Spatio-temporal modeling of odds of disease

Mahmoud Torabi *†

The analysis of geographical and temporal variability of binomial data, using generalized additive mixed models, are considered. In this class of models, spatially correlated random effects and temporal components are adopted. The frequentist analysis of these complex models is computationally difficult. Recently developed method of data cloning has overcome the computational challenges of the analysis of mixed models from the frequentist approach. We use data cloning, which yields to maximum likelihood estimation, to propose frequentist analysis of spatio-temporal modeling of odds of disease. The advantages of the data cloning approach are that the prediction and prediction interval of the smoothing odds over space and time are easily obtained. We illustrate this approach using a real dataset of yearly asthma physician visits by children in the province of Manitoba, Canada, during 2000-2010. The performance of the proposed approach is also studied through a simulation study.

Keywords: binomial data; conditional autoregressive; geographic epidemiology; penalized spline; random effects; time series

*Correspondence to: Mahmoud Torabi, Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, R3E 0W3, Canada. E-mail: torabi@cc.umanitoba.ca

†Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada

1. INTRODUCTION

The analysis of disease disparities over space and time has received considerable attention since it is crucial for health systems to assess the distributive impact within systems of public policies in relation to spatial or health status inequalities. The use of generalized linear or additive mixed Poisson model is very common for rare events in a population health context to construct new methodologies or to apply existing methodologies for spatial and temporal models. However, in studies of health system services, using the mixed Poisson baseline structure may not be appropriate as events are not typically rare or contagious. We then need to consider spatio-temporal modeling of odds and study the risk factors in the mixed binomial model setting.

The idea behind developments on spatial and spatio-temporal modeling of odds is essentially to model variations in true odds and better separate systematic variability from random noise, a component that usually overshadows crude odds maps. Maps of regional odds of incidence and mortality over time are useful tools in determining spatial and temporal patterns of disease. The odds of incidence or mortality may differ substantially across geographical regions. A reliable estimate of the underlying risk of disease is usually provided by *borrowing strength* from neighboring geographic sub-regions.

Waller *et al.* (1997) proposed Bayesian Poisson spatio-temporal models to account for spatial and temporal random effects as well as spatio-temporal interactions. A unified approach for a Bayesian analysis of incidence or mortality data in space and time was proposed by Knorr-Held (2000). MacNab and Dean (2001), Torabi and Rosychuk (2011, 2012) and Torabi (2013) proposed Poisson spatio-temporal models that use autoregressive (AR) local smoothing across the spatial effects and B-spline smoothing over the temporal effects. Martínez-Beneito *et al.* (2008) suggested an AR Poisson spatio-temporal model based on time series and spatial modeling using a Bayesian approach to link information in time and space. In some contexts (e.g., health conditions), the underlying rates may

change over seasons within a given year. Torabi and Rosychuk (2010) proposed a Poisson spatio-temporal model using a frequentist approach (generalized estimation equation approach) to account for spatial, temporal as well as seasonal effects. Torabi (2012a) proposed spatio-temporal models that use AR smoothing across the spatial effects, random walk smoothing over the temporal effects, and a smoothing function to account for seasonal effects.

It is also important to study inference for additive mixed models in mapping health service data (e.g., asthma physician visits) when using a Poisson distribution may not be appropriate as an approximate to binomial distribution. MacNab (2003) studied spatial random effects modeling for non-rare diseases using a hybrid algorithm as a computational alternative to fully Bayesian spatial analysis of binomial data. Knorr-Held and Besag (1998) considered spatio-temporal binomial data with incorporating time as a categorical variable and studied the role of time- and space-varying covariate effects from a Bayesian perspective. Recently, Silva *et al.* (2008) developed a fully Bayesian binomial spatio-temporal approach by using an AR local smoothing across the spatial effects and a B-spline smoothing over the temporal effects to study trends of odds, and produced smoothed maps including regional effects.

There are many different ways to perform inference in mixed models, however, the frequentist approach has been computationally difficult particularly for generalized additive mixed models (GAMMs). In the last two decades, many approximate approaches have been proposed. For instance, one may use penalized quasi-likelihood (PQL) (Breslow and Clayton, 1993) which may work well for Poisson mixed models. However, such methods have been shown to perform poorly for the binomial setting (Breslow and Clayton, 1993; Lin and Breslow, 1996).

Lele *et al.* (2007) recently introduced an alternative frequentist approach, called data cloning (DC), to compute the maximum likelihood estimates (MLE) and their standard

errors for general hierarchical models. The DC approach, which uses Bayesian tools, avoids high dimensional numerical integration and requires neither maximization nor differentiation of a function. In the method of DC, non-estimable parameters are also flagged automatically. Lele *et al.* (2010) described an approach to compute prediction and prediction interval of random effects in the context of generalized linear mixed models (GLMMs). Recently, Torabi (2012b, 2013) considered the use of DC method in the areas of Poisson spatial and spatio-temporal models, respectively. Thus, the DC approach is well suited to offer a frequentist approach for the analysis of binomial spatio-temporal models.

In this paper, we propose the DC approach in the context of binomial spatio-temporal models. In our spatio-temporal model, the well-known conditional autoregressive (CAR) model (Besag *et al.*, 1991) and penalized spline (P-spline) models (Eilers and Marx, 1996) are used for the spatial and temporal effects, respectively (Section 2). We then describe how DC can be used to obtain estimate of model parameters and also to predict the odds (Section 3). In Section 4, the performance of the proposed approach is evaluated using a real dataset of yearly number of asthma physician visits by children in the province of Manitoba, Canada, during 2000-2010, and also by a simulation study. Concluding remarks are given in Section 5.

