

Article

A Multilevel Model for Comorbid Outcomes: Obesity and Diabetes in the US

Peter Congdon

Department of Geography and Centre for Statistics, Queen Mary University of London, Mile End Rd, London E1 4NS, UK; E-Mail: p.congdon@qmul.ac.uk

Received: 16 November 2009 / Accepted: 21 January 2010 / Published: 27 January 2010

Abstract: Multilevel models are overwhelmingly applied to single health outcomes, but when two or more health conditions are closely related, it is important that contextual variation in their joint prevalence (e.g., variations over different geographic settings) is considered. A multinomial multilevel logit regression approach for analysing joint prevalence is proposed here that includes subject level risk factors (e.g., age, race, education) while also taking account of geographic context. Data from a US population health survey (the 2007 Behavioral Risk Factor Surveillance System or BRFSS) are used to illustrate the method, with a six category multinomial outcome defined by diabetic status and weight category (obese, overweight, normal). The influence of geographic context is partly represented by known geographic variables (e.g., county poverty), and partly by a model for latent area influences. In particular, a shared latent variable (common factor) approach is proposed to measure the impact of unobserved area influences on joint weight and diabetes status, with the latent variable being spatially structured to reflect geographic clustering in risk.

Keywords: diabetes; obesity; multilevel; multinomial; latent variable; spatial; poverty

1. Introduction

Two of the major risk factors for cardiovascular disease are obesity and diabetes, and analysis of geographic patterning in the variation and interrelation of these two major conditions is important for ensuring that resources for prevention and care match need and are effectively targeted. The close link between obesity and diabetes is well established [1,2], and increases in the prevalence of obesity and

overweight are a major factor in the growth of diabetes [3,4]. In the US there is evidence of wide geographic contrasts in the prevalence of both obesity and diabetes, and of clear differences in relative risk between age and ethnic groups, and between socioeconomic groups [5,6].

It is important to establish whether geographic variations are simply the result of differences between populations in their age, social and ethnic composition (compositional effects), or whether there are distinct geographic effects that account for part of the variation. The distinct impacts of area on health, and also those of interactions between area variables and individual level risk factors, are often denoted as "contextual variation". Thus, area-based measures of socioeconomic status may affect health outcomes even after control for individual risks [7-10], while interactions between geography and individual risk factors are exemplified by the study of Subramanian *et al.* [11], which considers "geographic variation in the individual relationship between race/ ethnicity and mortality". In the present paper, contextual variation is assessed by the significance (after allowing for major individual level risk factors) of both known area variables, and latent area effects, on chances of diabetes and/or excess weight.

This paper develops a multilevel multinomial regression model for diabetes and weight category as joint outcomes. The model framework allows for subject level risk factors, and contextual (area) effects including known area influences (e.g., poverty, race composition, and population density), and unmeasured area influences at two levels (county and state). The latter are modelled using a latent variable approach that results in a summary index of latent contextual effects shared across multinomial outcomes. Because of clustering in both diseases, the latent variables are assumed to be spatially structured [12].

A case study application is based on the 2007 Behavioral Risk Factor Surveillance System (BRFSS) survey, which is an annual random-digit-dialed telephone survey to determine the prevalence among adults (ages 18 and over) of major illnesses and health risk behaviors. The results described in this paper are based on 128,150 male respondents to the 2007 BRFSS, and living in the continental United States. The main object is to demonstrate unique aspects of the methodology such as the use of a common spatial factor with a multinomial health outcome, and within a multilevel analysis that also allows for the impact of individual risk factors. The method transfers straightforwardly to other cross-sectional settings, including (say) joint obesity-diabetes prevalence for females in 2007, and no distinct methodological elements would be involved in considering females. Therefore the analysis is confined to males. Distinct methods would certainly be involved if time were introduced as an extra feature (e.g., how has the joint obesity-diabetes multilevel relationship evolved since 2000), but this is left for another study.

Obesity is defined as a body mass index over 30, based on self-reported height and weight, with overweight defined as BMI between 25 and 29.9. To determine diabetes status, respondents were asked "Have you ever been told by a doctor that you have diabetes?", encompassing both types of diabetes. Women with gestational diabetes were excluded. The BRFSS does include other questions on diabetes such as age of onset, and whether or not taking insulin; answers to these questions have been combined in some studies [13] to informally differentiate diabetes 1 (onset before age 30 combined with current insulin use) from diabetes 2. However, for the majority (94%) of diabetic survey subjects who have onset after age 30, obesity is an established risk factor for diabetes.

2. Multinomial Regression Combining Individual and Geographic Risk Factors

A multinomial response is involved when there are three or more sub-categories of a single condition, or may be obtained by combining sub-categories over two or more conditions. In the present BRFSS case study, the response y_i ($i = 1, \dots, n$) is based on combining sub-categories of two conditions: there are $J + 1 = 6$ categories defined by diabetic status and weight status, namely diabetic and obese ($y = 1$), diabetic and overweight ($y = 2$), diabetic and normal weight ($y = 3$), non-diabetic and obese ($y = 4$), non-diabetic and overweight ($y = 5$), and non-diabetic and normal weight ($y = 6$). So categories 1 to 5 all show some form of morbidity relative to the final non-morbid category who are normal weight and not diabetic.

A multinomial regression is applied with the final (non-morbid) category as reference, so that

$$Pr(y_i = j) = \pi_{ij} = \frac{\exp(\phi_{ij})}{1 + \sum_{j=1}^J \exp(\phi_{ij})} \quad j = 1, \dots, J \quad (1.1)$$

$$Pr(y_i = J + 1) = \pi_{i,J+1} = \frac{1}{1 + \sum_{j=1}^J \exp(\phi_{ij})}. \quad (1.2)$$

where the ϕ_{ij} are J regression terms. Let $d_{ij} = 1$ if subject i is in the j^{th} category. Then for equally weighted subjects the likelihood L would take the form

$$L = \prod_{i=1}^n \prod_{j=1}^{J+1} \pi_{ij}^{d_{ij}}, \quad (2.1)$$

with log-likelihood

$$\log L = \sum_{i=1}^n \sum_{j=1}^{J+1} d_{ij} \log(\pi_{ij}). \quad (2.2)$$

However, with population survey data, such as the BRFSS [14], it is necessary also to incorporate survey weights w_i for respondents i to account for differential response between demographic groups and regions. Then a weighted likelihood L_w is obtained as

$$L_w = \prod_{i=1}^n \prod_{j=1}^{J+1} [\pi_{ij}^{d_{ij}}]^{w_i}, \quad (3.1)$$

with weighted log-likelihood

$$\log L_w = \sum_{i=1}^n \sum_{j=1}^{J+1} w_i d_{ij} \log(\pi_{ij}). \quad (3.2)$$

Three classes of predictors are used in the multinomial regression defined by (1.1)–(1.2), and with weighted likelihood as in (3.1)–(3.2). As well as subject level risk variables R , the regression model includes known geographic influences G_K , and latent geographic influences, G_L , so the regression terms have the generic form $\phi_{ij} = \phi_{ij}(R, G_K, G_L)$.

Predictor effects are modelled either as fixed or random effects, in a form of general linear mixed model, in particular one with a multinomial outcome [15]. Random effects are used to pool strength (e.g., over areas or age groups) and to incorporate anticipated correlations in the age or spatial profiles of prevalence for categories $j = 1, \dots, J$ (see Appendix 1). For example, while the levels of the different combinations of diabetic and weight status are different (e.g., obesity without diabetes is much more common than diabetes with normal weight), one would expect their age profiles to be similar (*i.e.*, correlated).

