Information Fusion xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Information Fusion

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

Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice

David Banks [a,*], Gauri Datta [b], Alan Karr [c], James Lynch [d], Jarad Niemi [a], Francisco Vera [e]

[a] Dept. of Statistical Science, Duke University, Box 90251, Durham, NC 27708, United States
[b] Department of Statistics, University of Georgia, Athens, GA 30602, United States
[c] National Institute of Statistical Sciences, Research Triangle Park, NC 27709, United States
[d] Department of Statistics, University of South Carolina, Columbia, SC 29208, United States
[e] Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, United States

ARTICLE INFO

Article history:
Received 6 June 2008
Received in revised form 19 July 2009
Accepted 30 October 2009
Available online xxxx

Keywords:
CAR models
Bayes
Syndromic surveillance
Gibbs distribution
Markov random field

ABSTRACT

Syndromic surveillance has, so far, considered only simple models for Bayesian inference. This paper details the methodology for a serious, scalable solution to the problem of combining symptom data from a network of US hospitals for early detection of disease outbreaks. The approach requires high-end Bayesian modeling and significant computation, but the strategy described in this paper appears to be feasible and offers attractive advantages over the methods that are currently used in this area. The method is illustrated by application to ten quarters worth of data on opioid drug abuse surveillance from 636 reporting centers, and then compared to two other syndromic surveillance methods using simulation to create known signal in the drug abuse database.

1. Introduction

There is enormous social benefit from early discovery of disease. Quick detection of emerging geographical clusters of disease has important benefits for public health; swift intervention can prevent a pandemic. The stakes are immense: the US congressional Office of Technology Assessment estimates that a release of 100 KG of anthrax spores upwind of Washington, DC, would, if not detected rapidly, lead to as many as three million deaths and nearly a trillion dollars of life and economic loss [33].

However, the first few cases of a disease are easy to overlook, since they may be misdiagnosed as atypical presentations of a relatively benign illness or isolated cases caused by rare transmission vectors. Definitive diagnoses, especially of rare diseases, require time. To overcome the lag implied by traditional public health notification structures based on final diagnoses, the Department of Homeland Security and the Centers for Disease Control have discussed implementation of a monitoring system in which all hospitals in the United States report, in real or near-real-time, admissions and out-patient visits associated with several kinds of symptoms. The expectation is that an outbreak of, say, anthrax will be visible first as a regional up-tick in the number of people treated for severe coughs, or that epidemic flu is signaled by a surge in patients with fever.

The statistical theory and methodology underlying a successful syndromic surveillance system must satisfy many criteria. Specifically, the system must:

1. *Be scalable.* Scalability requires that the method be computable, in appropriate time, on data sets of the size expected for disease surveillance. Ultimately, an effective program would link all major US hospitals into the reporting network. This implies data from about 7500 hospitals.
2. *Be multivariate.* The input data will be in terms of multivariate counts; hospitals will report admissions according to syndromic categories, such as fever, cough, rash, and combinations of these. Such counts are not independent. When there are additional sources of data, more subtle dependencies can arise. For example, there is overlap between the people absent from work and those purchasing over-the-counter medications, and both kinds of data have been proposed as useful data sets for an integrated syndromic surveillance system. A successful system must handle multivariate inference.
3. *Incorporate covariates.* Different geographical regions have different disease profiles. Hospitals in Miami, FL service an elderly population, and will tend to report different kinds of illnesses. Less affluent populations make greater use of emergency room

[*] Corresponding author. Tel.: +1 919 684 3743.
E-mail address: banks@stat.duke.edu (D. Banks).

1566-2535/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.inffus.2009.10.005

facilities. Statistical analysis must account for demographic information such as the size of the local population, its age structure, and perhaps local temperature or pollution levels [22].

4. *Change with time.* There are well-known weekly and daily patterns in hospital admissions and emergency room visits, as well as seasonal trends in disease, short-term demographic changes (e.g., summer vacation in college towns), all of which must be accounted for when distinguishing bioterrorism from natural outbreaks from random fluctuation.
5. *Accommodate low quality data.* Data quality is a complex, multifaceted issue [40]. The multiplicity of sources requires particular attention to data quality issues.
6. *Support decision-makers.* In order to be a usable real-time tool, a syndromic surveillance system must produce easily interpretable results. Especially given public, media and political hypersensitivity about disease outbreaks, false alarm rates must be low, and also commensurate with the resources available for post-signal investigation. To enable confidence, the modeling in the system must be understandable, at least at a heuristic level.

Statistically, the unifying theme for these criteria is principled calculation of uncertainties.

Satisfying these criteria poses daunting challenges, leading to the arrival of syndromic surveillance as a new interdisciplinary research area. No current approach to syndromic surveillance satisfies all (or even most) of these requirements.