2. SPATIO-TEMPORAL MODEL OF ODDS

Let y_{it} be the number of what will be broadly termed "successes" out of n_{it} "trials" for the ith geographic region at time $t, (i = 1, ..., I; t = 1, ..., T)$, where y_{it} is the total number of individuals for whom a binary value is unity. For example, y_{it} is a physician visit in our application and n_{it} stands for the size of a population for whom the binary outcome was recorded at the individual level. We assume that y_{it}, conditional on p_{it}, has a binomial distribution with parameters n_{it} and p_{it}, where p_{it} denotes the probability of success. One

may use a Poisson approximation to the binomial if events are rare. However, in many health service settings, events such as physician visits are not rare. A generalized additive spatio-temporal model is then given by

$$logit(p_{it}) \equiv \log\{p_{it}/(1-p_{it})\} = m + S(t) + \eta_i + \theta_{it}, \tag{1}$$

where m reflects the overall mean log-odds over time and region, $S(t)$ is the effect of time t, η_i represents spatial structure of region i, and θ_{it} is the interaction between the spatial and temporal effects. The P-spline models (Eilers and Marx, 1996) are used to account for the temporal effects denoted by $S(t)$. With the overall mean log-odds, m, in our model, the P-spline is provided without an intercept. In this case, $S(t)$ is given by

$$S(t) = \beta_1 t + \beta_2 t^2 + ... + \beta_p t^p + \sum_{k=1}^{K} \gamma_k (t-\kappa_k)_+^p, \tag{2}$$

where p is the degree of spline, $(t)_+^p$ denotes the function $t^p I_{\{t>0\}}, \kappa_1 < ... < \kappa_K$ is a set of fixed knots, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_K)'$ with $\gamma_k \overset{i.i.d.}{\sim} N(0, \sigma_{sp}^2)$ are the coefficient vectors for the parametric and spline parts of the model, respectively. It is well known that with spreading out the location of knots sufficiently over the range of t and with large enough K, the class of functions $S(t)$ is very large and can approximate most smooth functions (Eilers and Marx, 1996; De Boor, 2001). It is recommended to use the number of spline knots (K) as the minimum of 40 and the number of unique t's divided by 4 (Ruppert, 2002); we also use this criterion in our paper. We refer to Ruppert *et al.* (2003) for more details of P-spline models.

To capture the spatial random effects η_i, a CAR model is employed. A variety of CAR models may be used by taking a collection of mutually compatible conditional distributions $p(\eta_i|\eta_{-i}), i = 1, ..., I$, where $\eta_{-i} = \{\eta_j : j \neq i, j \in \partial_i\}$ and ∂_i refers a set of neighbors for the ith region (Besag *et al.*, 1991). We consider the following general model

for the spatial effects η_i,

$$\boldsymbol{\eta} = (\eta_1, ..., \eta_I)' \sim N(0, \Sigma_\eta), \tag{3}$$

$$\Sigma_\eta = \sigma_\eta^2 (N_I - \lambda_\eta D)^{-1} P,$$

where P is a $I \times I$ diagonal matrix with elements $P_{ii} = 1/e_i$; D is a $I \times I$ matrix with elements $D_{ij} = 1/e_i$ if region i and j are adjacent and $D_{ij} = 0$ otherwise, where e_i is the number of regions which are adjacent to region i; σ_η^2 is the spatial dispersion parameter; λ_η measures the conditional spatial dependence, $\lambda_{min} \leq \lambda_\eta \leq \lambda_{max}$, where $\lambda_{\min}^{-1}$ and λ_{max}^{-1} are the smallest and largest eigenvalues of $P^{-1/2} D P^{1/2}$; and N_I is an identity matrix of dimension I (Cressie and Chand, 1989; Stern and Cressie, 1999). One may define the interaction effect of space and time, θ_{it}, as $\delta_i t$ or $S_i(t)$ or simply i.i.d. Normal distribution, depending on the nature of dataset (Bernardinelli *et al.*, 1995; MacNab and Dean, 2001; Silva *et al.*, 2008; Torabi and Rosychuk, 2012). Note that δ_i is the coefficient of the linear temporal effect related to the ith region, and $S_i(t)$ is a temporal P-spline for specific region i.

3. LIKELIHOOD-BASED INFERENCE

Let $\boldsymbol{y} = (y_{11}, ..., y_{1T}, ..., y_{I1}, ..., y_{IT})'$ be the observed data vector and assume that the elements of $\boldsymbol{y}$ are conditionally independent given the random effects , $\boldsymbol{V} = (\eta_1, ..., \eta_I, \gamma_1, ..., \gamma_K, \theta_{11}, ..., \theta_{IT})'$, and drawn from a binomial distribution with parameters $\boldsymbol{\alpha}_1 = (m, \beta_1, ..., \beta_p)$. It is also assumed that distribution for $\boldsymbol{V}$ depends on parameters $\boldsymbol{\alpha}_2$ which includes $\lambda_\eta, \sigma_\eta^2, \sigma_{sp}^2$ and related parameter(s) from θ_{it}. The goal of the analysis is to estimate the model parameters $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)'$, and provide prediction (and prediction interval) of the odds over space and time as a function of $\boldsymbol{V}$.

As the DC approach uses Bayesian tools, we first review standard Bayesian approach

for inference in hierarchical models. We denote $L(\boldsymbol{\alpha};\boldsymbol{y})$ as likelihood of $\boldsymbol{\alpha}$ given $\boldsymbol{y}$ as

$$L(\boldsymbol{\alpha};\boldsymbol{y}) = \int f(\boldsymbol{y}|\boldsymbol{V}=\boldsymbol{v},\boldsymbol{\alpha_1})g(\boldsymbol{V}=\boldsymbol{v}|\boldsymbol{\alpha_2})d\boldsymbol{v},$$

where $f(\cdot)$ and $g(\cdot)$ denote binomial and multivariate Normal distributions, respectively. Let $\pi(\boldsymbol{\alpha})$ be prior distribution of model parameters $\boldsymbol{\alpha}$. One can then write the posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ as

$$\pi(\boldsymbol{\alpha}|\boldsymbol{y}) = \frac{L(\boldsymbol{\alpha};\boldsymbol{y})\pi(\boldsymbol{\alpha})}{C(\boldsymbol{y})}, \tag{4}$$

where $C(\boldsymbol{y}) = \int L(\boldsymbol{\alpha};\boldsymbol{y})\pi(\boldsymbol{\alpha})d\boldsymbol{\alpha}$ is the normalizing constant. One can use Markov chain Monte Carlo (MCMC) algorithms to generate random variates from the posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ without computing the integrals in the numerator or the denominator of (4)(Gilks *et al.*, 1996; Spiegelhalter *et al.*, 2004).