3. Subject Level Predictors

There has been extensive research on variations in obesity and diabetes over demographic and socioeconomic categories, such as age, socioeconomic status and race. A pronounced gradient in diabetes prevalence by age is reported by CDC [16], though obesity may reduce slightly among the very old. Paeratakul *et al.* [5] also report impacts of socioeconomic status (SES) on obesity and its comorbidities. For example, obese subjects with lower education reported higher rates of diabetes compared to those with higher education; these differentials were more marked than those between high and low income individuals; see Table 3 in Paeratakul *et al.* [5]. Freudenberg & Ruglis [17] argues that “although education is highly correlated with income and occupation, evidence suggests that education exerts the strongest influence on health”, and Zhang *et al.* [18] and Maty *et al.* [19] also argue the benefit of using education as a measure of SES risk. For the 2007 BRFSS data, Table 1 shows morbidity rates due to obesity/overweight and/or diabetes by education level, namely percentages of subjects at each education level located in the six diabetic-weight categories (including the reference category). Higher levels of combined morbidity, namely suffering diabetes combined with obesity or overweight, occur for the less well educated. For example, 5.6% of subjects with less than high school education (namely, 794 of 14,158 survey participants at this education level) are obese and diabetic, compared to 3.2% of college graduates (namely, 1562 of 48186 survey participants at this education level).

Table 1. Percentage prevalence of obesity-diabetes status, by education.

	Less than high school	High school graduate	Some college	College graduate	All
Obese & Diabetic	5.6	4.4	4.7	3.2	4.2
Overweight & Diabetic	3.8	3.3	2.8	2.9	3.1
Normal weight & Diabetic	2.2	1.4	1.2	1.3	1.4
Obese non-diabetic	21.3	24.9	25.1	19.0	22.4
Overweight & non-diabetic	36.3	37.3	39.4	44.0	40.2
Normal weight & non-diabetic	30.8	28.6	26.9	29.6	28.8

Ethnic group variability in levels of obesity and diabetes are well established. Paeratakul *et al.* [5] found the prevalence of overweight and obesity to be higher among black and Hispanic groups compared to whites, and the prevalence of obesity comorbidities (including diabetes) was also found to be higher in blacks than whites. For the 2007 BRFSS data, Table 2 shows morbidity rates due to obesity/overweight and/or diabetes by race, namely percentages of subjects in the six diabetic-weight categories. Compared to other races, black non-Hispanics are more likely to be located in the first two categories, and also have the highest proportion who are both obese and non-diabetic. The other race category (penultimate column in Table 2) has a relatively high proportion who are diabetic but of normal weight. As Zhang *et al.* [18] mention, racial disparities in diabetes are not entirely explained by racial/ethnic differences in the prevalence of common risk factors such as obesity: racial differences in diabetes risk remain after controlling for body mass and socioeconomic status. Hence cross tabulation such as in

Table 2 does not control for interrelations between risk factors (e.g., between race and education), and a regression is required.

Table 2. Percentage prevalence of obesity-diabetes status, by race.

	White non-Hispanic	Black non-Hispanic	Hispanic	Other	All
Obese & Diabetic	4.1	5.7	3.7	3.7	4.2
Overweight & Diabetic	3.0	4.1	2.8	2.9	3.1
Normal weight & Diabetic	1.3	1.6	1.3	2.1	1.4
Obese non-diabetic	22.1	27.1	25.2	14.7	22.4
Overweight & non-diabetic	41.0	35.3	41.5	37.1	40.2
Normal weight & non-diabetic	28.5	26.3	25.4	39.5	28.8

Subject level risks are here represented in the regression terms ϕ_{ij} by:

(a) overall intercepts (α_j),

(b) differential risks by ethnic group r , namely $r = 1$ for white non-Hispanic, $r = 2$ for black, $r = 3$ for Hispanic, and $r = 4$ for other races (mainly American Asians and native Americans); these are modelled as fixed effects within each ϕ_{ij} , with unknown parameters β_{jr} , $r = 2, 3, 4$, and with $\beta_{j1} = 0$ as reference under an identifying corner constraint;

(c) differential risks by education attainment e , namely $e = 1$ for less than high school; $e = 2$ for high school graduate; $e = 3$ for some college or technical school; and $e = 4$ for college graduate; these are also modelled as fixed effects, with unknowns γ_{je} , $e = 2, \dots, 4$, and $\gamma_{j1} = 0$ as reference;

(d) differential risks by age group $a = 1, \dots, A$ (with $A = 12$ for ages 18–24, 25–29, 30–34, ..., 70–74, and 75+), represented by unknowns η_{ja} . These are modelled using a random effects approach that allows correlation in the age profiles over the first J prevalence categories (see Appendix 1); an identifying constraint is applied that ensures these effects sum to zero within outcomes, so that $\sum_{a=1}^A \eta_{ja} = 0$.

4. Area Level Predictors

Health disparities not explained by population composition (*i.e.*, by considering subject level risk factors alone) may be linked to area effects. For example, Do *et al.* [20] seek to estimate the share of racial health disparities that can be explained by differences in residential context. There are a wide range of potential area level risk factors for obesity, diabetes and related conditions that have been suggested or applied in the literature. These include area poverty and income levels [21], area racial composition [22,23], climate [24], income inequality [25,26], social cohesion [27,28], type of place (e.g., level of urbanicity) [29–31], and urban sprawl [32–34].

As to geographic effects in the BRFSS, these are defined by the lowest spatial scale identified by that study, namely the county of residence. In fact this means that there are two potentially relevant spatial divisions for the BRFSS data considered here, namely states and counties. There are 3,110 counties across the mainland US, albeit varying considerably in population size. The choice of known area predictors (G_K) in the current study is defined partly by availability of a complete and contemporary

profile of county level indicators; for example, some studies suggest potential effects of environmental pollution on diabetes [35,36], but a comprehensive US wide index of environmental quality is not available at county level.

Known county level predictors included in the regression in this paper are 2007 county poverty rate x_{1c} (as a proportion between 0 and 1), county population density x_{2c} (logarithmically transformed), and binary indicators (x_{3c}, x_{4c}, x_{5c}) for counties with proportions in the top decile of county population who are black, Hispanic and other nonwhite. Thus the 311 counties with proportions black exceeding the 90th percentile (over all 3,110 continental counties) are coded $x_{3c} = 1$, and other counties coded zero, $x_{3c} = 0$. All these variables have the advantage of being updateable between censuses, whereas a number of more complex indices (urban sprawl, social cohesion, etc) rely on 2,000 census variables in their construction. There are therefore five county predictors, each with outcome specific effects. These are represented by fixed effect parameters ($\delta_{j1}, \delta_{j2}, \delta_{j3}, \delta_{j4}, \delta_{j5}$), for categories $j = 1, \dots, J$, applying to the five county level predictors $\{x_{jc}, j = 1, \dots, 5; c = 1, \dots, 3110\}$.

To account for unmeasured (*i.e.*, omitted) area effects (G_L), a latent variable strategy is adopted. Given considerable evidence of spatial clustering in high levels of diabetes and obesity, this feature should be incorporated in the latent variable specification. One option is a separate random effect for each area and each outcome, but this would involve heavy parameterisation. The object of the method adopted here is a parsimonious summary of risks that tend to produce the well known clustering of both high obesity/overweight and high diabetes in certain parts of the US (*e.g.*, in the South East and Appalachians).

Specifically, a spatially correlated county effect v_c for counties $c = 1, \dots, 3110$ is adopted with loadings λ_j defining the impact of the shared county effect on weight-diabetes category j (see Appendix 1 for the form of the spatial dependence). A second set of spatially structured random effects u_s is defined according to state s of residence ($s = 1, \dots, 49$ including District of Columbia), with loadings κ_j defining the impact of that effect on category j . The latter model relatively broad scale and unmeasured effects for states. Identifiability is obtained by setting the first category loadings to 1, namely $\lambda_1 = \kappa_1 = 1$, so that the conditional variances of v_c and u_s are unknowns.

