The Annals of Applied Statistics
2016, Vol. 10, No. 3, 1427–1446
DOI: 10.1214/16-AOAS941

QUANTIFYING THE SPATIAL INEQUALITY AND TEMPORAL TRENDS IN MATERNAL SMOKING RATES IN GLASGOW[1]

BY DUNCAN LEE AND ANDREW LAWSON

University of Glasgow and Medical University of South Carolina

Maternal smoking is well known to adversely affect birth outcomes, and there is considerable spatial variation in the rates of maternal smoking in the city of Glasgow, Scotland. This spatial variation is a partial driver of health inequalities between rich and poor communities, and it is of interest to determine the extent to which these inequalities have changed over time. Therefore in this paper we develop a Bayesian hierarchical model for estimating the spatio-temporal pattern in smoking incidence across Glasgow between 2000 and 2013, which can identify the changing geographical extent of clusters of areas exhibiting elevated maternal smoking incidences that partially drive health inequalities. Additionally, we provide freely available software via the R package `CARBayesST` to allow others to implement the model we have developed. The study period includes the introduction of a ban on smoking in public places in 2006, and the results show an average decline of around 11% in maternal smoking rates over the study period.

1. Introduction. The detrimental effect of maternal smoking on birth outcomes is well known [see Wang et al. (2002) and Cnattingius (2004)], with epidemiological evidence linking it to increased rates of still birth and small-for-gestational-age babies. Tappin et al. (2010) estimate that over 20% of pregnant mothers in Scotland smoked in 2005, although this was far from uniform across the country with self-reported rates varying geographically between 3% and 53%. Also in Scotland, Gray et al. (2009) estimate that maternal smoking accounted for 38% of the spatial inequality in stillbirths and 31% of the inequality in infant deaths. The harmful effects of smoking on these and other health outcomes led to the introduction of The Smoking, Health and Social Care (Scotland) Act 2005, which banned smoking in any enclosed public space in Scotland from the 26 March 2006. The ban followed soon after the first national ban of this type in Ireland in March 2004, and now bans exist in many countries including Australia, Brazil, Canada and South Korea. Numerous research has found links between these bans and improved public health, with Mackay et al. (2012) reporting a significant association between the Scottish ban and small-for-gestational-age babies. However, Mackay et al. (2012) note that assessing the impact of a ban is nontrivial be-

Received July 2015; revised April 2016.
[1]Supported in part by Medical Research Council Grant MR/L022184/1.
Key words and phrases. Cluster detection, maternal smoking, spatial inequality, spatio-temporal modelling.

cause there was widespread advertising of the ban before its introduction, resulting in an anticipatory effect as people changed their behaviours before the deadline.

Therefore, this paper investigates the changing spatio-temporal dynamics of maternal smoking in Scotland between 2000 and 2013, which is an era that included the ban in March 2006. We focus on the city of Glasgow because it has a high smoking incidence, a large inequality in health between rich and poor, and one of the poorest health records in Europe [the "*Glasgow Effect*," Bauld et al. (2005), Gray and Leyland (2009), Gray et al. (2012)]. Specifically, we address the following questions: (i) what is the overall temporal trend in the maternal smoking incidence in Glasgow between 2000 and 2013; (ii) how has the magnitude of the spatial inequality in maternal smoking incidence changed between 2000 and 2013; (iii) where were the clusters of areas with high maternal smoking incidences in 2000 that partially drive these inequalities, and which of them have seen a reduction in incidences by 2013; and (iv) what impact does socioeconomic deprivation have on maternal smoking rates? Answering these questions provides key public policy information on the extent to which maternal smoking is driving health inequalities, and whether these inequalities have gotten wider or narrower over the 14 years considered in this study. The identification of clusters of high incidence areas also allows future health resources to be targeted appropriately at areas in greatest need of reducing maternal smoking levels.

A range of models have been developed for estimating spatio-temporal patterns in areal unit data [see Knorr-Held (2000) and Lawson (2009), Chapter 12], while scan statistics have been proposed for cluster detection [see Kulldorff et al. (2005)]. However, these approaches have fundamentally different goals, as the former estimates a smoothed spatio-temporal incidence surface, while the latter only identifies a small number of high incidence clusters. Charras-Garrido et al. (2013) propose a two-stage approach in a purely spatial setting for achieving both goals, which applies a clustering algorithm to the incidence surface estimated from a spatial smoothing model. However, identifying clusters (i.e., step changes in incidence between neighbouring areas) from a spatially smoothed surface is inherently problematic, and Anderson, Lee and Dean (2014) show this does not lead to good cluster recovery. Alternatively, Forbes et al. (2013), Gangnon and Clayton (2000), Green and Richardson (2002), Knorr-Held and Raßer (2000), Wakefield and Kim (2013) and Anderson, Lee and Dean (2014) propose integrated approaches in a purely spatial context. The identification of clusters of areas exhibiting elevated incidence compared to their geographical neighbours would seem to violate the common assumption of a single global level of spatial smoothness (autocorrelation), as some pairs of neighbouring areas will have similar values while those on the edge of a cluster will not. Choi et al. (2011), Lawson et al. (2012) and Li et al. (2012) have extended clustering-type models to the spatio-temporal domain, but only focus on detecting shared latent structures and unusual temporal trends, and an integrated modelling framework for spatio-temporal estimation and cluster detection is yet to be proposed.

Therefore, this paper has two key contributions. First, we fill the methodological gap described above by proposing a novel modelling approach for cluster detection and spatio-temporal estimation that can quantify the changing nature of health inequalities. The model is able to detect clusters dynamically so that cluster membership can evolve over time. Inference is based on Markov chain Monte Carlo (MCMC) simulation, and, unlike the majority of existing models in this field, we provide software for others to use via the R package `CARBayesST`. Second, we provide the first in-depth investigation into the changing dynamics of the spatial inequalities in maternal smoking incidence in Scotland, in an era that included government legislation aimed at reducing smoking levels. The data are presented in Section 2, while our methodological and software contribution is outlined in Section 3. Section 4 quantifies the performance of our methodology by simulation, while the results of the data analysis are presented in Section 5. Finally, Section 6 concludes the paper.

2. Data and existing spatio-temporal models.

2.1. *Data description.* The study region is the Greater Glasgow and Clyde Health board displayed in Figure 1, which contains the city of Glasgow and has a population of around 1 million. The region is split into $N = 271$ Intermediate Geographies (IG), which have an average population of around 4000 people. The study period is from 2000 until 2013, and the data we model are available from

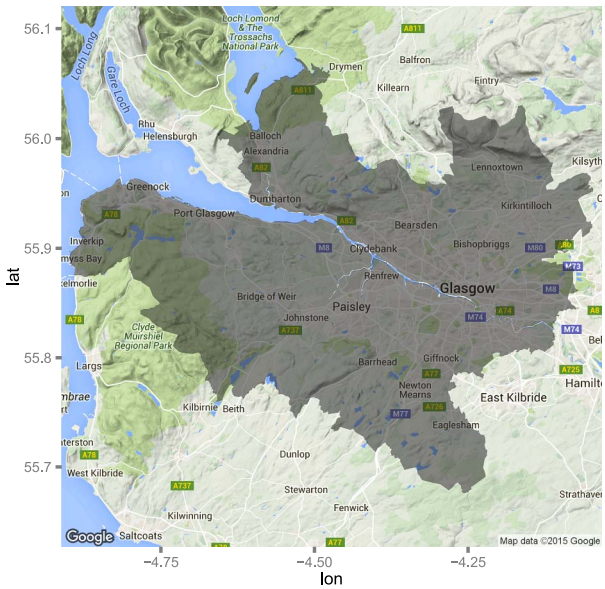

FIG. 1. *The study region of the Greater Glasgow and Clyde Health Board (shaded region) overlaid on a Google map.*