In the DC method, we use the Bayesian computational approach for frequentist purposes. In particular, the observations $\boldsymbol{y}$ is repeated independently by J different individuals and all these individuals obtain exactly the same set of observations $\boldsymbol{y}$ called $\boldsymbol{y}^{(J)} = (\boldsymbol{y}, \boldsymbol{y}, ..., \boldsymbol{y})$. The likelihood function for the combination of the data from these J independent experiments is then given by $\{L(\boldsymbol{\alpha};\boldsymbol{y})\}^J$. Note that this likelihood function has two important features: a) the location of the maximum of this function is exactly equal to the location of the maximum of $L(\boldsymbol{\alpha};\boldsymbol{y})$ and b) the Fisher information matrix based on this likelihood is J times the Fisher information matrix based on $L(\boldsymbol{\alpha};\boldsymbol{y})$, (Lele *et al.*, 2010). Denote $\hat{\boldsymbol{\alpha}}$ as MLE and $I(\hat{\boldsymbol{\alpha}})$ as corresponding Fisher information matrix based on $L(\boldsymbol{\alpha};\boldsymbol{y})$. It is assumed that the parameters are identifiable and that there is a unique mode (but possibly multiple smaller peaks) to the likelihood function. The posterior distribution of $\boldsymbol{\alpha}$ conditional on the data $\boldsymbol{y}^{(J)}$ is then given by

$$\pi_L(\boldsymbol{\alpha}|\boldsymbol{y}^{(J)}) = \frac{\{L(\boldsymbol{\alpha};\boldsymbol{y})\}^J\pi(\boldsymbol{\alpha})}{C(\boldsymbol{y}^{(J)})}, \tag{5}$$

where $C(\boldsymbol{y}^{(J)}) = \int \{L(\boldsymbol{\alpha}; \boldsymbol{y})\}^J \pi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}$ is the normalizing constant. The following Theorem 1 shows that how one can use the likelihood of J copies of the original data to make an inference based on the MLE.

Theorem 1. Consider the general models (1)-(3). Under some regularity conditions, the distribution in (5) converges to a multivariate Normal distribution with mean equal to the MLE of the model parameters and variance-covariance matrix equal to $1/J$ times the inverse of the Fisher information matrix for the MLE.

Proof: The proof follows along the lines of Walker (1969) and Lele *et al.* (2010), and is omitted for simplicity.

Hence, the sample mean vector of the generated random numbers from (5) provides the MLE of the model parameters and J times their sample variance-covariance matrix is an estimate of the asymptotic variance-covariance matrix for the MLE $\hat{\boldsymbol{\alpha}}$. There are also various checks to determine the adequate number of clones J (Lele *et al.*, 2010). For instance, one may plot the largest eigenvalue of the posterior variance as a function of the number of clones J to determine if the posterior distribution has become nearly degenerate. As another criterion, it is approximately correct that as we increase the number of clones, we have

$$(\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}})' \boldsymbol{W}^{-1} (\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}}) \sim \chi^2_d, \tag{6}$$

where $\boldsymbol{W}$ is the variance of the posterior distribution and d is the dimension of $\boldsymbol{\alpha}$. One may also compute the following two statistics: (a)$\zeta = \frac{1}{B} \sum_{q=1}^{B} (O_q - Q_q)^2$, where O_q and Q_q are observed and estimated quantiles for χ^2_d random variable, and (b) $\tilde{r}^2 = 1 - \rho^2$, where ρ is the correlation between (O, Q). If these two statistics are close to zero, it indicates that the approximation (6) is feasible.

In this paper, the independent Normal distribution prior is assigned for fixed effects

with zero mean and variance 10^6 and gamma distribution for the inverse of variance components with shape and scale parameter 0.001. Because the DC is invariant to the priors, one may use different priors. To monitor the convergence of the model parameters, we use several diagnostic methods implemented in the Bayesian output analysis (BOA) program (Smith, 2007), a freely available package created for R (R Development Core Team, 2012). We also use three diagnostic methods implemented in the R package dclone (Sólymos, 2010), which were described in this section, to monitor the convergence of the model parameters in terms of number of clones J (Lele *et al.*, 2010).

3.1. Prediction of odds

Prediction of odds (or random effects), particularly from the frequentist viewpoint, is usually problematic. If the parameters $\boldsymbol{\alpha}$ are known, one can use the conditional distribution of $\boldsymbol{O} = (O_{11}, ..., O_{1T}, ..., O_{I1}, ..., O_{IT})'$, the latent variables, given the observed data to predict the odds where $O_{it} = p_{it}/(1 - p_{it})$ is odds at region i and time t. That is, one can use $\pi(\boldsymbol{O}|\boldsymbol{y}, \boldsymbol{\alpha}^*)$ where $\boldsymbol{\alpha}^*$ is the true value of the parameter. A naive approach, when $\boldsymbol{\alpha}$ is estimated using the data, is to use $\pi(\boldsymbol{O}|\boldsymbol{y}, \hat{\boldsymbol{\alpha}})$. This approach, however, does not take into account the variability introduced by the model parameters estimate. An approach that has been suggested in the literature (e.g., Hamilton, 1986; Lele *et al.*, 2010) is to use the following density:

$$\pi(\boldsymbol{O}|\boldsymbol{y}) = \frac{\int f(\boldsymbol{y}|\boldsymbol{O}) g(\boldsymbol{O}|\boldsymbol{\alpha}) \phi(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}, I^{-1}(\hat{\boldsymbol{\alpha}})) d\boldsymbol{\alpha}}{C(\boldsymbol{y})}, \tag{7}$$

where $f(\cdot)$ and $g(\cdot)$ are binomial and multivariate Normal distributions, respectively, and $\phi(., \mu, \Sigma)$ denotes a multivariate Normal density with mean μ and variance Σ, which are equal to the MLE and inverse of the Fisher information matrix here. In this paper, we obtain prediction and prediction interval of the $\boldsymbol{O}$ using the density in equation (7) along with MCMC sampling; noting that one can also use the same approach to predict, for

example, $\exp(\eta_i + \theta_{it}), (i = 1, ..., I; t = 1, ..., T)$.