Let C_i and S_i denote the county and state of residence for respondents $i = 1, \dots, n$ where $n = 128, 150$, and let $\{a_i, r_i, e_i\}$ denote the age, race and education status of individual respondents. Then the regression terms ϕ_{ij} ($i = 1, \dots, n; j = 1, \dots, J$) defining the multinomial logit regression are represented in full form as:

$$\begin{aligned} \phi_{ij} = & \alpha_j + \beta_{j,r_i} + \gamma_{j,e_i} + \eta_{j,a_i} + \delta_{j1}x_{1,C_i} + \delta_{j2}x_{2,C_i} + \delta_{j3}x_{3,C_i} + \delta_{j4}x_{4,C_i} + \delta_{j5}x_{5,C_i} \\ & + \lambda_j v_{C_i} + \kappa_j u_{S_i}. \end{aligned} \quad (4)$$

Thus the model provides estimates both of the impacts of individual level risk factors and of area effects. Let $S_c \in (1, \dots, 49)$ denote the state that county c is located in. Then the composite county latent effect for joint prevalence category j is defined by the sum

$$t_{jc} = \lambda_j v_c + \kappa_j u_{S_c}, \quad (5.1)$$

and will incorporate both localized county effects, but also distinctive state level influences. In particular, the total county effect for obesity and diabetes combined is

$$t_{1c} = v_c + u_{S_c}. \quad (5.2)$$

5. Modelling Strategy and Distinct Geographic Effects

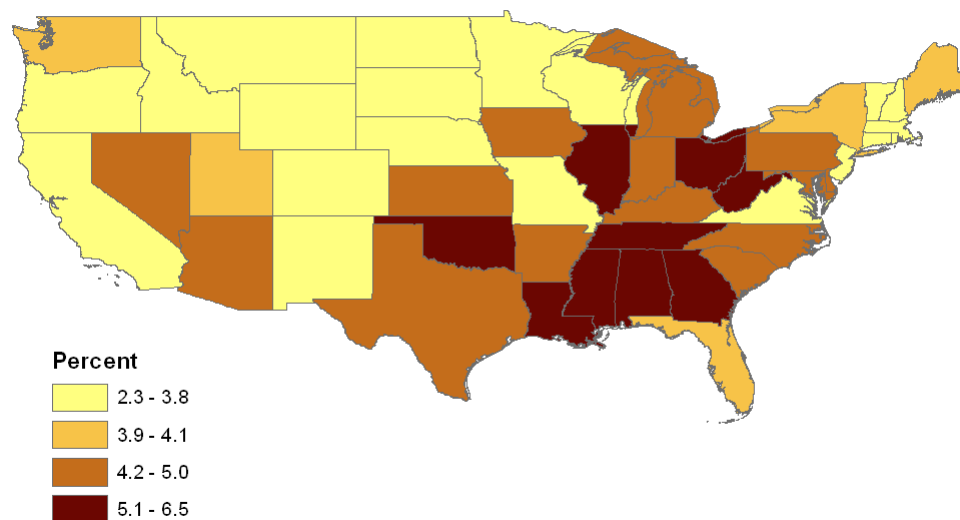
A major question of interest in multilevel modelling of health data is the presence (or otherwise) of distinct area effects, both effects of known area indicators (such as county poverty or population density) and effects of latent unmeasured area characteristics [20,37]. The presence of geographic contrasts is apparent from Table 3, which contains age standardized prevalence rates (as percents) for the six conditions by state; the nine census divisions of the states are also listed. For example, the highest rates of obesity & diabetes combined exceed 5.5% (e.g., in Tennessee, Mississippi, and Illinois) while the lowest rates are under 3%, for example, in Colorado and Montana (see Figure 1). Such variation may be due largely to differences in population composition, or there may be substantial area effects, and the role of such area effects is assessed here by using an incremental modelling strategy. Distinct area effects (sometimes called contextual effects), due either to known area covariates or latent area effects, are those remaining after the influence on prevalence of individual level attributes has been controlled for.

Table 3. Percentage prevalence of obesity-diabetes status, by race.

State	Census Division	Obese & Diabetic	Overweight & Diabetic	Normal weight & Diabetic	Obese non-diabetic	Overweight non-diabetic	Normal weight non-diabetic
Alabama	E South Central	5.7	2.8	1.3	23.6	39.6	27.0
Arizona	Mountain	4.2	3.6	1.9	25.9	37.8	26.6
Arkansas	W South Central	4.6	2.8	0.9	24.7	41.2	25.8
California	Pacific	3.6	2.8	2.0	20.5	39.8	31.2
Colorado	Mountain	2.3	2.8	1.0	18.0	42.3	33.7
Connecticut	New England	3.5	3.3	1.2	20.8	43.6	27.7
Delaware	South Atlantic	4.2	3.5	0.9	26.9	39.5	25.0
District of Columbia	South Atlantic	3.3	2.9	1.3	16.1	35.0	41.5
Florida	South Atlantic	4.0	3.2	1.5	21.6	41.4	28.3
Georgia	South Atlantic	5.1	3.2	1.7	20.8	39.8	29.3
Idaho	Mountain	3.6	2.9	1.0	21.1	43.5	27.8
Illinois	E North Central	6.0	4.2	1.2	20.4	40.0	28.2
Indiana	E North Central	4.7	2.8	1.5	20.7	42.2	28.2
Iowa	W North Central	4.2	2.4	1.3	25.1	38.4	28.5
Kansas	W North Central	4.4	2.8	0.9	24.7	39.9	27.3
Kentucky	E South Central	4.6	4.6	1.0	24.3	46.9	18.7
Louisiana	W South Central	5.7	3.4	1.0	26.1	35.9	27.8
Maine	New England	4.1	3.0	0.8	23.1	41.0	27.9
Maryland	South Atlantic	4.7	3.2	1.2	20.9	40.6	29.3
Massachusetts	New England	3.4	2.8	1.2	20.1	42.1	30.4
Michigan	E North Central	5.0	3.6	1.3	24.5	38.4	27.2
Minnesota	W North Central	3.6	2.1	0.7	24.4	40.5	28.7

Thus a baseline model estimates county level prevalence rates from a reduced version of the full model (4), including only subject level age and latent county and state effects. The resulting estimates of county prevalences of the different weight-diabetes categories are adjusted for age [38], but not for population differences in race and education composition, or for the effect of measured county level factors. Thus the baseline model (model 1) involves the regression terms

$$\phi_{ij} = \alpha_j + \eta_{j,a_i} + \lambda_j v_{C_i} + \kappa_j u_{S_i}, \quad j = 1, \dots, J \quad (6)$$

Figure 1. Percent of adults both diabetic & obese, BRFSS 2007.

which account for the differing age composition of survey subjects in different areas. Defining

$$\omega_{cj} = \exp(\alpha_j + \lambda_j v_c + \kappa_j u_{S_c}), \quad j = 1, \dots, J \quad (7)$$

$$\omega_{c,J+1} = 1,$$

age standardised proportions in counties $c = 1, \dots, 3110$, for the $J + 1 = 6$ weight-diabetes categories are then estimated as

$$p_{cj} = \omega_{cj} / \sum_{j=1}^{J+1} \omega_{cj}. \quad (8)$$

A second model (model 2) adds the effect of measured area predictors, namely county poverty, race composition and population density to the baseline model. Thus in model 2

$$\phi_{ij} = \alpha_j + \eta_{j,a_i} + \delta_{j1}x_{1,C_i} + \delta_{j2}x_{2,C_i} + \delta_{j3}x_{3,C_i} + \delta_{j4}x_{4,C_i} + \delta_{j5}x_{5,C_i} + \lambda_j v_{C_i} + \kappa_j u_{S_i}. \quad (9)$$

