2011; Benichou, 2001; Steenland and Armstrong, 2006). Another estimator is

$$P_{E|D}(RR-1)/RR = P_{E|D}(1-1/RR), \tag{1b}$$

where $P_{E|D}$ is the proportion of diseased subjects exposed.

Variations in attributable risks over demographic categories (e.g. age categories, ethnic groups) have been considered in some studies (e.g. Okosun and Boltri, 2006; Oteng-Ntim et al., 2013), and contributions to the Global Burden of Disease study such as Ferrari et al., 2014) use the estimator (1a) to derive attributable risks varying both by demographic group and over nations. Variations in attributable risks at subnational geographic scales, down to relatively small area scale, and the underlying methodological issues, have, however, been little explored. One approach (Tanuseputro et al., 2005) assumes confounder adjusted relative risks based on national epidemiological surveys to be transferable across lower scale geographic settings. However, relative risks may vary across such geographic settings. Estimation of ARs from multilevel data, after adjustment for contextual risks, either measured neighbourhood confounders (e.g. area deprivation), or unmeasured spatially structured risk factors, has not been

E-mail address: p.congdon@qmul.ac.uk

http://dx.doi.org/10.1016/j.sste.2015.02.002
1877-5845/

considered in previous studies. Allowance for spatial heterogeneity (e.g. effects of the focus risk varying by area type) is also not considered in existing studies.

The present paper is particularly concerned with the AR for diabetes in relation to excess weight. The association between elevated bodyweight and diabetes risk has been explored in many studies based on patient level data, either using categorical forms of the bodyweight predictor (Ganz et al., 2014; Field et al., 2001), or with linear regression of diabetes risk on BMI (Wong et al., 2014). However, a few multilevel studies have also considered spatial aspects of rising obesity prevalence or diabetes (e.g. Krokstad et al., 2013; Liu and Núñez, 2014), and a multilevel perspective on obesity is advocated by Huang et al. (2009). Taking account of geographic context is important as an increasing number of studies link obesity (and hence diabetes) to environmental influences (Hill and Peters, 1998).

The present paper seeks to assess the potential importance of spatial effects (spatial heterogeneity, spatial clustering) in the estimation of context sensitive attributable risks, and their relevance in estimating area disease rates specific to particular risk factors, specifically small area diabetes rates attributable to excess weight. Spatial variation in the latter is particularly relevant for policy purposes. A subsidiary aim is to demonstrate the utility of a Bayesian approach to estimation using a logistic regression method in which ARs are based on a ratio estimator, while also selecting out significant influences on the disease outcome using Bayesian variable selection.

1.1. Attributable risks in a multilevel setting

The application in this paper considers estimation of ARs for diabetes prevalence in relation to excess weight, using multilevel data from health registers or health surveys. Oriented to such data, the paper considers adjustment for both patient confounders (e.g. other diseases, age, ethnicity), and observed and unobserved neighbourhood (contextual) confounders.

Multilevel data presupposes subjects nested within clusters, and observations for subjects within clusters areas may be correlated (Chen and Dey, 2003). As mentioned by Diez-Roux et al. (1997) "correlation between individuals within neighborhoods ... may persist even after controlling for [observed] individual level and neighborhood level variables." Existing multilevel disease risk models generally consider spatial effects in terms of (a) effects of observed area variables, and (b) randomly varying intercepts (typically assumed iid) over areas to represent unobserved area influences. Contextual effects are then assessed in terms of the relative proportion of variation explained by areas (e.g. Pickett and Pearl, 2001; Merlo et al., 2006). Some analyses go beyond this to allow for spatial clustering in unmeasured neighbourhood influences on disease levels (Dasgupta et al., 2014; Xu, 2014; Chaix et al., 2005).

However, as discussed in Goodchild (2011), spatial effects encompass spatial heterogeneity as well as spatial clustering. There is an extensive literature on spatially varying regression relationships with both Bayesian approaches (Assunçao, 2003), and classical approaches often based on generalized weighted regression (Fotheringham et al., 2003). This paper considers a relatively simple form of heterogeneity in regression effects, namely varying impacts of individual risk factors according to area type. In terms of the framework provided by Anselin (2010, p. 6) the form of heterogeneity considered here involves discrete heterogeneity, or spatial regimes.

Such heterogeneity can also be seen as a spatially defined form of effect modification or "hazard heterogeneity" (Ezzati et al., 2006, p. 245), applicable "when the assumption of constant relative risk [is] not appropriate". The potential importance of effect modification in estimating ARs is considered by Flegal et al. (2004).

Specifically the analysis below accordingly considers estimation of ARs via multilevel models that admit the potential for (a) spatially correlated but unmeasured risk factors, and (b) neighbourhood group heterogeneity in impacts of bodyweight on diabetes. Regarding the first feature, and as discussed above, multilevel data presupposes subjects nested within clusters, and observations or residuals for subjects within clusters areas may be correlated (Chen and Dey, 2003). When areas constitute the clusters, residuals may show spatial correlation.

There is an extensive literature on modeling spatially correlated residual effects on health outcomes. Such spatial effects often proxy unobserved risk factors (e.g. environmental or cultural), which vary smoothly over space (Best, 1999). As mentioned by Wakefield et al. (2000), modeling of spatially correlated errors, denoted v_j ($j = 1,\ldots J$) for J areas, may proceed by initially specifying either the joint multivariate distribution of the vector $v = (v_1,\ldots,v_J)$, or the univariate density of each areas error, v_j, conditional on errors in other areas. A widely adopted scheme known as the convolution prior, but with potential identification issues, involves an intrinsic autoregressive effect (Besag et al., 1991) combined with an iid (non-spatial) effect. Lee (2011) compares the properties of alternative conditional priors for spatial errors, and recommends instead the method of Leroux et al. (1999), on the grounds of including a measure of spatial dependence, and in providing a rational form of conditional variance.

Regarding spatial heterogeneity, a focus here is on the potential interaction between area deprivation category and the effects of overweight, a cross-level interaction in the terminology of multilevel analysis. Possible mechanisms for such interaction are suggested by the large number of studies linking obesity (and diabetes itself) to environmental influences, such as access to healthy food and exercise opportunities (Hill and Peters, 1998; Feng et al., 2012; Salois, 2012). For example, obesity may be related to aspects of food environment (e.g. density of facility types, such as fast food outlets) which adversely influence diet, with less healthy food environments characterized by high consumption of processed food, high in fat and sugar (Lake and Townshend, 2006). Less healthy food environments tend to be in less affluent areas, that is areas with high deprivation (Morland et al., 2002). Exercise has independent effects on diabetes as well as through its effect on obesity (Kriska et al., 2003; De Feo et al., 2006), and exercise access is typically lower in