4. APPLICATION

4.1. Data analysis

To evaluate the performance of the proposed approach, a yearly dataset of childhood (age $\leq$ 20 years) asthma physician visits in the Canadian province of Manitoba during the 2000-2010 fiscal years is used. The population of Manitoba was stable during the study period from 1.15 million in 2000 to 1.20 million in 2010, with an average population of children of around 335 000. The province consisted of eleven Regional Health Authorities that were responsible for the delivery of health care services. These eleven regions were further sub-divided into 56 Regional Health Authority Districts (RHADs) and these RHADs are the geographic units used in our model and all data were linked to these geographic boundaries. For simplicity, we call these RHADs 1,2,...,56. The number of children asthma physician visits totaled 736 106 over the study period with mean and median number of yearly visits per region of 1 314 and 450 (range 59 to 44 090), respectively. The regional child population sizes varied from 290 to 175 300, with mean and median numbers of 5 998 and 2 488, respectively. We first present the provincial rate of children asthma physician visits over time. Figure 1 shows that the rates slightly decreased over time from 0.26 in 2000 to 0.22 in 2010.

"Figure 1 around here"

We then fit the model (1) to the dataset of children asthma physician visits, with cubic P-spline and two knots, using the DC method. The adequacy of spatio-temporal model (1) for this dataset was checked by comparing the model (1) with spatial or temporal model. Following Ponciano *et al.* (2009), we use model selection based on the information criteria

for this purpose. In particular, to compare two models, one can write $AIC_1 - AIC_2 = -2ln(\frac{\hat{L}_1}{\hat{L}_2}) + 2(d_1 - d_2)$, where AIC stands for Akaike Information Criteria, L_i is the likelihood for model $i(i = 1, 2), d_1$ and d_2 are the number of parameters estimated under models 1 and 2, respectively (Burnham and Anderson, 2002; Ponciano *et al.*, 2009). We define the spatial model (only m and η_i in (1)) as model 1, the temporal model (only m and $S(t)$ in (1)) as model 2, and the spatio-temporal model (1) as model 3. We then have the following results for our dataset: $AIC_1 - AIC_3 = 2.18, AIC_2 - AIC_3 = 3.48$. As indicated in Ponciano *et al.* (2009), because these differences are positive we can conclude that the spatio-temporal model (1) provides a better description of the data than models with only spatial or only temporal effects. The AIC differences greater than 2 are generally thought to be significant, and differences greater than 3 very significant (Burnham and Anderson, 2002; Taper, 2004; Ponciano *et al.*, 2009).

It is also well-known that the variance components can be heavily biased if there is a large number of fixed effects in the mixed model. Hence, we offer the bias correction of variance components using jackknife ($\hat{\boldsymbol{\alpha}}_{2JK}$) as $\hat{\boldsymbol{\alpha}}_{2JK} = Z\hat{\boldsymbol{\alpha}}_2 - (Z-1)\bar{\hat{\boldsymbol{\alpha}}}_{2JK}$, where $\bar{\hat{\boldsymbol{\alpha}}}_{2JK} = Z^{-1}\sum_{z=1}^{Z}\hat{\boldsymbol{\alpha}}_{2,-z}$, $Z = IT$, and $\hat{\boldsymbol{\alpha}}_{2,-z}$ is the estimates of $\boldsymbol{\alpha}_2$ using DC with deleting the information of zth cell (Jiang, 2007). The PQL approach is also studied in our data analysis because this approximation method has been widely used in the context of binomial spatio-temporal model as a frequentist approach. Table 1 reports the model parameters estimates and corresponding standard errors for DC, Hierarchical Bayes (HB), and PQL approaches; noting that we used $\theta_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\theta^2)$ which was found useful in our exploration of the data. It seems that the standard errors for some model parameters in DC method are smaller than the other two methods, noting that the PQL based estimate for the conditional spatial dependence (λ_η) is too high compared to the corresponding values in the DC and HB approaches. Also, the estimates of spatial dispersion (σ_η^2) and spline dispersion (σ_{sp}^2) in the PQL method are not comparable with the corresponding

estimates in the DC and HB approaches. The standard error of the spline dispersion ($\hat{\sigma}^2_{sp}$) for the HB is also unusually high compared to the corresponding values in the DC and PQL methods. We will further evaluate the performance of these methods in section 4.2 through a simulation study.

"Table 1 around here"

The necessary requirements were also checked in terms of the convergence of model parameters and the number of clones. For instance, Figure 2 shows that the scaled variances decrease at a $1/J$ rate and reach a lower bound (say < 0.05) which shows that the DC approach has converged. For this application, the number of clones was $J = 40$ to obtain MLE, and 50,000 iterations for the convergence of the model parameters. We also compared the computational costs for the DC and HB methods. For example, the computing time for the HB method was about 48 s, whereas for the DC method with $J = 40$ was about 1516 s. It is true that the DC computation is more intensive than the HB approach, however, it is the cost of getting a likelihood-based method (MLE) and also providing prediction (and prediction interval) of random effects which has been a challenge from the frequentist perspective.

"Figure 2 around here"

For a diagnostic analysis, we calculated the *deviance residual* (McCullagh and Nelder, 1989) as

$$d_{it} = sgn(y_{it} - n_{it}p_{it})\Big[2\big\{y_{it}\log(\frac{y_{it}}{n_{it}p_{it}}) + (n_{it} - y_{it})\log(\frac{n_{it} - y_{it}}{n_{it} - n_{it}p_{it}})\big\}\Big]^{1/2},$$

where

$$sgn(z) = \begin{cases} 1 & z > 0 \\ 0 & z = 0 \\ -1 & z < 0 \end{cases} .$$

Figure 3 gives the residuals versus predicted odds diagnostic plot based on the DC approach. It is clear from Figure 3 that there is no serious lack of fit in our model.

"Figure 3 around here"

Figure 4 reveals estimates of the overall odds over time, $\exp(m+\sum_{j=1}^{3}\beta_j t^j+\sum_{k=1}^{2}\gamma_k(t-\kappa_k)_+^3)$, using the DC approach. It shows that over the study period, departures from linearity are observed with an overall decrease in asthma visit odds. The change in the estimated probability of asthma shows a sharp decrease particularly from 2000 to 2001 and also from 2004 to 2008, and an increase from 2008 to 2010; noting that the RHADs 29, 33, and 38 had high influence to increase the overall odds from 2008 to 2010.