Age standardised prevalence rates by county and category under model 2 are estimated via $p_{cj} = \omega_{cj} / \sum_{j=1}^{J+1} \omega_{cj}$, where now

$$\omega_{cj} = \exp(\alpha_j + \delta_{j1}x_{1,C_i} + \delta_{j2}x_{2,C_i} + \delta_{j3}x_{3,C_i} + \delta_{j4}x_{4,C_i} + \delta_{j5}x_{5,C_i} + \lambda_j v_{C_i} + \kappa_j u_{S_i}). \quad (10)$$

These models are compared to the full model (model 3) including all subject level predictors (age, race, education) and both types of area effect (known and latent), namely

$$\phi_{ij} = \alpha_j + \beta_{j,r_i} + \gamma_{j,e_i} + \eta_{j,a_i} + \delta_{j1}x_{1,C_i} + \delta_{j2}x_{2,C_i} + \delta_{j3}x_{3,C_i} + \delta_{j4}x_{4,C_i} + \delta_{j5}x_{5,C_i} + \lambda_j v_{C_i} + \kappa_j u_{S_i}. \quad (11)$$

Of particular interest are changes in the level of variance of the latent area effects as county predictors and individual risk variables are added to the model. Also of interest are changes in the impact (and statistical significance) of known area predictors when individual level risk variables are added. For example, are there distinct county poverty effects on prevalence after individual level race and education level are allowed for?

6. Case Study Application

Fitting of the regression models and assessment of their goodness of fit follows a Bayesian approach. A Bayesian strategy is advantageous for estimating models with several sets of random effects, including random effects which are spatially clustered, especially when the responses (as here) are not continuous variables but discrete, namely a multinomial category. Under the Bayesian approach, prior densities are specified on all parameters in the model, and final (or posterior) estimates of parameters are based on the combination of the data likelihood and the prior densities.

Estimation uses iterative Monte Carlo Markov Chain (MCMC) sampling methods [39], as provided in the WINBUGS program [40]. Goodness of fit is assessed by a measure of fit that penalizes model complexity, known as the Deviance Information Criterion or *DIC* [41]. The *DIC* is obtained as the average deviance, using the definition (3.2), plus a measure of complexity. Lower values of the *DIC* indicate better fitting models. Posterior summaries of parameters are based on the 2nd half of runs of 10,000 iterations, using two chains starting from dispersed starting values. Convergence was achieved in all models using Brooks-Gelman-Rubin criteria [42].

Figure 2. County latent effects.

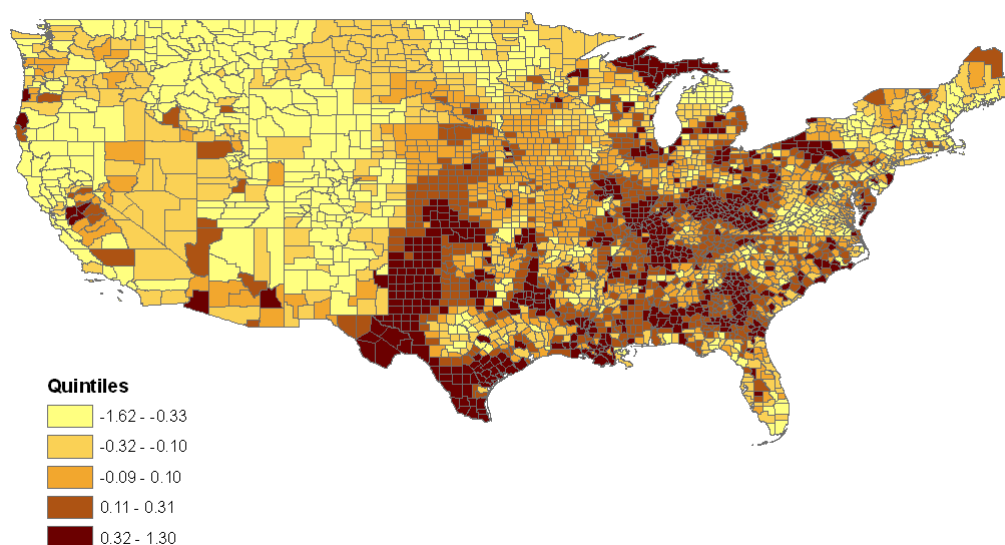
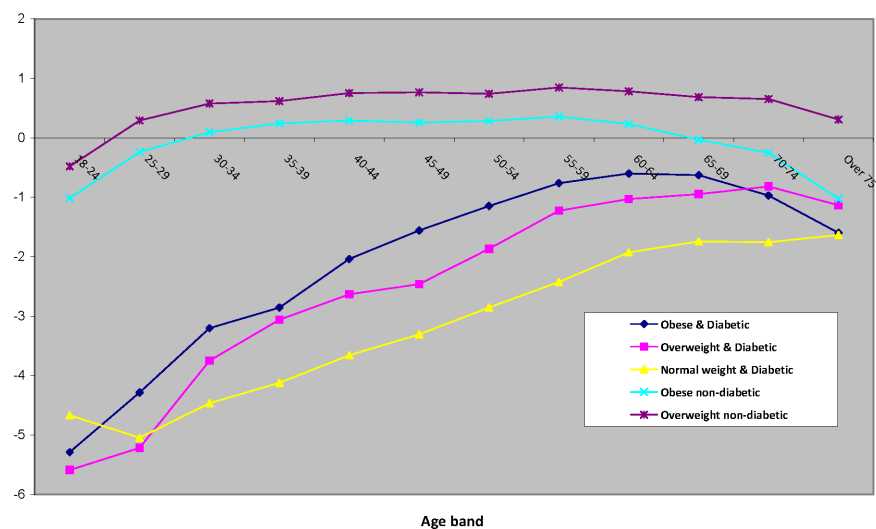


Figure 2 maps the composite latent county effects $t_{1c} = v_c + u_{S_c}$ from the baseline model 1 (these are posterior means from the MCMC sample). For example, $c = 1$ for Autauga County in Alabama, and Alabama is the first state alphabetically among the 49 states in the analysis, so $S_c = S_1 = 1$. The effects t_{1c} summarise varying risks for the jointly obese-diabetic condition between counties before controlling for factors such as county poverty and race composition. They show higher risks in the East South Central states (Kentucky, Tennessee, Alabama, Mississippi), and in some East North Central states (e.g., Illinois, Ohio) [43]. Model 1 also provides age profiles for the five diagnostic groups, plotted in Figure 3 as log-odds coefficients relative to the reference category. An increasing prevalence with age is confined to the categories obese & diabetic, overweight & diabetic, and normal weight & diabetic.

Figure 3. Age profiles of the different diagnostic groups.

One useful feature of this initial analysis is that the county effects can be profiled against known county and state level characteristics. For example, Figure 4 shows the profile of the average t_{1c} according to county poverty decile (defined by grouping counties into ten categories according to their ranked poverty rates).

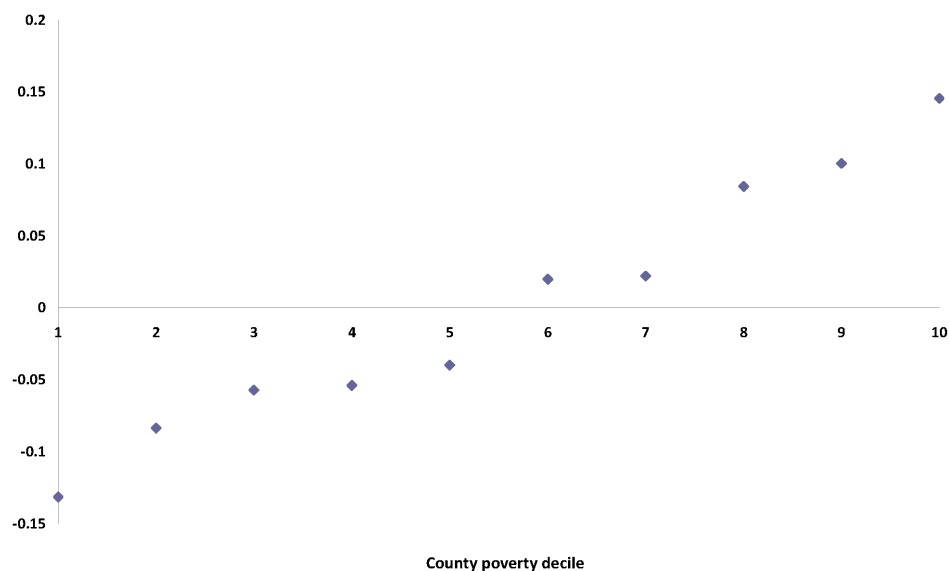
Figure 4. Total latent area effect, model 1, by county poverty decile.