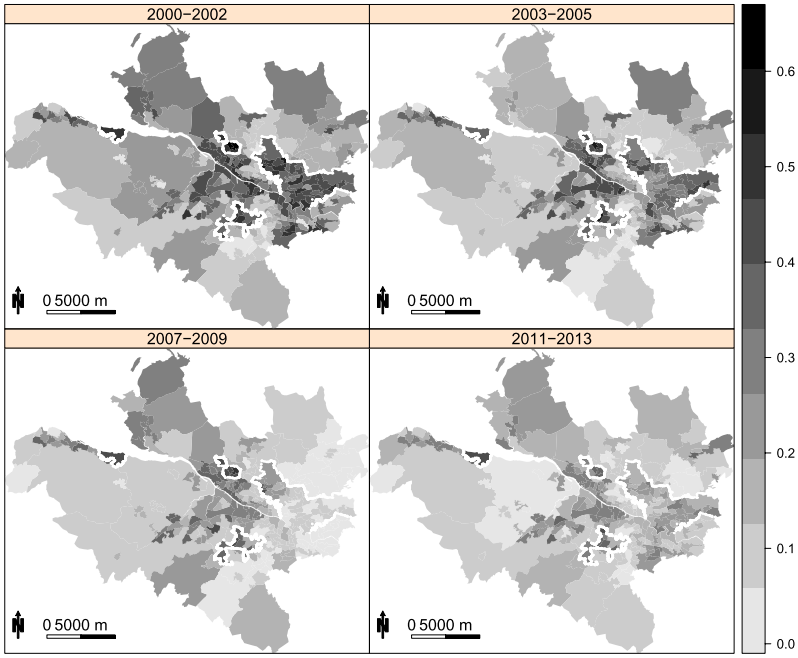

FIG. 2. *The spatial patterns in the raw smoking incidence for Greater Glasgow for* 4 *time periods* (*blacker means higher incidence*), *namely*, 2000–2002, 2003–2005, 2007–2009 *and* 2011–2013. *The white lines depict high incidence clusters in* 2000–2002.

Statistics.Gov.Scot (http://statistics.gov.scot). The data are self-reported current smoking status (smoker/nonsmoker), recorded at each pregnant woman's first antenatal visit to the hospital. The number of pregnant women (m_{it}) and the number of those that smoke (v_{it}) are available over the $i = 1, \ldots, N$ IGs as yearly three-year rolling totals centred between 2001 and 2012, resulting in $t = 1, \ldots, T = 12$ time periods. For example, (v_{i1}, m_{i1}) respectively denote the number of pregnant women that smoke and the number of pregnant women in the ith IG between 2000 and 2002. Finally, socioeconomic deprivation is likely to have a large effect on maternal smoking rates, and here we represent it by two proxy measures: (i) the proportion of the working age population who are in receipt of Job Seekers Allowance (JSA), a benefit paid to people who are unemployed in the UK; and (ii) the natural log of the median property price in each IG. A natural log transformation is applied to the latter, as exploratory analyses suggested it was a better predictor of maternal smoking.

Figure 2 summarises the spatio-temporal pattern in the three-yearly incidences $p_{it} = v_{it}/m_{it}$ by presenting the spatial pattern for the first (2000–2002), last (2011–2013) and two intervening (2003–2005 and 2007–2009) time periods. The figure shows a noticeable decline in smoking incidence amongst pregnant women overall during the study period, as the median incidence across Glasgow in 2000–

2002 was 29.3%, which dropped to a low point of 11.7% in 2007–2009, before increasing back to 15.5% in 2011–2013. The maps also show clear evidence of spatial clusters in 2000–2002 exhibiting higher smoking incidences than their neighbours, with examples including Port Glasgow in the west, Drumchapel in the north and Nitshill in the south (outlined by white dots in the figure). These high incidence clusters in 2000–2002 typically exhibit a much reduced incidence by the end of the study period (2011–2013).

2.2. *Spatio-temporal models for areal unit data.* There is an active development of spatio-temporal models for areal unit data, with examples including Knorr-Held (2000), Lawson (2009) (Chapter 12), Ugarte et al. (2012) and Rushworth, Lee and Mitchell (2014). The most common modelling approaches either utilise a main effects and interaction decomposition such as Knorr-Held (2000) or an autoregressive structure such as Rushworth, Lee and Mitchell (2014), and variants of both are compared against the model proposed here. Data augmentation techniques are used to estimate the unobserved yearly numbers of pregnant women that smoke, y_{it}, and pregnant women, n_{it}, from the available three-year rolling totals (v_{it}, m_{it}), and details are given in Section 3. The main effects and interaction model we consider here on the yearly scale is similar to that proposed by Knorr-Held (2000) and is given by

$$
\begin{aligned}
y_{it} &\sim \text{Binomial}(n_{it}, \theta_{it}),\\
\ln\left(\frac{\theta_{it}}{1-\theta_{it}}\right) &= \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + \phi_i + \delta_t + \gamma_{it},\\
\phi_i|\boldsymbol{\phi}_{-i}, \mathbf{W}, \tau_\phi^2, \rho_\phi &\sim \text{N}\left(\frac{\rho_\phi \sum_{j=1}^{N} w_{ij}\phi_j}{\rho_\phi \sum_{j=1}^{N} w_{ij} + 1 - \rho_\phi}, \frac{\tau_\phi^2}{\rho_\phi \sum_{j=1}^{N} w_{ij} + 1 - \rho_\phi}\right), \qquad (2.1)\\
\delta_t|\boldsymbol{\delta}_{-t}, \mathbf{D}, \tau_\delta^2, \rho_\delta &\sim \text{N}\left(\frac{\rho_\delta \sum_{s=1}^{T} d_{ts}\delta_s}{\rho_\delta \sum_{s=1}^{T} d_{ts} + 1 - \rho_\delta}, \frac{\tau_\delta^2}{\rho_\delta \sum_{s=1}^{T} d_{ts} + 1 - \rho_\delta}\right),\\
\gamma_{it}|\tau_\gamma^2 &\sim \text{N}(0, \tau_\gamma^2),
\end{aligned}
$$

where θ_{it} is the estimated probability of smoking in the ith IG and tth year, which depends on terms including a vector of p covariates (including an intercept term) denoted by $\mathbf{x}_{it}$. The remaining terms in the linear predictor include $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_T)$, which are the overall spatial and temporal trends in the estimated probability $\{\theta_{it}\}$. Both are modelled by the conditional autoregressive (CAR) prior proposed by Leroux, Lei and Breslow (2000), which is able to capture dependence structures ranging from independence ($\rho_\phi = \rho_\delta = 0$) through to strong autocorrelation ($\rho_\phi = \rho_\delta = 1$). The dependence is represented by binary spatial and temporal neighbourhood matrices $\mathbf{W}_{N\times N}$ and $\mathbf{D}_{T\times T}$, respectively, where $w_{ij} = 1$ if areal units (i, j) share a common border and is zero otherwise, while $d_{st} = 1$ if the time periods are one unit apart (that is,