"Figure 4 around here"

One of the main features of the DC approach is the ability to predict the random effects from the frequentist perspective. To have a better understanding of the spatial variation, we obtained the spatial distribution of the area effects over time as $\exp(\eta_i+\theta_{it})$ using the DC approach; this provides identification of those RHADs experiencing large differing temporal effects with respect to the overall spline. Figure 5 presents maps of the estimated spatial effects based on the fitted model, where the regional risk factor of asthma visits corresponds to some selected years (2001, 2006, and 2010). The overall spatial pattern suggests that some RHADs in the south part of the province had relatively high odds of children asthma visits. Generally, the spatial pattern does not change much over time; although some RHADs had higher odds estimate in 2010 compared to 2001

(e.g., RHADs 37, 38, and 40) and some RHADs had lower odds estimate in 2010 compared to 2001 (e.g., RHADs 14, 18, and 20).

"Figure 5 around here"

We also provided the regional odds estimate of asthma visit obtained from fitting the spatio-temporal mixed model given by $\exp(\eta_i + \theta_{it})$. Figure 6 plots the fitted odds of asthma visits with corresponding 95% prediction intervals, for example, for RHADs 7, 21, 24, and 39. The standardized crude odds estimates are $y_{it}/r_t(n_{it} - y_{it})$, where $r_t = \sum_i y_{it} / \sum_i (n_{it} - y_{it})$, and are also plotted in Figure 6. As expected in Figure 6, our odds estimates of asthma visits provide smoothed estimates while crude odds are unstable over time.

"Figure 6 around here"

4.2. Simulation study

We also conducted a simulation study to evaluate the performance of ML estimates, via DC approach, and compare it with the PQL and HB approaches using a scenario similar to our children asthma dataset. More specifically, data are generated from the model (1) with the parameters $(m, \beta_1, \beta_2, \beta_3, \sigma^2_\eta, \lambda_\eta, \sigma^2_{sp}, \sigma^2_\theta)$ listed in Table 2. The neighborhood structure and the population sizes are exactly as for the asthma dataset. Estimates are obtained for the DC, PQL, and HB methods using 1000 datasets generated from the mixed binomial model (1).

The mean values of the model parameters estimates, the standard deviation of the estimated parameters and mean values of the estimated standard errors are presented in Table 2. It seems that the estimates of model parameters in DC approach are reasonably unbiased, and their standard errors are also estimated well with comparing the

estimated standard errors with the corresponding simulated values, noting that the variance components were corrected using the jackknife method. However, the estimate of model parameters is heavily biased for the PQL method, and the standard errors of model parameters estimate are not comparable with the corresponding simulated values. The model parameters estimate of the HB method is also biased, although the standard errors of model parameters estimate are generally comparable with the corresponding simulated values except the spline dispersion. Overall, it seems that the DC approach, which yields to MLE, provides good point estimates and standard errors for this data analysis.

"Table 2 around here"

5. Conclusion

For fitting complex models in the context of spatio-temporal, the Poisson approximate to binomial distribution may be appropriate for rare disease. However, in studies of health system services, using Poisson baseline structure may not be appropriate as events are not typically rare. Consequently, we need to use spatio-temporal modeling of odds and study risk factors in the mixed binomial model setting.

The frequentist analysis of the mixed binomial model is computationally difficult. It has been shown that, for example, penalized quasi-likelihood performs poorly for the binomial setting. Analysis based on data cloning (DC) has overcome the computational difficulties of the maximum likelihood (ML) method.

Using the DC, we have proposed a frequentist approach for spatio-temporal analysis of the binomial setting that focused on the mapping of area level odds of disease. The model accommodated a CAR model for the spatial random effects and penalized spline smoothing over the temporal effects. The model can be also easily extended to include some covariates directly, which may be required for some applications. We used the DC

method which yields to MLE to estimate the model parameters with correcting variance components using the jackknife method, and also to provide prediction (and prediction interval) of the odds over space and time. As another advantage of the DC approach is that the non-estimable parameters are flagged automatically (Lele *et al.*, 2010). In particular, we considered the model (1) with incorporating an i.i.d. regional random variable with Normal distribution and then fitted it to the dataset in Section 4.1 and observed that with increasing number of clones, the variances of posterior distribution also increase.

Based on the model estimates, it was suggested that the odds of children asthma physician visits were decreasing over the study period. However, some RHADs in the south part of province had slightly higher odds of asthma for children compared to the north. It may be due to limited health services such as general physicians in the north part of the province. These findings may represent real changes or different distributions of important covariates that are unmeasured and unadjusted for in our modeling. Further investigation is needed to explore these findings.

In our childhood asthma physician visits, we assumed that the physician visits were independent from each other, however, we had some cases who visited physicians multiple times within a given year. In fact, the mean and median number of yearly readmissions per region were 1.9 and 1.1 (range 1 to 4.4), respectively. We have planned to study this kind of data more appropriately in a separate manuscript.

Our proposed DC approach is very general in the context of mixed binomial model setting. In this paper, we used CAR and penalized spline models for spatial and temporal effects, respectively; however, one may consider other variants of spatial and temporal effects; for example, CAR model (MacNab and Dean, 2001) for spatial random effects and AR model (Torabi and Rosychuk, 2010) for temporal random effects.

Acknowledgements

I would like to thank two referees and an associate editor for constructive comments and suggestions, which led to the improvement of the paper. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC). Disclaimer: The interpretations, conclusions and opinions expressed in this paper are those of the author and do not necessarily reflect the position of Manitoba Health. This study is based in part on data provided by Manitoba Health through the Manitoba Centre for Health Policy. The interpretation and conclusions contained herein are those of the researcher and do not necessarily represent the views of the government of Manitoba.

REFERENCES

Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. 1995. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* **14**:2433–2443.

Besag J, York J, Mollièe A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**:1–21.

Breslow NE, Clayton DG. 1993. Approximate inference in generalized linear mixed models. *Journal of American Statistical Association* **88**:9–25.

Burnham, KP, Anderson DR. 2002. *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*, Second edition, New York: Springer-Verlag.

Cressie NA, Chan NH. 1989. Spatial modeling of regional variables. *Journal of the American Statistical Association* **84**: 393–401.

De Boor C. 2001. *A Practical Guide to Splines*, New York: Springer-Verlag.