As with many fields, syndromic surveillance has grown opportunistically. Initial work relied heavily upon off-the-shelf statistics, such as CUSUM charts (CDC), regression and exponentially weighted moving averages (the RODS project), and standard time series models [8]. Some “cocktail” approaches were developed, such as a combination of time series and control charts for early detection of anthrax outbreaks by tracking over-the-counter medication sales [33]. The most statistically advanced procedures involve the spatial scan statistic [41] which counts the number of disease reports in a given geographic window (for instance, a circle of fixed radius) and then “slides” that window over the region to determine whether there is any center-point whose window contains a count that is significantly higher than expected under the null model of no epidemic. This was extended to detect space-time clustering of disease [42], allowing one to find hotspots that are concentrated in time as well as space. Further extensions are given by [1,15,23,24,43,44]. These methods require some work to establish the null distribution of their test statistics; permutation methods are often used, but these make independence assumptions that are problematic for contagious disease.

These methods use classical frequentist statistics, and consequently suffer from issues of multiple testing, interpretability of results, high false alarm rates versus low power, and inadequate use of covariate information. Only initial work has been done from a Bayesian perspective; Neill et al. [49,50] describe an elementary model with convenient priors, but do not draw upon the full power of Bayesian hierarchical models for complex spatial and temporal dependence, nor do they address the need for flexible models that incorporate essential covariate information on population size and seasonality. For these and other reasons, all previous work in this area fails on at least one of five criteria that we have listed.

To address these shortfalls, Section 2 lays out a sophisticated, fully Bayesian approach to syndromic surveillance for disease outbreaks. Section 3 applies that methodology to the problem of drug abuse surveillance; our data set and model are slightly smaller and simpler than is needed for national hospital surveillance, but we establish feasibility and highlight practical issues. Section 3 also compares the CAR results to surveillance systems based on CUSUM charts and paired comparisons. Section 4 discusses the results, and indicates the technical extensions that are needed to address large-scale problems.

2. Bayesian syndromic surveillance

Bayesian models for spatial and spatio-temporal processes have become prominent in applied statistics [4,5,10,2,56]. These models pose special computational issues, and each application requires individual modeling. Nonetheless, there is growing consensus in the applied statistics community that when such models are carefully constructed and tested, they represent a powerful approach to big, important problems.

We propose a hierarchical model for the disease reporting process. At the bottom level of the hierarchy are vectors of counts from each hospital of patients with specific symptoms. Above that level, the spatial structure is determined by a graph whose vertices are hospitals and whose edges link “cliques” of hospitals that are sufficiently close to experience a common outbreak. Similarly, the temporal structure is determined by dependence of the reported counts upon previous reports from the same hospital and those to which it is connected in the graph. At the top level of the hierarchy, one has general location effects. Such systems are called conditional autoregressive models, or (CAR) models.

CAR models have their roots in the Ising models of physics, now generalized as Markov random field models. Inference is done through (MCMC), but problems of the scale considered in this application pose significant computational challenges. The research issues in the model and the computation are described in the following subsections.

2.1. The basic model

For clarity, we start with a simple case, assuming univariate reports on a single symptom (say, fever), a simple ARMA model for the noise term in the model for fever counts, and covariates that are measured without error.

The measurement at hospital i on day t is $Y_i(t)$—the number of reported patients who have, say, high fevers. We suppose that there are m hospitals and T days of data. From epidemiological considerations, for most diseases the signals will manifest within 7–14 days.

To start, we assume that the number of cases $Y_i(t)$ at hospital i on day t has a Poisson distribution. (This can be extended to the more realistic case of multivariate records and extra-Poisson variation.) In the absence of an epidemic, the mean function of the Poisson count at hospital i is $\mu_i(t)$; when there is a disease, there are additional cases and the mean of the additional count is $\lambda_i(t)$. An indicator function $\delta_i(t)$ marks whether a disease is present. Thus the model for the data at hospital i is:

$$Y_i(t) \sim \text{Pois}(\mu_i(t) + \delta_i(t)\lambda_i(t)). \tag{1}$$

Using this, we want to make an inference about the probability that one or more of the $\delta_i(t)$ are non-zero, i.e., the posterior probability that there has been a disease outbreak.

We can compute the posterior probability for a well-constructed hierarchical Bayes model; the first stage of this model is based on the Poisson distribution specified by Eq. (1). The spatial dependence is captured through the CAR model. Unlike the scan statistic, our models can easily and directly incorporate additional covariate information. Also unlike the scan statistic, which concentrates on finding only the primary cluster, and occasionally, secondary cluster(s), our method identifies all important regions and time windows based on the posterior distribution of relevant model parameters. Indeed, as in any realistic Bayesian application, our

Bayesian inference is driven by posterior summaries of relevant model parameters. Such summaries are obtained via MCMC. In particular, we rely on Gibbs sampling [29] and the Metropolis–Hastings algorithm [47,36].