Table 4 compares the fit from the baseline model against the other two models, and also presents the variances of the latent spatial effects. These need to account for differential loadings $\{\lambda_j, \kappa_j\}$ of the area effects by category (see Table 5), and for the distribution of subjects between categories, and are obtained marginally as $\text{var}(\lambda_j v_{C_i} + \kappa_j u_{S_i})$. It can be seen that adding known area predictors in Model 2 results in improved fit (a reduced DIC), and also in a (relatively slight) reduction in the variance of the latent spatial effects.

Table 4. Model fit summary.

	Average Deviance	Effective Parameters (Complexity)	DIC	Variance Spatial Effects
Model 1	260644	802	261446	0.329
Model 2	260581	821	261402	0.290
Model 3	256924	846	257769	0.279

Table 6 contains the δ_{jk} coefficients from model 2 ($j = 1, \dots, J; k = 1, \dots, 5$). It can be seen that the county poverty rate has a strong influence in raising chances of being both obese and diabetic (the first category). It is also an important positive influence on area relativities in the joint normal weight-diabetic category, and on obesity without diabetes. As Table 3 shows, the latter condition applies to around 22% of the US male population and occurs across relatively evenly the age spectrum. Table 6 also shows (via the coefficients δ_{j2}) higher rates of morbidity in lower density areas, typically non-metropolitan areas, for four of the five categories. This is consistent with findings that lower density areas, with greater sprawl and lower "walkability", have higher rates of obesity and overweight [44]. The exception to this effect is diabetes combined with normal weight, which is higher in more densely populated areas. As to effects of county ethnic structure, high concentrations of blacks or Hispanics remain a positive influence on the three morbidity categories involving diabetes, even after controlling for county poverty.

Figure 5. Rate of obesity & diabetes jointly, US Counties, Model 2.

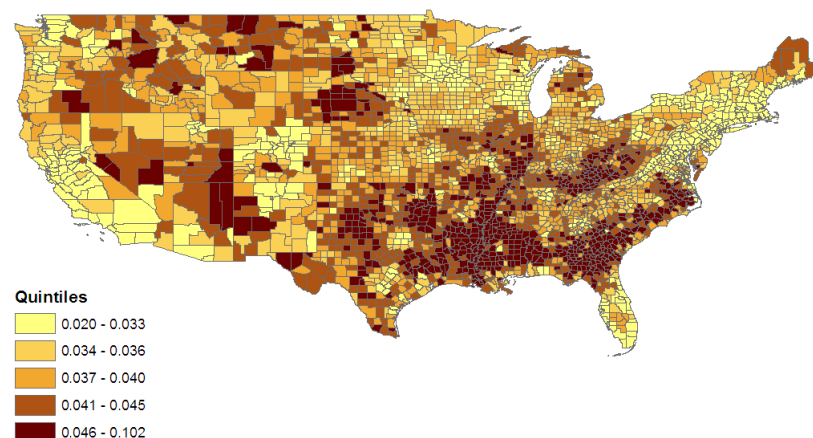


Table 5. Loadings on shared latent effects, by model.

Model	Level	Category	Mean	2.5%	97.5%
1	County	Obese & Diabetic	1		
		Overweight & Diabetic	0.72	0.58	0.87
		Normal weight & Diabetic	0.49	0.32	0.67
		Obese non-diabetic	1.42	1.32	1.54
		Overweight non-diabetic	0.88	0.81	0.96
	State	Obese & Diabetic	1		
		Overweight & Diabetic	0.88	0.62	1.20
		Normal weight & Diabetic	0.44	0.08	0.84
		Obese non-diabetic	0.16	0.03	0.31
		Overweight non-diabetic	0.05	0.00	0.13
2	County	Obese & Diabetic	1.00		
		Overweight & Diabetic	1.07	0.97	1.18
		Normal weight & Diabetic	1.53	1.35	1.69
		Obese non-diabetic	2.56	2.50	2.65
		Overweight non-diabetic	1.75	1.67	1.85
	State	Obese & Diabetic	1		
		Overweight & Diabetic	1.08	0.96	1.29
		Normal weight & Diabetic	0.14	-0.02	0.30
		Obese non-diabetic	0.86	0.80	0.93
		Overweight non-diabetic	0.42	0.37	0.48
3	County	Obese & Diabetic	1		
		Overweight & Diabetic	1.25	1.02	1.52
		Normal weight & Diabetic	1.27	0.95	1.55
		Obese non-diabetic	2.44	2.16	2.83
		Overweight non-diabetic	1.65	1.44	1.91
	State	Obese & Diabetic	1		
		Overweight & Diabetic	1.09	0.68	1.74
		Normal weight & Diabetic	0.45	0.12	0.64
		Obese non-diabetic	0.48	0.32	0.69
		Overweight non-diabetic	0.30	0.14	0.48

Figure 5 maps county level variations in proportions jointly obese & diabetic from model 2, namely

$$p_{c1} = \omega_{c1} / \sum_{j=1}^6 \omega_{cj}, \quad (12)$$

where the ω_{cj} are as in (10). The role of county poverty in defining levels of the joint obese-diabetic category under model 2 results in isolated high prevalence clusters in West North Central states such as North Dakota, Montana and Nebraska. These may, for example, be low income rural areas or counties with concentrations of native Americans [45]. Figure 6 maps variations in the prevalence of obesity

Table 6. Effects of county predictors by diabetes-weight category, model 2.

Category	Predictor	Posterior Mean	Standard Devn	MC Error	2.5%	97.5%
Obese & Diabetic	Poverty Rate	1.10	0.08	0.01	0.93	1.21
	Popn Density	-0.11	0.01	0.00	-0.12	-0.10
	Top decile Black	0.31	0.05	0.01	0.21	0.41
	Top decile Hispanic	0.17	0.06	0.01	0.93	0.31
	Top decile Other Ethnicity	-0.12	0.06	0.01	-0.23	-0.02
Overweight & Diabetic	Poverty Rate	-0.10	0.09	0.02	-0.27	0.09
	Popn Density	-0.04	0.01	0.00	-0.06	-0.03
	Top decile Black	0.02	0.09	0.02	-0.23	0.16
	Top decile Hispanic	0.30	0.05	0.01	0.19	0.40
	Top decile Other Ethnicity	-0.16	0.05	0.01	-0.24	-0.06
Normal weight & Diabetic	Poverty Rate	0.46	0.12	0.02	0.23	0.71
	Popn Density	0.04	0.01	0.00	0.01	0.06
	Top decile Black	0.19	0.08	0.02	0.01	0.33
	Top decile Hispanic	0.37	0.07	0.01	0.27	0.49
	Top decile Other Ethnicity	-0.14	0.05	0.01	-0.25	-0.04
Obese non-diabetic	Poverty Rate	0.16	0.14	0.02	0.01	0.41
	Popn Density	-0.06	0.01	0.00	-0.08	-0.05
	Top decile Black	0.02	0.02	0.00	-0.03	0.08
	Top decile Hispanic	0.09	0.03	0.01	0.04	0.18
	Top decile Other Ethnicity	-0.15	0.05	0.01	-0.23	-0.05
Overweight non-diabetic	Poverty Rate	-0.42	0.10	0.02	-0.64	-0.24
	Popn Density	-0.01	0.01	0.00	-0.02	0.00
	Top decile Black	-0.08	0.03	0.01	-0.16	-0.03
	Top decile Hispanic	0.01	0.02	0.00	-0.04	0.06
	Top decile Other Ethnicity	-0.08	0.02	0.00	-0.12	-0.03