Eilers PHC, Marx BD. 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* **11**:89–121.

Gilks WR, Richardson S, Spiegelhalter DJ. (eds.). 1996. *Markov Chain Monte Carlo in Practice*, New York: Springer - Verlag.

Hamilton JD. 1986. A standard error for the estimated state vector of a state-space model. *Journal of Econometrics* **33**:387–397.

Jiang J. 2007. *Linear And Generalized Linear Mixed Models And Their Applications*, New York: Springer-Verlag.

Knorr-Held L. 2000. Bayesian modeling of inseparable space-time variation in disease risk. *Statistics in Medicine* **19**:2555–2567.

Knorr-Held L, Besag J. 1998. Modelling risk from a disease in time and space. *Statistics in Medicine* **17**:2045–2060.

Lele SR, Dennis B, Lutscher F. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters* **10**:551–563.

Lele SR, Nadeem K, Schmuland B. 2010. Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association* **105**:1617–1625.

Lin X, Breslow NE. 1996. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* **91**:1007–1016.

MacNab YC. 2003. Hierarchical Bayesian spatial modelling of small-area rates of non-rare disease. *Statistics in Medicine* **22**:1761–1773.

MacNab YC, Dean CB. 2001. Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics* **57**:949–956.

Martinez-Beneito MA, Lopez-Quilez A, Botella-Rocamora P. 2008. An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine* **27**:2874–2889.

McCullagh P, Nelder JA. 1989. *Generalized Linear Models*, 2nd ed., London: Chapman and Hall.

Ponciano JM, Taper ML, Dennis B, Lele SR. 2009. Hierarchical models in ecology:

Confidence intervals, hypothesis testing, and model selection using data cloning. *Ecology* **90**:356–362.

R Development Core Team. 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria ISBN 3-900051-07-0,http://www.R-project.org, [accessed on January 17, 2013].

Ruppert D. 2002. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11**:735–757.

Ruppert D, Wand MP, Carroll RJ. 2003. *Semiparametric Regression*. Cambridge: Cambridge University Press.

Silva GL, Dean CB, Niyonsenga T, Vanasse A. 2008. Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines. *Statistics in Medicine* **27**:2381–2401.

Smith BJ. 2007. *BOA User Manual (version 1.1.7)*, Department of Biostatistics, College of Public Health, University of Iowa, Ames.

Sólymos P. 2010. dclone: data cloning in R. *The R Journal* **2**:29–37.

Spiegelhalter D, Thomas A, Best N, Lunn D. 2004. *WinBUGS version 1.4 User Manual*. MRC Biostatistics unit, Institute of Public Health,

Stern HS, Cressie NA. 1999. Inference for extremes in disease mapping. In Disease mapping and risk assessment for public health. A. Lawson, A. Biggeri, D. Bohning, E. Lesaffre, J-F. Viel and R. Bertollini (eds.), Chichester: Wiley, 63–84.

Taper ML. 2004. Model identification from many candidates. In The nature of scientific evidence: statistical, philosophical and empirical considerations. ML. Taper, SR Lele (eds.), Chicago: The University of Chicago Press, 448–524.

Torabi M. 2012a. Hierarchical Bayes estimation of spatial statistics for rates. *Journal of Statistical Planning and Inference* **142**:358–365.

Torabi M. 2012b. Spatial modeling using frequentist approach for disease mapping.

Journal of Applied Statistics **39**:2431–2439.

Torabi M. 2013. Spatio-temporal modeling for disease mapping using CAR and B-spline smoothing. *Environmetrics* **24**:180–188.

Torabi M, Rosychuk RJ. 2010. Spatio-temporal modelling of disease mapping of rates. *The Canadian Journal of Statistics* **38**:698–715.

Torabi M, Rosychuk RJ. 2011. Spatio-temporal modeling using spline for disease mapping: Analysis of childhood cancer trends. *Journal of Applied Statistics* **38**:1769–1781.

Torabi M, Rosychuk RJ. 2012. Hierarchical Bayesian spatiotemporal analysis of childhood cancer trends. *Journal of Geographical Analysis* **44**:109–120.

Walker AM. 1969. On the asymptotic behavior of posterior distributions. *Journal of Royal Statistical Society, Series B* **31**:80–88.

Waller LA, Carlin BP, Xia H, Gelfand AE. 1997. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**:607–617.

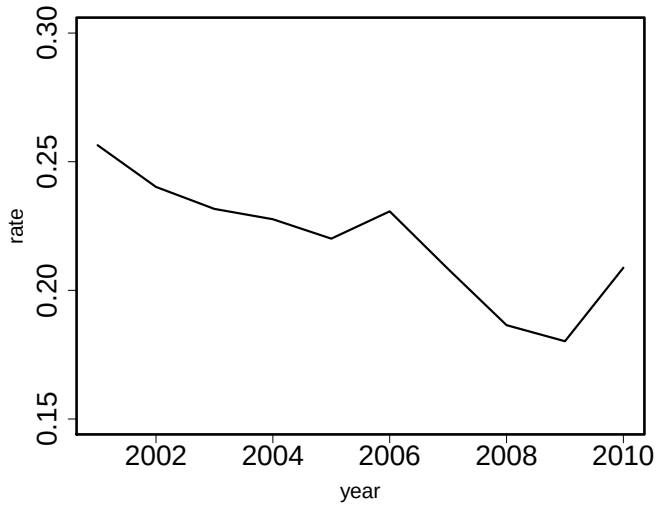


Fig.1. Provincial childhood asthma physician visit rates over time

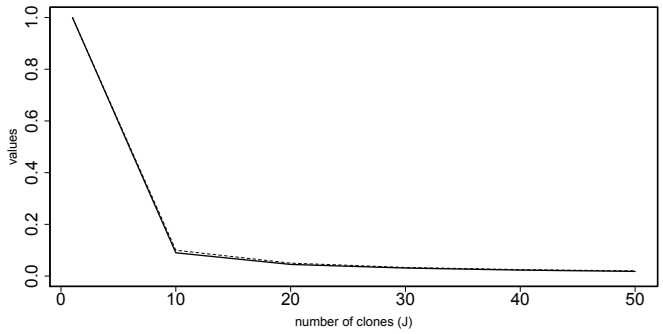


Fig.2. Data cloning convergence diagnostics for asthma physician visit study dataset. The standardized maximum eigenvalues (solid line) converge to zero at the expected rate $1/J$ (dashed line)