$|s-t|=1$) and zero otherwise. These additive main effects give the model a separable spatio-temporal structure, and this separability assumption is relaxed by adding a set of independent and identically distributed spatio-temporal interaction terms $\boldsymbol{\gamma}=(\gamma_{11},\ldots,\gamma_{NT})$ to the model. Other specifications for $\boldsymbol{\gamma}$ are possible; for details see Knorr-Held (2000). The second model we use as a comparator is the autoregressive decomposition described by Rushworth, Lee and Mitchell (2014) and given by

$$\begin{aligned} y_{it} &\sim \text{Binomial}(n_{it},\theta_{it}),\\ \ln\left(\frac{\theta_{it}}{1-\theta_{it}}\right) &= \mathbf{x}_{it}^{\top}\boldsymbol{\beta}+\phi_{it},\\ \boldsymbol{\phi}_1 &\sim \mathrm{N}\big(\mathbf{0},\tau^2\mathbf{Q}(\mathbf{W},\rho)^{-1}\big),\\ \boldsymbol{\phi}_t|\boldsymbol{\phi}_{t-1} &\sim \mathrm{N}\big(\xi\boldsymbol{\phi}_{t-1},\tau^2\mathbf{Q}(\mathbf{W},\rho)^{-1}\big) \qquad \text{for } t=2,\ldots,T. \end{aligned} \tag{2.2}$$

Here the spatial surface at time t, $\boldsymbol{\phi}_t=(\phi_{1t},\ldots,\phi_{Nt})$, evolves over time via a first order autoregressive process, whose precision matrix $\mathbf{Q}(\mathbf{W},\rho)=\rho[\text{diag}(\mathbf{W1})-\mathbf{W}]+(1-\rho)\mathbf{I}$ corresponds to the CAR prior proposed by Leroux, Lei and Breslow (2000). Here $(\mathbf{1},\mathbf{I})$ are a vector of ones and the identity matrix, respectively. For both models weakly informative inverse-gamma, uniform and Gaussian priors are specified for the variance, dependence and regression parameters, that is,

$$\begin{aligned} \tau_\phi^2,\tau_\delta^2,\tau_\gamma^2,\tau^2 &\sim \text{Inverse-Gamma}(1,0.01),\\ \rho_\phi,\rho_\delta,\rho,\xi &\sim \text{Uniform}(0,1),\\ \boldsymbol{\beta} &\sim \mathrm{N}(0,1000\mathbf{I}), \end{aligned}$$

where $\mathbf{I}$ is the $p\times p$ identity matrix.

3. Methodology. This section proposes a novel Bayesian spatio-temporal localised smoothing model for identifying clusters of elevated probability areas (Section 3.1), outlines the data augmentation strategy to account for the temporally overlapping nature of the data (Section 3.2), and describes the accompanying software that has been developed (Section 3.3). Inference for this model is based on MCMC simulation.

3.1. *Proposed model.* We initially describe the proposed model without the data augmentation because the overlapping nature of the data is specific to our maternal smoking application. Letting (y_{it},n_{it}) denote the number of pregnant women that smoke and the number of pregnant women, respectively, in the ith IG and tth year, we propose the following likelihood model:

$$\begin{aligned} y_{it} &\sim \text{Binomial}(n_{it},\theta_{it}),\\ \ln\left(\frac{\theta_{it}}{1-\theta_{it}}\right) &= \mathbf{x}_{it}^{\top}\boldsymbol{\beta}+\lambda_{Z_{it}}+\phi_{it}. \end{aligned} \tag{3.1}$$

As before, a weakly informative multivariate Gaussian prior is assigned to $\boldsymbol{\beta}$. The logit probability surface is modelled by a linear combination of covariates $\mathbf{x}_{it}^{\top}\boldsymbol{\beta}$ and two sets of latent effects, where $\{\phi_{it}\}$ are correlated and evolve smoothly in space and time, while $\{\lambda_{Z_{it}}\}$ is a piecewise constant intercept term. Thus, after adjusting for covariate effects spatially and temporally, adjacent probabilities $(\theta_{it}, \theta_{js})$ will be autocorrelated if $\lambda_{Z_{it}} = \lambda_{Z_{js}}$, but could exhibit very different values (a step change) if $\lambda_{Z_{it}} \neq \lambda_{Z_{js}}$. This formulation can thus be seen as a localised smoother, where the $\{\theta_{it}\}$ surface can exhibit areas of spatio-temporal smoothness separated by distinct step changes, the latter allowing spatially or temporally neighbouring areas to have very different probabilities of maternal smoking. Thus, the piecewise constant intercept term $\{\lambda_{Z_{it}}\}$ can identify clusters of IGs with unusually high (or low) probabilities of maternal smoking because if a group of adjacent IGs have a different $\lambda_{Z_{it}}$ value than their geographical neighbours, then they are likely to have markedly different estimated smoking probabilities.

The piecewise constant intercept term comprises at most G distinct levels $(\lambda_1, \ldots, \lambda_G)$, which are ordered via the prior

$$\lambda_j \sim \text{Uniform}(\lambda_{j-1}, \lambda_{j+1}) \qquad \text{for } j = 1, \ldots, G,$$

where $\lambda_0 = -\infty$ and $\lambda_{G+1} = \infty$. This order constraint ensures that $\lambda_1 < \lambda_2 < \cdots < \lambda_G$, which helps mitigate against the label-switching problem common in mixture models. The assignment of data point (i, t) to one of the G intercept terms is controlled by the indicator variable $Z_{it} \in \{1, \ldots, G\}$, and we note that the set of all NT indicators $\{Z_{it}\}$ does not have to cover the set $\{1, \ldots, G\}$, meaning that G is the maximum number of different intercept terms in the model. In the extreme case that $Z_{it} = k$ for all (i, t) for some value k, then the model reduces to a special case of the global smoothing model proposed by Rushworth, Lee and Mitchell (2014).

Here we fix the maximum number of intercept terms G in the model rather than estimating it using a reversible jump McMC algorithm similar to that used by Knorr-Held and Raßer (2000), partly because such algorithms can be slow to converge and exhibit poor mixing. Additionally, it is unlikely that G would be well identified in our setting because different values of G could result in identical $\{Z_{it}\}$ parameter sets. For example, the set $Z_{it} = 1$ if $t < 4$ and $Z_{it} = 2$ if $t \geq 4$ for modelling a region-wide temporal step change at time 4 could be obtained from all values of $G \geq 2$. This occurs because G is the maximum and not the actual number of intercept terms in the model, which we note is not the case in the model of Knorr-Held and Raßer (2000), where G represents the actual number of clusters in the model.

Our choice of prior $f(\mathbf{Z})$, where $\mathbf{Z} = \{Z_{it} | i = 1, \ldots, N, t = 1, \ldots, T\}$, is guided by two considerations. First, one may expect the probability of maternal smoking to evolve smoothly over time, which suggests a temporally autocorrelated prior such as a Markov model. We do not assume $\mathbf{Z}$ is spatially autocorrelated because $\{\lambda_{Z_{it}}\}$ captures a localised structure not captured by the spatially smooth

$\{\phi_{it}\}$. Additionally, Figure 2 shows that high incidence areas appear on opposite sides of Greater Glasgow that are spatially disconnected. The second consideration when constructing a prior for $\mathbf{Z}$ is that G is the maximum number of different intercept terms in the model, and thus we specify a value of G that is larger than the expected number of intercept terms required and use a penalty prior to encourage each Z_{it} towards the middle class. This middle class is $G^* = (G+1)/2$ if G is odd and $G^* = G/2$ if G is even, and this penalty ensures that Z_{it} is only estimated to be in one of the extreme classes if supported by the data. This penalty-based approach can be viewed as a discrete random variable analogue of ridge regression or penalised splines [Eilers and Marx (1996)], where in the latter too many basis functions are specified and the corresponding coefficients are smoothed towards each other. These two considerations suggest the following Markov decomposition:

$$f(\mathbf{Z}) = \prod_{i=1}^{N}\left[f(Z_{i1}) \prod_{t=2}^{T} f(Z_{it}|Z_{it-1}) \right], \tag{3.2}$$

where the individual components are given by

$$f(Z_{it}|Z_{it-1}) = \frac{\exp(-\delta[(Z_{it}-Z_{it-1})^2 + (Z_{it}-G^*)^2])}{\sum_{r=1}^{G}\exp(-\delta[(r-Z_{it-1})^2+(r-G^*)^2])} \qquad \text{for } t = 2,\dots,T,$$