without diabetes, namely

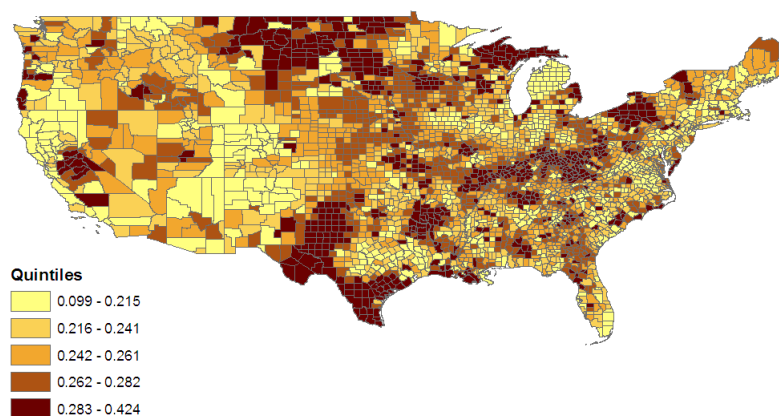
$$p_{c4} = \omega_{c4} / \sum_{j=1}^6 \omega_{cj}. \quad (13)$$

The geographic pattern of this condition broadly resembles that of the rarer joint obese-diabetic condition; the state level correlation between these two sets of prevalence rates is 0.50.

As might be expected, combining individual and county level predictors in model 3 produces the lowest DIC and a reduced spatial variance, though over 80% in the baseline spatial variance remains unexplained. Table 7 summarises the effects of individual level risk factors under model 3, in terms of relativities between education and race groups for each of the five morbidity categories. These are represented by the education parameters γ_{je} ($e = 2, \dots, 4$), and race parameters β_{jr} ($r = 2, 4$); the reference coefficient for education is $\gamma_{j1} = 0$ for less than high school, and the reference coefficient for education is $\beta_{j1} = 0$ for white non-Hispanics.

A notable feature from the education parameters is the lower morbidity among college graduates. Generally, morbidity is greater for subjects with lower education attainment, except for the overweight-non diabetic category.

The race parameters in Table 7 show that black and Hispanic males have higher morbidity than white non-Hispanic males for all conditions. By contrast, other ethnicity (primarily Asian Americans

Figure 6. Rate of obesity without diabetes, US counties, model 2.**Table 7.** Effects of county predictors by diabetes-weight category, model 2.

Education Coefficients		Posterior Mean	Standard Devn	MC Error	2.5%	97.5%
Obese & Diabetic	High School Graduate	-0.09	0.05	0.01	-0.18	0.01
	Some College	0.02	0.05	0.01	-0.07	0.13
	College Graduate	-0.66	0.05	0.01	-0.74	-0.56
Overweight & Diabetic	High School Graduate	0.13	0.05	0.01	0.03	0.23
	Some College	0.02	0.05	0.01	-0.08	0.12
	College Graduate	-0.25	0.05	0.01	-0.34	-0.14
Normal weight & Diabetic	High School Graduate	-0.12	0.08	0.01	-0.30	0.04
	Some College	-0.34	0.09	0.01	-0.53	-0.18
	College Graduate	-0.50	0.07	0.01	-0.64	-0.38
Obese non-diabetic	High School Graduate	0.26	0.03	0.00	0.20	0.32
	Some College	0.37	0.03	0.00	0.30	0.43
	College Graduate	-0.16	0.03	0.00	-0.23	-0.10
Overweight non-diabetic	High School Graduate	0.15	0.02	0.00	0.11	0.19
	Some College	0.29	0.02	0.00	0.24	0.34
	College Graduate	0.15	0.02	0.00	0.10	0.19
Race Coefficients		Posterior Mean	Standard Devn	MC Error	2.5%	97.5%
Obese & Diabetic	Black	0.60	0.06	0.01	0.48	0.73
	Hispanic	0.74	0.06	0.01	0.62	0.85
	Other	-0.12	0.05	0.01	-0.21	-0.01
Overweight & Diabetic	Black	0.81	0.08	0.01	0.63	0.95
	Hispanic	0.87	0.06	0.01	0.76	1.01
	Other	-0.01	0.06	0.01	-0.13	0.10
Normal weight & Diabetic	Black	0.56	0.11	0.01	0.35	0.79
	Hispanic	0.67	0.11	0.01	0.46	0.88
	Other	0.34	0.09	0.01	0.18	0.50
Obese non-diabetic	Black	0.39	0.03	0.00	0.33	0.44
	Hispanic	0.48	0.03	0.00	0.43	0.54
	Other	-0.63	0.03	0.00	-0.70	-0.57
Overweight non-diabetic	Black	0.08	0.03	0.00	0.03	0.14
	Hispanic	0.41	0.02	0.00	0.36	0.46
	Other	-0.31	0.03	0.00	-0.36	-0.26

and native Americans) enhances only the rate of diabetes without obesity. This is consistent with the original survey data (see Table 2) which shows that other ethnic groups have the highest proportion of all race groups in the non-morbid (normal weight, non-diabetic) category. The high rate of diabetes without excess weight among asian Americans has been shown by other studies [43,46]. Although other studies [47] show high obesity among native Americans, the results in Table 7 suggest this may to a considerable extent be explained by socioeconomic status (as measured by education) and by area effects.

Table 8. Effects of county predictors by diabetes-weight category, model 3.

Category	Predictor	Posterior Mean	Standard Devn	MC Error	2.5%	97.5%
Obese & Diabetic	Poverty Rate	2.07	0.15	0.02	1.80	2.39
	Popn Density	-0.11	0.01	0.00	-0.13	-0.09
	Top decile Black	0.07	0.06	0.01	-0.06	0.18
	Top decile Hispanic	-0.13	0.06	0.01	-0.24	-0.02
	Top decile Other Ethnicity	-0.08	0.05	0.01	-0.17	0.01
Overweight & Diabetic	Poverty Rate	0.58	0.26	0.03	0.10	1.06
	Popn Density	-0.07	0.01	0.00	-0.09	-0.04
	Top decile Black	-0.28	0.08	0.01	-0.41	-0.12
	Top decile Hispanic	0.03	0.06	0.01	-0.07	0.15
	Top decile Other Ethnicity	-0.12	0.06	0.01	-0.25	-0.03
Normal weight & Diabetic	Poverty Rate	0.07	0.33	0.04	-0.64	0.62
	Popn Density	0.01	0.01	0.00	-0.02	0.04
	Top decile Black	-0.04	0.11	0.01	-0.26	0.14
	Top decile Hispanic	0.12	0.07	0.01	-0.03	0.24
	Top decile Other Ethnicity	-0.15	0.06	0.01	-0.30	-0.05
Obese non-diabetic	Poverty Rate	1.06	0.13	0.02	0.79	1.28
	Popn Density	-0.08	0.02	0.00	-0.12	-0.06
	Top decile Black	-0.09	0.06	0.01	-0.19	0.04
	Top decile Hispanic	-0.25	0.10	0.01	-0.42	-0.08
	Top decile Other Ethnicity	-0.17	0.04	0.00	-0.25	-0.09
Overweight non-diabetic	Poverty Rate	0.15	0.17	0.02	-0.12	0.52
	Popn Density	-0.04	0.01	0.00	-0.06	-0.02
	Top decile Black	-0.08	0.04	0.00	-0.15	0.00
	Top decile Hispanic	-0.20	0.06	0.01	-0.31	-0.09
	Top decile Other Ethnicity	-0.09	0.03	0.00	-0.15	-0.04

Table 8 contains the δ_{jk} coefficients relating to county level predictors under model 3. The effects of county poverty rate remain pronounced, and are in fact enhanced for the joint obese-diabetic and obese-non-diabetic categories. The significantly higher prevalence of all conditions (except diabetic normal weight) in lower density counties is also still evident. Thus the effects of known area predictors have been largely maintained after allowing for subject level race and education, established as major individual level risk factors for the two conditions [48]. The reduction (relatively slight) in the variance of latent area effects (see Table 4) under model 3 may then be mainly attributable to control for population composition.