Table 1
Parameter estimates (and standard errors), spatio-temporal mixed binomial model for maximum likelihood estimation via data cloning (DC-MLE), penalized quasi-likelihood (PQL), and hierarchical Bayesian (HB) methods, childhood asthma physician visits in the province of Manitoba, Canada, during 2000–2010

Parameter	Estimate (standard error)		
	DC-MLE	PQL	HB
m	-1.110(0.812)	-1.101(0.821)	-1.123(0.818)
β_1	-0.114(0.064)	-0.124(0.080)	-0.118(0.058)
β_2	-0.183(0.063)	-0.195(0.081)	-0.188(0.060)
β_3	-0.131(0.067)	-0.113(0.076)	-0.137(0.069)
σ^2_η	0.253(0.091)	0.069(0.016)	0.499(0.106)
λ_η	0.814(0.001)	4.817(0.518)	0.997(0.003)
σ^2_{sp}	0.194(0.168)	0.0001(0.005)	1.686(17.707)
σ^2_θ	0.041(0.003)	0.118(0.007)	0.042(0.003)

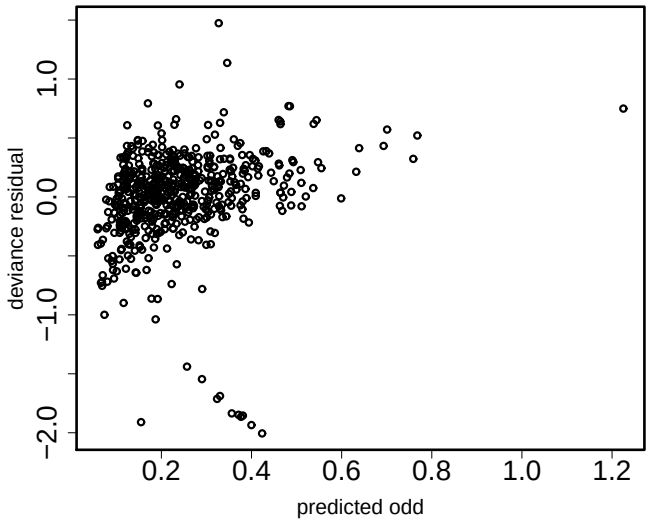


Fig.3. The deviance residuals versus predicted odds diagnostic plot of childhood asthma physician visits based on the data cloning approach

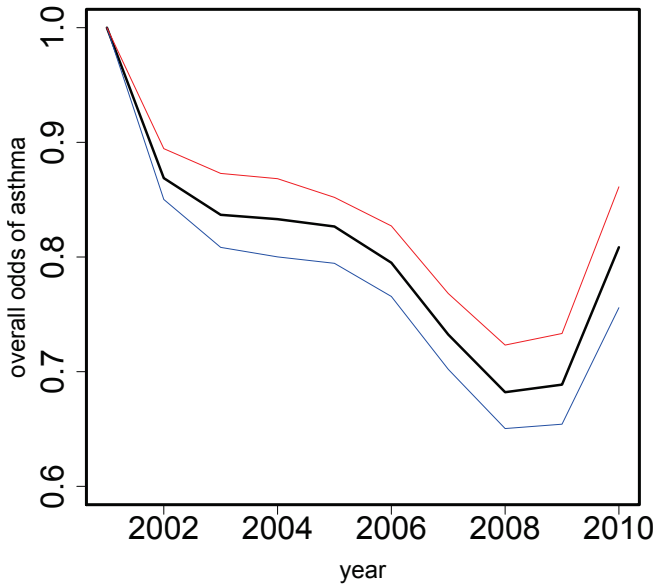
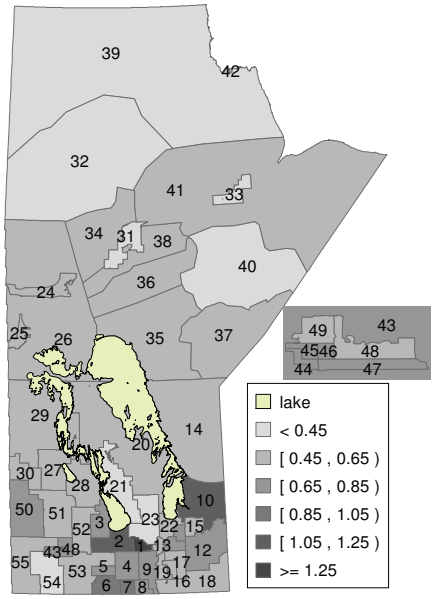


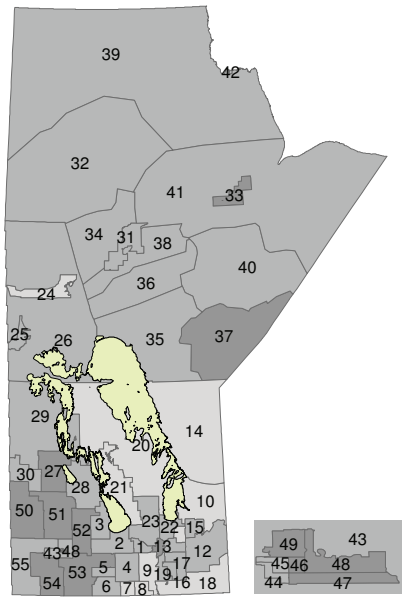
Fig.4. Overall odds of asthma physician visits by children and corresponding 95% prediction bands based on the data cloning approach

Table 2. Mean values of the model parameters estimates, the standard deviation of the estimated parameters and mean values of the estimated standard errors of maximum likelihood estimation via data cloning (DC-MLE), penalized quasi-likelihood (PQL), and hierarchical Bayes (HB) methods based on 1,000 simulated datasets