$$f(Z_{i1}) = \frac{\exp(-\delta(Z_{i1}-G^*)^2)}{\sum_{r=1}^{G}\exp(-\delta(r-G^*)^2)}, \tag{3.3}$$

$$\delta \sim \text{Uniform}(1, M = 100).$$

Temporal autocorrelation in $\mathbf{Z}$ is induced by the $(Z_{it}-Z_{it-1})^2$ component of the penalty, while the $(Z_{it}-G^*)^2$ component penalises Z_{it} towards the middle risk class G^*. The size of this penalty, and hence the amount of smoothing imparted on $\mathbf{Z}$, is controlled by δ, which is assigned a uniform prior on a large range. To ensure some smoothing is imposed as G is larger than necessary, we set the lower limit of the prior for δ equal to one corresponding to exponential decay. A number of variations were investigated when developing this model, such as separate coefficients for the two penalty components, only having one of the two penalty components and using an L1 rather than an L2 penalty, but all performed poorer in initial simulations than the model proposed here.

The smoothing component ϕ_{it} models spatially and temporally autocorrelated variation in the logit of the probability surface $\{\theta_{it}\}$, via the multivariate autoregressive process:

$$\begin{aligned} \boldsymbol{\phi}_1 &\sim \mathrm{N}(\mathbf{0}, \tau^2\mathbf{Q}(\mathbf{W})^-), \\ \boldsymbol{\phi}_t|\boldsymbol{\phi}_{t-1} &\sim \mathrm{N}(\xi\boldsymbol{\phi}_{t-1}, \tau^2\mathbf{Q}(\mathbf{W})^-) \qquad \text{for } t = 2,\dots,T, \end{aligned} \tag{3.4}$$

where $\boldsymbol{\phi}_t = (\phi_{1t},\dots,\phi_{Nt})$. The joint distribution for $\boldsymbol{\phi}$ corresponding to (3.4) is a zero-mean Gaussian Markov Random field with precision matrix $\mathbf{Q}(\mathbf{W})^* =$

$\mathbf{C} \otimes \mathbf{Q}(\mathbf{W})$, where $\mathbf{C}$ is a $T \times T$ first order autoregressive matrix. As before, weakly informative Inverse-Gamma$(1, 0.01)$ and Uniform$(0, 1)$ priors are specified for (τ^2, ξ), respectively. The only difference from (2.2) is that $\rho = 1$, enforcing strong spatial smoothing on $\boldsymbol{\phi}_t$ so that any step changes in the surface are captured by $\{\lambda_{Z_{it}}\}$. We note that if ρ was estimated, it could be zero, resulting in both $(\phi_{it}, \lambda_{Z_{it}})$ being independent in space and thus competing for the same variation in the data. In implementing this model $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_T)$ are mean centred within the MCMC algorithm separately for data points with distinct $\{\lambda_{Z_{it}}\}$ values so that λ_j represents the mean logit probability for all data points in the jth intercept group. Thus, the posterior median of $\mathbf{Z}$ represents a grouping of the data into at most G groups, and is the mechanism by which clusters are identified.

3.2. *Data augmentation*. The model described above is not directly applicable to the maternal smoking data because the yearly data (y_{it}, n_{it}) are not available. Instead, three-year running totals $(v_{it} = y_{it-1} + y_{it} + y_{it+1}, m_{it} = n_{it-1} + n_{it} + n_{it+1})$ for $t = 2, \ldots, T-1$ are available for each IG, leading to the integer linear inverse problems $\mathbf{E}\mathbf{y}_i = \mathbf{v}_i$ and $\mathbf{E}\mathbf{n}_i = \mathbf{m}_i$ for each IG. Here the unknown yearly data are denoted by $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})_{T\times 1}$ and $\mathbf{n}_i = (n_{i1}, \ldots, n_{iT})_{T\times 1}$, while the known three-year totals are denoted by $\mathbf{v}_i = (v_{i2}, \ldots, v_{iT-1})_{T-2\times 1}$ and $\mathbf{m}_i = (m_{i2}, \ldots, m_{iT-1})_{T-2\times 1}$, respectively. The constraint matrix $\mathbf{E}$ is given by

$$\mathbf{E} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 1 & 1 & 0 & \ldots & 0 \\ & \vdots & & & & \vdots & \\ 0 & \ldots & 0 & 0 & 1 & 1 & 1 \end{bmatrix}_{T-2\times T} \qquad \tilde{\mathbf{E}} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{E} \\ \mathbf{e}_T \end{bmatrix}_{T\times T}.$$

However, if one specifies $(y_{i1}, y_{iT}, n_{i1}, n_{iT})$, then the remaining yearly data can be recovered via the equations $\mathbf{y}_i = \tilde{\mathbf{E}}^{-1}\tilde{\mathbf{v}}_i$ and $\mathbf{n}_i = \tilde{\mathbf{E}}^{-1}\tilde{\mathbf{m}}_i$, where $\tilde{\mathbf{v}}_i = (y_{i1}, \mathbf{v}_i, y_{iT})$, $\tilde{\mathbf{m}}_i = (n_{i1}, \mathbf{m}_i, n_{iT})$, $\mathbf{e}_1 = (1, 0, \ldots, 0)$ and $\mathbf{e}_T = (0, \ldots, 0, 1)$. Data augmentation is thus used to update $(y_{i1}, y_{iT}, n_{i1}, n_{iT})$ at each iteration of the MCMC algorithm, with the sampled $(\mathbf{y}_i, \mathbf{n}_i)$ further required to meet the binomial constraints $0 \le y_{it} \le n_{it}$ for all (i, t).

3.3. *Software*. The `R` [R Core Team (2013)] package `CARBayesST` has been developed in conjunction with this paper and can be downloaded from http://cran.r-project.org/. It can fit the localised smoothing model given by (3.1)–(3.4) as well as models (2.1) and (2.2). All these models can be applied to binomial (logistic link) and Poisson (log link) data, with a selection also being available for Gaussian data, making it widely useable beyond the specific application considered here. As the data augmentation outlined above is specific to this application, code to implement model (3.1)–(3.4) with data augmentation is available upon request from the first author. However, an example of using `CARBayesST` on simulated data is presented in Section 2 of the supplementary material [Lee and Lawson (2016)].

4. Model assessment via simulation. This section presents a simulation study, which assesses the performance of the clustering model proposed here across different values of G. The study generates and models yearly data without data augmentation, and an additional study exploring the model with data augmentation is presented in Section 3 of the supplementary material accompanying this paper [Lee and Lawson (2016)].

4.1. *Data generation and study design*. Simulated smoking incidence data are generated from binomial distributions for the $N = 271$ IGs and $T = 14$ time periods considered in the real study. The population sizes n_{it} are varied in this study to assess their impact on model performance. The logit probability surface is generated from a multivariate Gaussian distribution, with a piecewise constant mean (for clustering) and a spatially and temporally smooth variance matrix. The latter induces smooth spatio-temporal variation into the logit probability surface within a cluster, and is defined by a combination of a spatial exponential correlation function and a temporal first order autoregressive process. Clusters are induced into these data by the piecewise constant mean function, and we consider two different base templates:

- *Template A* is a constant vector corresponding to a probability of 0.25, and corresponds to generating no clusters in the spatio-temporal probability surface.
- *Template B* is a clustered surface with three levels, low probability of 0.07, medium probability of 0.25 and high probability of 0.46, which are similar to the real data. The spatial pattern in this cluster structure mimics the real data in the first time period, and is displayed in Section 4 of the supplementary material [Lee and Lawson (2016)]. IGs with a raw proportion less than 0.1 in the real data are in the low probability cluster, those with a raw proportion greater than 0.4 are in the high proportion group and those in between are in the middle group.