7. Conclusion

This paper has considered an approach to modelling prevalence variations in diseases or conditions considered jointly, taking account of both area effects and characteristics of survey subjects. The influence of area is represented partly by known variables (e.g., county poverty), and partly by spatially clustered latent area influences. The application has focussed on the joint prevalence of diabetes and weight status, so providing a geographic perspective on weight-related diabetes prevalence. However, the approach is generic and potentially extends to more than two conditions.

Geographic variability in chronic conditions whether considered singly or jointly will partly reflect variations in the socio-demographic characteristics of area populations, sometimes termed ‘compositional’ effects [49]. However, a number of studies find evidence for prevalence variations between different areas even after controlling for population composition, illustrating what are sometimes termed ‘contextual’ effects [50]. The present study adds to this evidence by showing enduring geographic contrasts in prevalence of different joint obesity-diabetes categories after taking account of individual level age, race and education status.

In the present paper contextual effects have been represented by shared latent effects over the joint obesity-diabetes categories. These are spatially structured random effects for counties and states, and the consistently positive loadings in Table 5 demonstrate that a shared univariate effect is appropriate. Elaborations to the model presented above are possible, such as ethnic group differentiation in age or education gradients, or additional subject level predictors, though those included (age, race, education) are established as the major dimensions of variation for diabetes and obesity [5,48]. One might also assume spatially varying impacts of the known area predictors, such as the county poverty rate [51].

References

1. Balluz, L.; Okoro, C.; Mokdad, A. Association between selected unhealthy lifestyle factors, body mass index, and chronic health conditions among individuals 50 years of age or older, by race/ethnicity. *Ethn. Dis.* **2008**, *18*, 450-457.
2. Yach, D.; Stuckler, D.; Brownell, K. Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. *Nat. Med.* **2006**, *12*, 62-663.
3. Mokdad, A.; Ford, E.; Bowman, B.; Nelson, D.; Engelgau M.; Vinicor F.; Marks J. Diabetes trends in the US: 1990–1998. *Diabetes Care* **2001**, *24*, 1278-1283.
4. Gregg, E.; Cheng, Y.; Narayan, K.; Thompson, T.; Williamson, D. The relative contributions of different levels of overweight and obesity to the increased prevalence of diabetes in the United States: 1976–2004. *Prev. Med.* **2007**, *45*, 348-352.
5. Paeratakul, S.; Lovejoy, J.; Ryan, D.; Bray, G. The relation of gender, race and socioeconomic status to obesity and obesity comorbidities in a sample of US adults. *Int. J. Obes. Relat. Metab. Disord.* **2002**, *26*, 1205-1210.
6. Cowie, C.; Rust, K.; Ford, E.; Eberhardt, M.; Byrd-Holt, D.; Li, C.; Williams, D.; Gregg, E.; Bainbridge, K.; Saydah, S.; Geiss, L. Full accounting of diabetes and pre-diabetes in the US population in 1988–1994 and 2005–2006. *Diabetes Care* **2009**, *32*, 287-294.

7. Krieger, N.; Chen, J.; Waterman, P.; Rehkopf, D.; Subramanian, S. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures. *Amer. J. Public Health* **2003**, *93*, 1655-1671.
8. Drewnowski, A.; Rehm, C.; Solet, D. Disparities in obesity rates: analysis by ZIP code area. *Soc. Sci. Med.* **2007**, *65*, 2458-2463.
9. Lee, R.; Cubbin, C.; Winkleby, M. Contribution of neighbourhood socioeconomic status and physical activity resources to physical activity among women. *J. Epid. Comm. Health* **2007**, *61*, 882-890.
10. Cubbin, C.; Hadden, W.; Winkleby, M. Neighborhood context and cardiovascular disease risk factors: the contribution of material deprivation. *Ethn. Dis.* **2001**, *11*, 687-700.
11. Subramanian, S.; Chen, J.; Rehkopf, D.; Waterman, P.; Krieger, N. Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts. *Am. J. Public Health* **2005**, *95*, 260-265.
12. Schuurman, N.; Peters, P.; Oliver, L. Are obesity and physical activity clustered?. A spatial analysis linked to residential density. *Obesity* **2009**, doi:10.1038/oby.2009.119.
13. Saaddine, J.; Narayan, K.; Engelgau, M.; Aubert, R.; Klein, R.; Beckles, G. Prevalence of self-rated visual impairment among adults with diabetes. *Am. J. Public Health* **1999**, *89*, 1200-1205.
14. Jiles, R.; Hughes, E.; Murphy, W.; Flowers, N.; McCracken, M.; Roberts, H.; Ochner, M.; Balluz, L.; Mokdad, A.; Elam-Evans, L.; Giles, W. Surveillance for certain health behaviors among states and selected local areas—Behavioral Risk Factor Surveillance System. *MMWR Surveill Summ.* **2005**, *54*, 1-116.
15. Hedeker, D. A mixed-effects multinomial logistic regression model. *Stat. Med.* **2003**, *22*, 1433-1444.
16. Center for Disease Control and Prevention (CDC). Prevalence of diabetes and impaired fasting glucose in adults—United States, 1999-2000. *MMWR Surveill Summ.* **2003** *52*, 833-837.
17. Freudenberg, N.; Ruglis, J. Reframing school dropout as a public health issue. *Prev. Chronic Dis.* **2007**, *7*, 63.
18. Zhang, Q.; Wang, Y.; Huang, E. Changes in racial/ethnic disparities in the prevalence of Type 2 diabetes by obesity level among US adults. *Ethn. Health* **2009**, *14*, 439-457.
19. Maty, S.; Everson-Rose, S.; Haan, M.; Raghunathan, T.; Kaplan, G. Education, income, occupation, and the 34-year incidence (1965–1999) of Type 2 diabetes in the Alameda County Study. *Int. J. Epid.* **2005**, *34*, 1282-1283.
20. Do, D.; Finch, B.; Basurto-Davila, R.; Bird, C.; Escarce, J.; Lurie, N. Does place explain racial health disparities? Quantifying the contribution of residential context to the Black/white health gap in the United States. *Soc. Sci. Med.* **2008**, *67*, 1258-1268.
21. Schwartz, F.; Ruhil, A.; Denham, S.; Shubrook, J.; Simpson, C.; Boyd, S. High self-reported prevalence of diabetes mellitus, heart disease, and stroke in 11 counties of rural Appalachian Ohio. *J. Rur. Health* **2009**, *25*, 226-230.
22. Mellor, J.; Milyo, J. Individual health status and racial minority concentration in US states and counties. *Am. J. Public Health* **2004**, *94*, 1043-1048.
23. Lopez, R. Neighborhood risk factors for obesity. *Obesity* **2007**, *15*, 2111-2119.