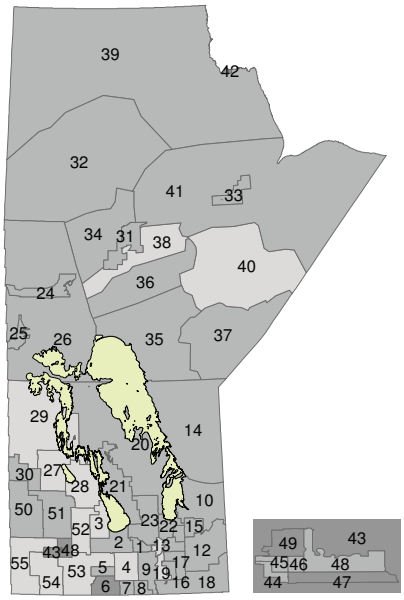
Parameter	DC-MLE			PQL			HB		
	Mean	Standard error		Mean	Standard error		Mean	Standard error	
		DC-MLE	Simulated		PQL	Simulated		HB	Simulated
m =-1.10	-1.101	0.815	0.815	-1.096	0.820	0.785	-1.104	0.818	0.816
β_1 =-0.12	-0.120	0.086	0.086	-0.122	0.085	$4e^{-6}$	-0.121	0.087	0.087
β_2 =-0.19	-0.193	0.082	0.083	-0.190	0.082	$4e^{-6}$	-0.194	0.084	0.084
β_3 =-0.14	-0.140	0.063	0.065	-0.143	0.063	$3e^{-6}$	-0.140	0.064	0.066
σ^2_η =0.50	0.528	0.164	0.167	-0.001	0.114	$1e^{-7}$	0.864	0.179	0.174
λ_η =0.40	0.416	0.258	0.270	0.400	0.037	$1e^{-7}$	0.466	0.263	0.254
σ^2_{sp} =0.10	0.096	0.030	0.046	0.0001	0.066	$1e^{-8}$	0.133	0.168	0.069
σ^2_θ =0.04	0.040	0.003	0.003	-0.001	0.003	$1e^{-7}$	0.041	0.003	0.003



2006



2010



2001

Fig.5. Maps of the odds estimate of asthma visits for the spatial effects for some selected years (2001, 2006, and 2010) based on the data cloning approach; Manitoba childhood asthma data (2000-2010). Major urban centre (Winnipeg region) is provided as inset

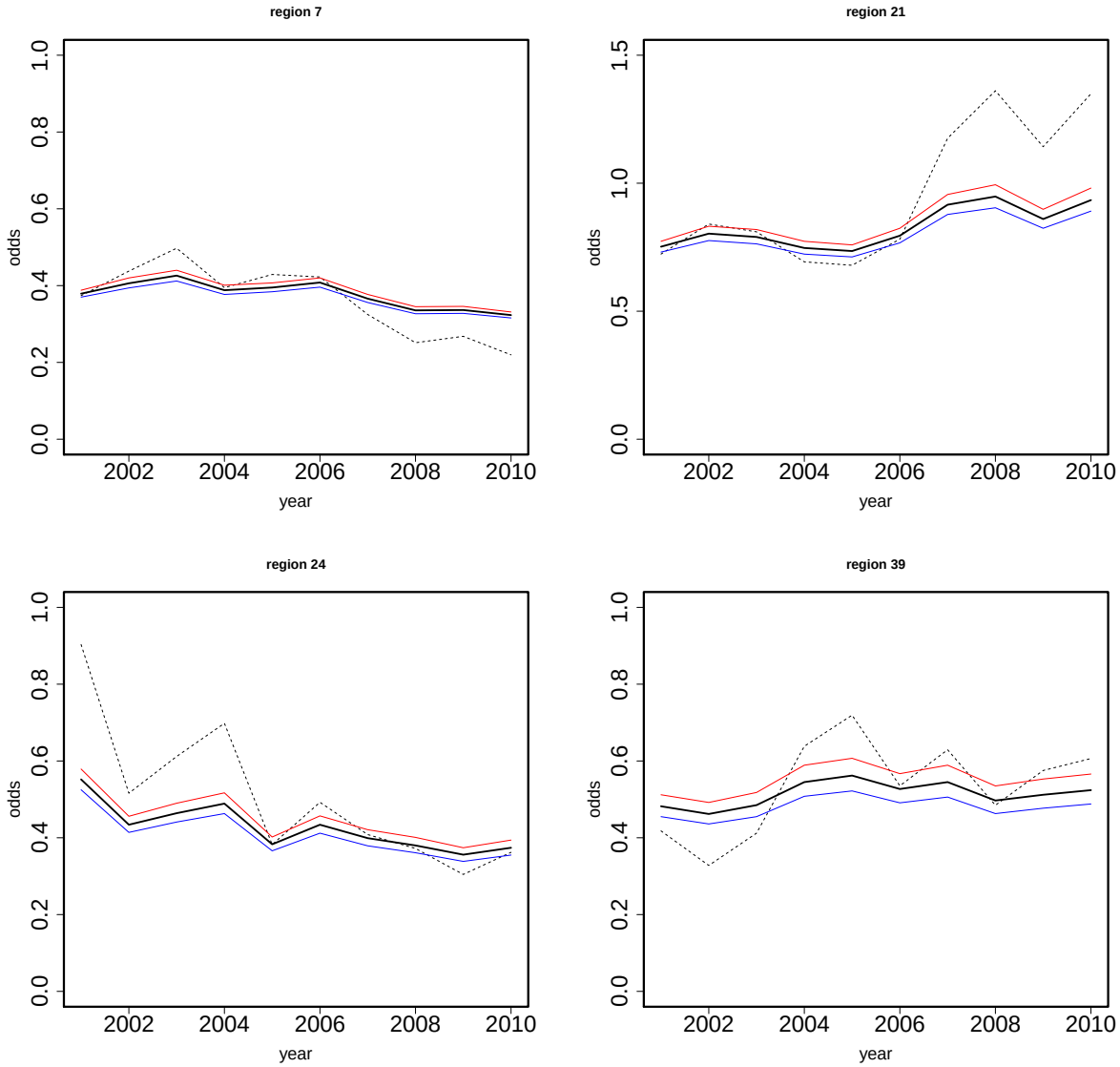


Fig.6. Plots of the standardized crude odds and estimated odds of asthma visits for the selected RHADs 7, 21, 24, and 39 during 2000–2010. The solid black line represents fitted odds with the blue and red lines as 95% prediction bands; the dashed line represents the crude odds