These two templates are combined to create 9 separate scenarios. Scenarios 1 to 3 are based on *Template A* with no clustering, and test whether the models falsely identify clusters when none are present. Scenarios 4 to 6 are based on *Template B*, and have the same cluster structure for all time periods. Finally, scenarios 7 to 9 correspond to temporally varying cluster structures, with *Template B* applying in the first 5 time periods, *Template A* in the next 5 and then finally *Template B* applies again for the last 4 time periods. In all three cases the number of pregnant women in each IG are 50, 100 and 200, respectively. Example realisations from both simulation templates under each value of n_{it} are displayed in Section 4 of the supplementary material [Lee and Lawson (2016)].

Two hundred data sets are generated under each of the 9 scenarios, and the model proposed here is applied to each data set with $G = 4, 5, 6, 7$ (the true values of G are 1 for *Template A* and 3 for *Template B*). We compare the performance of our clustering model to models (2.1), denoted *Model K*, and (2.2), denoted *Model R*, commonly used in the literature. Inference for each model is based on 20,000

McMC samples, which were generated following a burn-in period of 20,000 samples. Convergence was visually assessed to have been reached after 20,000 samples by viewing trace plots of sample parameters for a number of simulated data sets.

Model performance is summarised using two main metrics, the root mean square error (RMSE) of the estimated probability surface and the Rand index [Rand (1971)] of the estimated cluster structure. RMSE is computed as $\text{RMSE} = \sqrt{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(\theta_{it} - \hat{\theta}_{it})^2}$, where $\hat{\theta}_{it}$ is the posterior median for θ_{it}. The Rand Index quantifies a model's ability to correctly identify the true cluster structure in the data, and measures the proportion of agreement between the true and estimated cluster structures from each model, with a value of one indicating the structures are identical. The cluster structure estimated by the model proposed here is summarised by the posterior median of $\{Z_{it}\}$. In contrast, *Model K* and *Model R* do not have inbuilt clustering mechanisms, so we implement the posterior classification approach described in Charras-Garrido, Abrial and de Goer (2012), which applies a Gaussian mixture model to the posterior median probability surface to obtain the estimated cluster structure. Additionally, we also present the coverage probabilities of the 95% uncertainty intervals for the clustering indicators $\{Z_{it}\}$.

4.2. *Results*. The results of this study are displayed in Table 1, where the top panel displays the RMSE, the middle panel displays the Rand index, and the bottom panel displays the coverage probabilities. In all cases the median values over the 200 simulated data sets are presented. The table shows a number of key messages. First, the clustering model proposed here is not sensitive to the choice of the maximum number of clusters G, as all results are largely consistent over G. For example, the median (over the 200 simulated data sets) Rand index varies by at most 0.014, while the median RMSE varies by at most 0.008. Second, the clustering model has consistently excellent cluster identification, as the median Rand index ranges between 0.969 and 1 across all scenarios and values of G. Third, this excellent clustering is at odds with that observed by applying a posterior classification approach to the fitted proportions estimated from *Model K* and *Model R*. These models illustrate good clustering performance if there are true clusters in the data (scenarios 4–9), showing comparable results to the clustering model proposed here. However, if there are no clusters in the data (scenarios 1 to 3), then these models identify clusters that are not present (they identify 2 or 3 clusters on average), as they have median Rand indexes between 0.504 and 0.599. This suggests that a posterior classification approach should not be used for cluster detection in this context due to the identification of false positives. Fourth, the clustering model proposed here produces comparable or better probability estimates $\{\hat{\theta}_{it}\}$ (as measured by RMSE) than *Model K* and *Model R* in all scenarios, with the improvement being most pronounced in scenarios 7 to 9. Finally, the coverage probabilities for the clustering indicators $\{Z_{it}\}$ are all above 90%, and typically are more conservative than the nominal 95% level.

TABLE 1

Results of the simulation study. The top panel displays the root mean square error (RMSE) for the estimated probability surface, the middle panel displays the Rand index and the bottom panel displays the coverage probabilities for $\{Z_{it}\}$. The first three rows relate to a probability surface with no clusters, the second three to a surface with temporally consistent clusters, and the last three to temporally inconsistent clusters. The localised model (with $G = 4, 5, 6, 7$) is compared to (2.1), denoted Model K, and (2.2), denoted Model R

			Localised mode					
Scenario	**Clustering**	n_{it}	$G = 4$	$G = 5$	$G = 6$	$G = 7$	**Model K**	**Model R**
RMSE								
1	None	50	0.008	0.008	0.008	0.008	0.008	0.007
2	None	100	0.007	0.007	0.007	0.007	0.007	0.006
3	None	200	0.006	0.006	0.006	0.006	0.006	0.006
4	Consistent	50	0.020	0.019	0.019	0.019	0.018	0.032
5	Consistent	100	0.019	0.012	0.016	0.011	0.013	0.026
6	Consistent	200	0.008	0.007	0.007	0.007	0.010	0.021
7	Inconsistent	50	0.027	0.026	0.026	0.026	0.044	0.040
8	Inconsistent	100	0.014	0.013	0.014	0.014	0.036	0.032
9	Inconsistent	200	0.007	0.007	0.007	0.007	0.028	0.025
Rand								
1	None	50	1.000	1.000	1.000	1.000	0.566	0.597
2	None	100	1.000	1.000	1.000	1.000	0.559	0.565
3	None	200	1.000	1.000	1.000	1.000	0.539	0.540
4	Consistent	50	0.985	0.987	0.987	0.988	1.000	0.973
5	Consistent	100	0.983	0.996	0.987	0.997	1.000	0.988
6	Consistent	200	0.998	0.999	0.999	0.999	1.000	0.797
7	Inconsistent	50	0.969	0.970	0.970	0.970	0.950	0.967
8	Inconsistent	100	0.993	0.994	0.992	0.992	0.989	0.993
9	Inconsistent	200	0.999	0.998	0.998	0.956	0.999	1.000
Coverage of Z_{it}								
1	None	50	1.000	1.000	1.000	1.000	–	–
2	None	100	1.000	1.000	1.000	1.000	–	–
3	None	200	1.000	1.000	1.000	1.000	–	–
4	Consistent	50	0.965	0.980	0.986	0.989	–	–
5	Consistent	100	0.931	0.952	0.955	0.969	–	–
6	Consistent	200	0.928	0.915	0.911	0.885	–	–
7	Inconsistent	50	0.999	0.999	0.999	0.999	–	–
8	Inconsistent	100	0.991	0.990	0.982	0.976	–	–
9	Inconsistent	200	0.964	0.943	0.934	0.916	–	–

5. Results of the Glasgow maternal smoking study. Three models were applied to the Glasgow maternal smoking data, the localised spatio-temporal smoothing model proposed in Section 3 with values of G between 4 and 7, as well as *Model K* and *Model R* outlined by (2.1) and (2.2), respectively. In all cases the data augmentation strategy outlined in Section 3.2 was applied to obtain inference on the yearly probability surfaces $\{\theta_{it}\}$ from the available three-year rolling totals. Inference in all cases was based on 25,000 MCMC samples generated from 5 parallel Markov chains that were burnt-in until convergence, the latter being assessed by examining trace plots of sample parameters. The supplementary material [Lee and Lawson (2016)] summarises the hyperparameters in the model (Section 5) as well as providing sensitivity analyses (Section 6) to the choice of some prior distributions.