24. Franz, K.; Bailey, S. Geographical variations in heart deaths and diabetes: effect of climate and a possible relationship to magnesium. *J. Amer. Coll. Nutr.* **2004**, *23*, 521S-524S.
25. Pickett, K.; Kelly, S.; Brunner, E.; Lobstein, T.; Wilkinson, R. Wider income gaps, wider waistbands? An ecological study of obesity and income inequality. *J. Epid. Comm. Health* **2005**, *59*, 670-674.
26. Fuller-Thomson, E.; Gadalla, T. Income inequality and limitations in activities of daily living: a multilevel analysis of the 2003 American Community Survey. *Public Health* **2008**, *122*, 221-228.
27. Holtgrave, D.; Crosby, R. Is social capital a protective factor against obesity and diabetes? Findings from an exploratory study. *Ann. Epidemiol.* **2006**, *16*, 406-408.
28. Kim, D.; Subramanian, S.; Gortmaker, S.; Kawachi, I. US state- and county-level social capital in relation to obesity and physical inactivity: a multilevel, multivariable analysis. *Soc. Sci. Med.* **2006**, *63*, 1045-1059.
29. Mainous, A.; King, D.; Garr, D.; Pearson, W. Race, rural residence, and control of diabetes and hypertension. *Ann. Fam. Med.* **2004**, *2*, 563-568.
30. Koopman, R. Mainous, A.; Geesey, M. Rural residence and Hispanic ethnicity: doubly disadvantaged for diabetes? *J. Rur. Health* **2005**, *22*, 63-68.
31. Lovasi, G.; Neckerman, K.; Quinn, J.; Weiss, C.; Rundle, A. Effect of individual or neighborhood disadvantage on the association between neighborhood walkability and body mass index. *Amer. J. Public Health* **2009**, *99*, 279-284.
32. Ewing, R.; Schmid, T.; Killingsworth, R. Relationship between urban sprawl and physical activity, obesity, and morbidity. *Amer. J. Health Promot.* **2003**, *18*, 47-57.
33. Joshi, C.; Boehmer, T.; Brownson, R.; Ewing, R. Personal, neighbourhood and urban factors associated with obesity in the United States. *J. Epid. Comm. Health* **2008**, *62*, 202-208.
34. Li, F.; Harmer, P.; Cardinal, B.; Bosworth, M.; Johnson-Shelton, D.; Moore, J.; Acock, A.; Vongjaturapat, N. Built environment and 1-year change in weight and waist circumference in middle-aged and older adults: Portland Neighborhood Environment and Health Study. *Amer. J. Epid.* **2009**, *169*, 401-408.
35. Ershow, A. Environmental influences on development of type 2 diabetes and obesity: challenges in personalizing prevention and management. *J. Diab. Sci. Tech.* **2009**, *3*, 727-734.
36. Schreinemachers D. Mortality from ischemic heart disease and diabetes mellitus (type 2) in four U.S. wheat-producing states: a hypothesis-generating study. *Environ. Health Perspect.* **2006**, *114*, 186-193.
37. Sastry, N.; Hussey, J. An investigation of race and ethnic disparities in birthweight in Chicago neighborhoods. *Demography* **2003**, *40*, 701-725.
38. Gregg, E.; Cheng, Y.; Cadwell, B.; Imperatore, G.; Williams, D.; Flegal, K.; Narayan, K.; Williamson, D. Secular trends in cardiovascular disease risk factors according to body mass index in US adults. *J. Amer. Med. Assoc.* **2005**, *293*, 1868-1874.
39. Gelfand, A.; Smith, A. Sampling based approaches to calculate marginal densities. *J. Amer. Statist. Assoc.* **1990**, *85*, 398-409.
40. Lunn, D.; Thomas, A.; Best, N.; Spiegelhalter, D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.* **2000**, *10*, 325-337.

41. Spiegelhalter, D.; Best, N.; Carlin, B.; van der Linde, A. Bayesian measures of model complexity and fit. *J. Roy. Stat. Soc. B* **2002**, *64*, 583-639.
42. Brooks, S.; Gelman, A. Alternative methods for monitoring convergence of iterative simulations. *J. Comp. Graph. Stat.* **1998**, *7*, 434-456.
43. Wang, Y.; Beydoun, M. The obesity epidemic in the United States—gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. *Epid. Rev.* **2007**, *29*, 6-28.
44. Papas, M.; Alberg, A.; Ewing, R.; Helzlsouer, K.; Gary, T.; Klassen, A. The built environment and obesity. *Epid. Rev.* **2007**, *29*, 129-143.
45. Kirschner, A. Poverty in the rural West. *Perspect. Poverty Policy Place* **2005**, *3*, 4-6.
46. McNeely, M.; Boyko, E. Type 2 diabetes prevalence in Asian Americans: results of a national health survey. *Diabetes Care* **2004**, *27*, 66-69.
47. Broussard, B.; Johnson, A.; Himes, J.; Story, M.; Fichtner, R.; Hauck, F.; Bachman-Carter, K.; Hayes, J.; Frohlich, K.; Gray, N. Prevalence of obesity in American Indians and Alaska Natives. *Amer. J. Clin. Nutr.* **1991**, *53*, 1535S-1542S.
48. Geiss, L.; Pan, L.; Cadwell, B.; Gregg, E.; Benjamin, S.; Engelgau, M. Changes in incidence of diabetes in U.S. adults, 1997-2003. *Amer. J. Prev. Med.* **2006**, *30*, 371-377.
49. Duncan, C.; Jones, K.; Moon, G. Context, composition and heterogeneity: using multilevel models in health research. *Soc. Sci. Med.* **1998**, *46*, 97-117.
50. Sacker, A.; Wiggins, R.; Bartley, M. Time and place: putting individual health into context. A multilevel analysis of the British household panel survey, 1991–2001. *Health Place* **2006**, *12*, 279-290.
51. Gamerman, D.; Moreira, A.; Rue, H. Space-varying regression models specifications and simulation. *Comput. Statist. Data Analysis* **2003**, *42*, 513-533.
52. Fahrmeir, L.; Lang, S. Bayesian inference for generalized additive mixed models based on Markov random field priors. *J. Roy. Stat. Soc. C* **2001**, *50*, 201-220.
53. Besag, J.; York, J.; Mollie, A. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **1991**, *43*, 1-59.

Appendix 1: Structured Age and Area Effects

Age and area effects follow autoregressive random effect schemes. To pool strength across the age profiles of different outcomes, a low order multivariate random walk prior [52] may be adopted for the J dimensional vector $\eta_a = (\eta_{1a}, \dots, \eta_{Ja})$, $a = 1, \dots, A$. For example, first and second order random walk priors have conditional forms

$$\eta_a \sim N_J(\eta_{a-1}, \Omega_\eta^{-1}),$$

$$\eta_a \sim N_J(2\eta_{a-1} - \eta_{a-2}, \Omega_\eta^{-1}),$$

where the $J \times J$ matrix Ω_η^{-1} represents covariation between age mortality profiles over different weight-diabetes status categories. Here a first order random walk is used, and the precision matrix Ω_η is assigned a Wishart prior with identity scale matrix and J degrees of freedom, namely $\Omega_\eta \sim \text{Wish}(I, J)$.

The county and state effects follow conditional autoregressive (CAR) priors [53], with pooling of strength over neighbouring areas. Suppose the locality L_c of county c (the counties adjacent to it) contains d_c counties. For v_c conditional on all remaining effects $v_{[c]} = (v_1, \dots, v_{c-1}, v_{c+1}, \dots, v_n)$, one has

$$v_c | v_{[c]} \sim N(V_c, \frac{\delta}{d_c})$$

where δ is a conditional variance parameter, and V_c is the average of the v_h ($h \neq c$) in locality L_c , namely

$$V_c = \sum_{h \in L_c} v_h / d_c.$$

The state level prior pools strength over neighbouring states in the same way.

© 2010 by the author; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license <http://creativecommons.org/licenses/by/3.0/>.