5.1. *Model fit*. The overall fit of each model to the data is summarised in Table 2, which displays results with and without the socio-economic deprivation covariates. The table displays the Watanabe–Akaike information criterion [WAIC, Watanabe (2010)], as well as an estimate of the effective number of parameters (P.W). The table shows that varying G between 4 and 7 in the localised smoothing model results in almost no difference in model fit, with WAIC differing by at most 15 out of a total of around 18,600. The localised smoothing model fits the data better than *Model K* and *Model R* with or without covariates, with differences of around 665 for *Model K* and between 58 and 135 for *Model R*. *Model R* is close to a simplification of the localised smoothing model without the piecewise constant intercept term, and the inclusion of the latter has reduced the random effects ($\{\phi_{it}\}$) variance τ^2 from around 0.279 to 0.206. Finally, we note that the inclusion of the covariates has not changed the overall fit of the localised smoothing model greatly, but has reduced the effective number of parameters due to a reduction in the random effects variance τ^2 from 0.206 to 0.109.

TABLE 2
Watanabe–Akaike information criteria (WAIC) and the effective number of parameters (P.W) for each model

	No covariates		**Covariates**	
Model	**WAIC**	**P.W**	**WAIC**	**P.W**
Model localised: $G = 4$	18,587	1379	18,610	1281
Model localised: $G = 5$	18,596	1380	18,606	1276
Model localised: $G = 6$	18,602	1385	18,606	1281
Model localised: $G = 7$	18,591	1390	18,610	1277
Model K	19,270	1599	19,269	1617
Model R	18,737	1389	18,666	1262

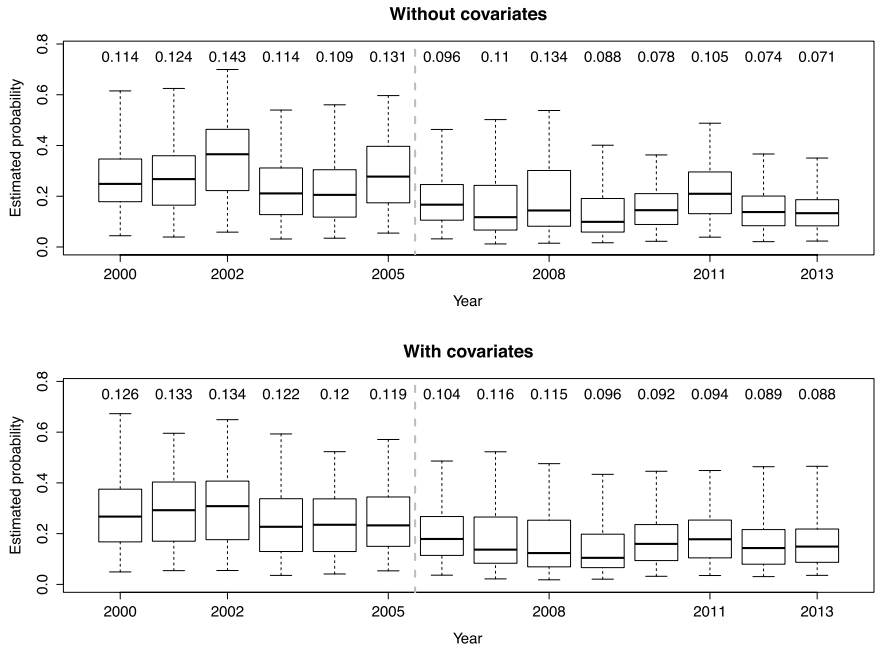

FIG. 3. *The figure shows boxplots of the spatial distribution of estimated smoking probabilities by year from the localised smoothing model with* $G = 4$ *with and without covariates. The dashed line marks the time of the smoking ban and the numbers are spatial standard deviations in smoking probabilities.*

5.2. *Covariate effects*. Both the socioeconomic deprivation covariates exhibited substantial effects on maternal smoking rates, with the following odds ratios and 95% credible intervals for a one standard deviation increase in the percentage of people claiming JSA (sd = 2.45) and the natural log of median property price (sd = 0.50): JSA—1.46 (1.41, 1.52); log price—0.73 (0.70, 0.76). These results relate to the localised smoothing model with $G = 4$, but results from the other models are almost identical. Thus, both results suggest that an increase in an areas level of socioeconomic deprivation results in a substantial increase in the odds of maternal smoking.

5.3. *Temporal trend and spatial inequalities*. The temporal trend in maternal smoking probabilities is displayed in Figure 3, which shows boxplots of the estimated probabilities across all IGs for each year. The dashed line denotes the time of the smoking ban, while the numbers at the top of the figure are spatial standard deviation quantifying the level of spatial inequality in estimated smoking probabilities. The results are presented for the localised smoothing model (with $G = 4$) with and without covariates because Table 2 shows it fits the data better than *Model K* or *Model R*. The results using other values of G are almost identical, having a

mean absolute difference of 0.004 on the probability scale. The figure shows clear evidence of an overall decline in smoking probabilities during the 14 years, with estimated reductions of 0.115 and 0.118 in the median smoking probabilities between 2000 and 2013 for the models without and with covariates, respectively. This suggests that in an era encompassing the smoking ban (March 2006) there was a reduction in maternal smoking probabilities by just below 12% on average in Glasgow, although the figure does not show a clear step change reduction between 2005 and 2006. Furthermore, these results do not show a monotonic decline and instead show some year-to-year variation, which may be due to random variation or the need to estimate the yearly data within the model using data augmentation. Reductions in the spatial inequality in estimated smoking probabilities show similar patterns, with the standard deviation falling by around 0.04 (a 35% reduction) between 2000 and 2013, which is broadly consistent to including or excluding covariates from the model.

5.4. *Localised spatio-temporal structure*. The piecewise constant intercept terms in the localised smoothing model allow spatially or temporally neighbouring data points to have very different estimated smoking probabilities, and thus a group of adjacent data points with a different intercept value from their neighbours could be viewed as a cluster with excessively high (or low) smoking probabilities. The model was fitted with the maximum number of different intercept terms $G = 4, \ldots, 7$, and in all cases only 3 different intercept terms were identified. The allocation of these three intercept terms to the NT data points was also consistent across G, with Rand index values that measure cluster agreement ranging between 0.98 and 0.99 (1 corresponds to complete agreement). For the model with no covariates the three different groups comprise low, medium and high maternal smoking groups, with average probabilities of 0.064, 0.172 and 0.413, respectively. Around 9% of IGs are in the high probability group in 2000 (based on the posterior median), which reduces to less than 1% in 2013.

Figure 4 displays the probability that each IG is in the high probability cluster in 2000, 2006 and 2013, where the left column relates to a model with no covariates, while the right panel is high unexplained values after adjusting for socioeconomic deprivation. Without adjusting for deprivation, the largest and most temporally preserving cluster of high probability IGs is the Nitshill/Pollockshaws part of the city, which is just north west of Giffnock (see Figure 1). This cluster reduces in size by 2013, but is about the only group of IGs that retain any substantial probability of being in the high probability group. The remaining clusters in 2000 include Port Glasgow in the west (see Figure 1) and Drumchapel in the mid-north (just below Bearsden in Figure 1), but both exhibit reduced probabilities of maternal smoking in 2013, and have posterior medians for $\{Z_{it}\}$ corresponding to being in the middle intercept group along with almost all of the other IGs. The right column of Figure 4 shows IGs that retain relatively high maternal smoking rates after removing the effects of socioeconomic deprivation, and some are in common with the left panel

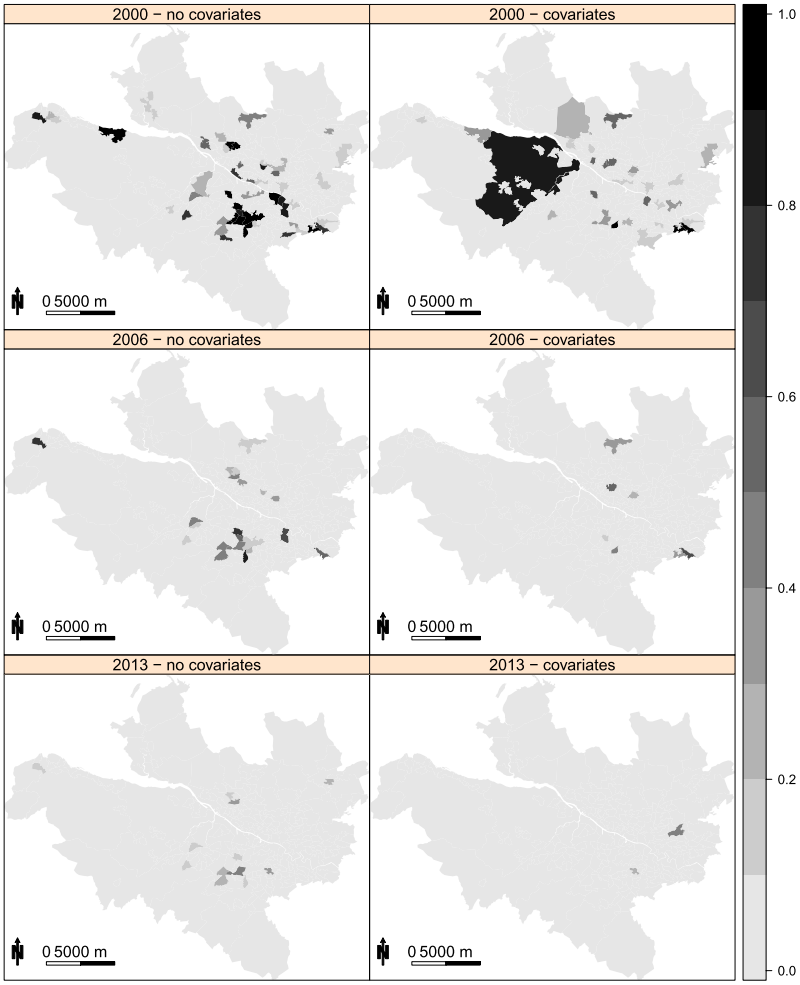

FIG. 4. *The probability of being in the high probability intercept group in* 2000, 2006 *and* 2013. *The left and right panels respectively relate to the model without and with the covariates.*

such as Cambslang (far southeast, east of Giffnock, see Figure 1), while others such as rural Renfrewshire (large black region) are not. In common with the left column, the majority of these areas exhibit reduced incidence by 2013, which is in keeping with the overall decline in maternal smoking observed in Figure 3.

6. Discussion. This paper has presented a new study investigating the changing temporal dynamics of small-area variation and inequality in maternal smoking during pregnancy in Glasgow, Scotland, during an era that included a ban on smoking in enclosed public spaces. To identify these dynamics, we have presented a novel localised smoothing model that is one of the first to simultaneously un-

dertake both cluster identification and spatio-temporal risk estimation in this epidemiological context. The model can be implemented by others using the `CARBayesST` package in the statitistical software `R` for both binomial and Poisson data models, making it widely applicable beyond the specific application considered here. A simulation study has shown the model performs consistently well across a range of scenarios, both in terms of cluster identification and risk estimation, and outperforms two commonly used competitor models.

The results from the Glasgow maternal smoking study show that overall smoking incidence in Glasgow has reduced by around 12%, being around 26% on average in 2000 and reducing to 14% in 2013. This is very similar to the results found in the literature, with, for example, Passmore et al. (2015) reporting a reduction in maternal smoking rates from 17% in 2000 to 11% in 2011 in Australia. Reductions in maternal smoking have also been seen in much earlier time periods, with, for example, Silveira et al. (2016) reporting a reduction from 36% in 1982 to 21% by 2011 in Brazil.

The reduction observed here coincides with a ban on smoking in 2006, although there does not appear to be a step change reduction in that year. This may be due to the anticipatory effect as described by Mackay et al. (2012), whereby people change their behaviours in advance of the ban in order to prepare for it. Alternatively, the available data were 3-year rolling totals, so the yearly data had to be estimated within the model using data augmentation, leading to greater uncertainty. Thus, while one cannot definitively say the observed reduction in maternal smoking is due to the ban as opposed to other factors such as more stringent controls on cigarette packaging and availability, the ban has coincided with a reduction in smoking rates over the 14 year study period which will have a knock-on effect in improving public health.

Our study also found strong relationships between maternal smoking rates and socioeconomic deprivation, the latter measured by average property price and the proportion of working age people claiming unemployment benefits. This relationship agrees with the existing literature [(Silveira et al. (2016), Williamson et al. (1989)], and a review by Kramer et al. (2000) shows that increased rates of cigarette smoking among poor communities is partially responsible for the socioeconomic disparities in poor pregnancy outcomes such as preterm birth. Additionally, Passmore et al. (2015) show a significant relationship between socioeconomic deprivation and smoking cessation during pregnancy, with those that smoke before pregnancy being more likely to quit during pregnancy if they are from affluent backgrounds. These effects recur across the world and in different time periods, and show that socioeconomic deprivation remains a key driver in maternal smoking prevalences.

The level of spatial inequality in maternal smoking incidence across Glasgow has also reduced over the study period, with the standard deviation in estimated smoking probability reducing from 0.114 in 2000 to 0.071 in 2013. This reduction results mainly from high smoking incidence areas reducing their levels, as

the maximum estimated smoking incidence has reduced from 61.8% to 38.1%, whereas the minimum estimated level has only changed from 4.4% to 2.5%. This reduction in the high incidence areas is also observed from the number of IGs estimated to be in the high incidence cluster (summarised in Figure 4), which is around 9% of IGs in 2000 but less than 1% in 2013. Thus, the city has become more equal in terms of smoking incidence over the 14 year study period, which has reduced the level of health inequality and should make for a fairer society in future health terms. For example, the spatial standard deviation in the standardised morbidity ratio for hospitalisation due to respiratory disease in the general population across Glasgow has reduced from 0.332 at the start of the study period to 0.305 at the end. However, whilst most areas have reduced their maternal smoking rates over the course of the study, the Nitshill/Pollockshaws part of the city still shows evidence of raised levels in 2013, and this would thus be an area for public health professionals to investigate further to understand why the reductions seen elsewhere have not happened in this area.

This paper presents a number of natural avenues for future work. On the public health side one would wish to investigate whether the reductions in maternal smoking incidences observed here have filtered through to similar improvements in birth outcomes, as well as whether the pattern in maternal smoking rates are mirrored by reductions in rates for the entire population. From a statistical viewpoint, it would be interesting to extend the clustering model so that the values of the piecewise constant intercept terms $\{Z_{it}\}$ were determined by covariates such as socioeconomic deprivation using an approach similar to Gormley and Murphy (2008).

Acknowledgements. The authors gratefully acknowledge the helpful comments from the editor and two reviewers which have greatly improved the paper. We also thank Dr. Ludger Evers for his helpful discussions regarding the data augmentation. The data and shapefiles used in this study were provided by the Scottish Government. *Conflict of Interest*: None declared.

SUPPLEMENTARY MATERIAL

Supplement A: Additional results and data analysis (DOI: 10.1214/16-AOAS941SUPP; .pdf). Section 2 illustrates the use of the (non-data augmented) proposed model via the `CARBayesST` package on simulated data. Section 3 provides an additional simulation study that assesses the efficacy of the model with data augmentation. Section 4 presents example realisations of the simulated data. Section 5 presents posterior summaries of the hyperparameters from the Glasgow study, while Section 6 presents additional sensitivity analyses.

Supplement B: Additional files for running the proposed model on simulated data (DOI: 10.1214/16-AOAS941SUPP; .zip). R code and data for running the model proposed in Section 3 on simulated data.

REFERENCES

ANDERSON, C., LEE, D. and DEAN, N. (2014). Identifying clusters in Bayesian disease mapping. *Biostat*. **15** 457–469.

BAULD, L., FERGUSON, J., LAWSON, L., CHESTERMAN, J. and JUDGE, K. (2005). Tackling smoking in Glasgow. Technical report, Glasgow Centre for Population Health.

CHARRAS-GARRIDO, M., ABRIAL, D. and DE GOER, J. (2012). Classification method for disease risk mapping based on discrete hidden Markov random fields. *Biostat*. **13** 241–255.

CHARRAS-GARRIDO, M., AZIZI, L., FORBES, F., DOYLE, S., PEYRARD, N. and ABRIAL, D. (2013). On the difficulty to delimit disease risk hot sports. *Journal of Applied Earth Observation and Geoinformation* **22** 99–105.

CHOI, J., LAWSON, A. B., CAI, B. and HOSSAIN, MD. M. (2011). Evaluation of Bayesian spatiotemporal latent models in small area health data. *Environmetrics* **22** 1008–1022. MR2861574

CNATTINGIUS, S. (2004). The epidemiology of smoking during pregnancy: Smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob. Res*. **6 Suppl 2** S125–S140.

EILERS, P. H. C. and MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci*. **11** 89–121. MR1435485

FORBES, F., CHARRAS-GARRIDO, M., AZIZI, L., DOYLE, S. and ABRIAL, D. (2013). Spatial risk mapping for rare disease with hidden Markov fields and variational EM. *Ann. Appl. Stat*. **7** 1192–1216. MR3113506

GANGNON, R. and CLAYTON, M. (2000). Bayesian detection and modeling of spatial disease clustering. *Biometrics* **56** 922–935.

GORMLEY, I. C. and MURPHY, T. B. (2008). A mixture of experts model for rank data with applications in election studies. *Ann. Appl. Stat*. **2** 1452–1477. MR2655667

GRAY, L. and LEYLAND, A. (2009). "Glasgow effect" of cigarette smoking explained by socioeconomic status? A multilevel analysis. *BMC Public Health* **9** 245.

GRAY, R., BONELLIE, S., CHALMERS, J., GREER, I., JARVIS, S., KURINCZUK, J. and WILLIAMS, C. (2009). Contribution of smoking during pregnancy to inequalities in stillbirth and infant death in Scotland 1994–2003: Retrospective population based study using hospital maternity records. *British Medical Journal* **339** b3754.

GRAY, L., MERLO, J., MINDELL, J., HALLQVIST, J., TAFFOREAU, J., O'REILLY, D., REGIDOR, E., NAESS, O., KELLEHER, C., HELAKORPI, S., LANGE, C. and LEYLAND, A. (2012). International differences in self-reported health measures in 33 major metropolitan areas in Europe. *European Journal of Public Health* **22** 40–47.

GREEN, P. J. and RICHARDSON, S. (2002). Hidden Markov models and disease mapping. *J. Amer. Statist. Assoc*. **97** 1055–1070. MR1951259

KNORR-HELD, L. (2000). Bayesian modelling of inseparable space–time variation in disease risk. *Stat. Med*. **19** 2555–2567.

KNORR-HELD, L. and RASSER, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* **56** 13–21.

KRAMER, M., SEGUIN, L., LYDON, J. and GOULET, L. (2000). Socio-economic disparities in pregnancy outcome: Why do the poor fare so poorly? *Paediatr. Perinat. Epidemiol*. **14** 194–210.

KULLDORFF, M., HEFFERNAN, R., HARTMAN, J., ASSUNCAO, R. and MOSTASHARI, F. (2005). A space-time permutation scan statistic for disease outbreak detection. *PLoS Med*. **2** 216–224.

LAWSON, A. B. (2009). *Bayesian Disease Mapping*: *Hierarchical Modeling in Spatial Epidemiology*. CRC Press, Boca Raton, FL. MR2484272

LAWSON, A. B., CHOI, J., CAI, B., HOSSAIN, M., KIRBY, R. S. and LIU, J. (2012). Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. *J. Agric. Biol. Environ. Stat*. **17** 417–441. MR2993274

LEE, D. and LAWSON, A. (2016). Supplement to "Quantifying the spatial inequality and temporal trends in maternal smoking rates in Glasgow." DOI:10.1214/16-AOAS941SUPPA, DOI:10.1214/16-AOAS941SUPPB.

LEROUX, B. G., LEI, X. and BRESLOW, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (*Minneapolis, MN*, 1997). *IMA Vol. Math. Appl.* **116** 179–191. Springer, New York. MR1731684

LI, G., BEST, N., HANSELL, A., AHMED, I. and RICHARDSON, S. (2012). BaySTDetect: Detecting unusual temporal patterns in small area data via Bayesian model choice. *Biostat.* **13** 695–710.

MACKAY, D., NELSON, S., HAW, S. and PELL, J. (2012). Impact of Scotland's smoke-free llegislation on pregnancy complications: Retrospective cohort study. *PLoS Med.* **9** e1001175.

PASSMORE, E., MCGUIRE, R., CORRELL, P. and BENTLEY, J. (2015). Demographic factors associated with smoking cessation during pregnancy in New South Wales, Australia, 2000–2011. *BMC Public Health* **15** 398.

RAND, W. (1971). Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* **66** 846–850.

RUSHWORTH, A., LEE, D. and MITCHELL, R. (2014). A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and Spatio-temporal Epidemiology* **10** 29–38.

SILVEIRA, M., MATIJASEVICH, A., MENEZES, A., HORTA, B., SANTOS, I., DARROS, A., BARROS, F. and VICTORA, C. (2016). Secular trends in smoking during pregnancy according to income and ethnic group: Four population-based perinatal surveys in a Brazilian city. *BMJ Open* **6** e010127.

TAPPIN, D., MACASKILL, S., BAULD, L., EADIE, D., SHIPTON, D. and GALBRAITH, L. (2010). Smoking prevalence and smoking cessation services for pregnant women in Scotland. *Substance Abuse Treatment, Prevention and Policy* **5** 1.

R CORE TEAM (2013). *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

UGARTE, D., ETXEBERRIA, J., GOICOA, T. and ARDANAZ, E. (2012). Gender-specific spatio-temporal patterns of colorectal cancer incidence in Navarre, Spain (1990–2005). *Cancer Epidemiol.* **36** 254–262.

WAKEFIELD, J. and KIM, A. (2013). A Bayesian model for cluster detection. *Biostat.* **14** 752–765.

WANG, X., ZUCKERMAN, B., PEARSON, C., KAUFMAN, G., CHEN, C., WANG, G., NIU, T., WISE, P., BAUCHNER, H. and XU, X. (2002). Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight. *J. Am. Med. Assoc.* **287** 195–202.

WATANABE, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594. MR2756194

WILLIAMSON, D., SERDULA, M., KENDRICK, J. and BINKIN, N. (1989). Comparing the prevalence of smoking in pregnant and nonpregnant women. *J. Am. Med. Assoc.* **261** 70–74.

D. LEE
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF GLASGOW
GLASGOW G12 8QQ
UNITED KINGDOM
E-MAIL: Duncan.Lee@glasgow.ac.uk

A. LAWSON
DIVISION OF BIOSTATISTICS
AND BIOINFORMATICS
DEPARTMENT OF PUBLIC HEALTH SCIENCES
MEDICAL UNIVERSITY OF SOUTH CAROLINA
CHARLESTON, SOUTH CAROLINA 29401-8350
USA
E-MAIL: lawsonab@musc.edu