The model given by (1) is obtained by extending [57], where the goal is to test which of many normal populations have zero means. That study is motivated by the need to analyze DNA microarray data, for which the expression level for most genes is assumed to be zero Scott and Berger [57]; that view accords with disease surveillance, since one assumes that nearly all locations are not undergoing an epidemic outbreak. This leads to a model in which observations are drawn from a $N(\delta_i\mu_i, \sigma^2)$ distribution, $i = 1, \ldots, n$, where the δ_i's are i.i.d. Bernoulli random variables with parameter p. The objective is to determine which of the means are zero. A zero mean is equivalent to the corresponding δ_i being zero. Priors are placed on the μ_i and p. Their study found that the Bayesian procedure introduces a penalty that adjusts for multiple testing. In particular, the posterior probability of $\delta_i = 1$ "decreases as the number of 'noise' observations grows, so that the same observation is viewed as implying less evidence of a non-zero mean when more tests are simultaneously considered".

We now discuss possible models for the parameters $\mu_i(t)$, $\lambda_i(t)$ and $\delta_i(t)$. These adapt models used in the disease mapping literature [12,10,18,59,60], statistical image analysis [4,6,38], and models used in small area estimation that incorporate a cross-sectional and time series approach [17,30,51,54]. While we need to be fairly specific in modeling $\mu_i(t)$ and $\delta_i(t)$, we can be less specific, for reasons explained later, in modeling $\lambda_i(t)$.

2.1.1. Modeling $\mu_i(t)$

Let $\theta_i(t) = \log(\mu_i(t))$. We use the representation $\theta_i(t) = \boldsymbol{X}_i^T(t)\boldsymbol{\beta} + \xi_i(t)$, where the covariate vector $\boldsymbol{X}_i(t)$ for hospital i at time t is included to capture seasonality and day-of-week effects, as well as weather and holiday-related covariates that affect the baseline rate. In this representation of $\theta_i(t)$ one can include a fixed "offset" term to account for the known base population.

We consider a series of multivariate normal models for the $\xi_i(t)$'s:

Model 1: Let $\xi_i(t) = \psi_i + \gamma_t$. Here the spatio-temporal effect is expressed through an additive model. To explain spatial dependence, we employ a CAR model for $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_m)^T$ as in, for example, Besag [4] and Clayton and Kaldor [13]. One possible model for γ_t is an AR(1) model.

Model 2: We assume a joint normal distribution and specify certain conditional mean and variance functions. Let $E[\xi_i(t)\,|\,\xi_j(t), j\neq i, \boldsymbol{\xi}(s), s < t] = \rho_1\sum_j w_{ij}\xi_j(t) + \rho_2\xi_i(t-1)$, and $\mathrm{Var}[\xi_i(t)\,|\,\xi_j(t), j\neq i, \boldsymbol{\xi}(s), s < t] = \sigma^2$, where $\boldsymbol{\xi}(t) = (\xi_1(t), \ldots, \xi_m(t))^T$ and $\boldsymbol{W} = (w_{ij})$, the adjacency matrix correlating neighboring hospitals. As a first step, the elements of $\boldsymbol{W}$ are taken to be 1 if hospitals i and j are neighbors, and zero otherwise. In an alternative model, w_{ij} can depend on the geographical distance and other measures of proximity, but that requires much more computation. From the conditional mean and variance specification above, it is possible to get the mean and the variance of the joint multivariate distribution after appropriate modification of the argument in Rao [53, p. 208] or Carlin and Banerjee [10].

Model 3: Let $\xi_i(t) = v_i + h_i(t)$, where $\boldsymbol{v} = (v_1, \ldots, v_m)^T$, $\boldsymbol{h}_i = (h_i(1), \ldots, h_i(T))^T$, and $\boldsymbol{h} = (\boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_m^T)^T$. Here we put independent prior distributions on the random components $\boldsymbol{v}$ and $\boldsymbol{h}$. For a simple model, we can assume that components $v_1, \ldots, v_m$ which capture random effects due to the hospitals are independent and identically normally distributed as is often assumed in small area estimation [16,20,52,53]. However, since spatial dependence is expected in the hospital random effects, a better option would be to replace the i.i.d. assumption by a CAR model as developed for disease mapping [13,53]. While systematic trends in the time series variation in the longitudinal data can be explained through the covariate vector, the random vector $\boldsymbol{h}$ can be modeled as in Datta et al. [17] to account for dependence in the series of observed counts.

To complete a hierarchical Bayesian formulation, we need to specify appropriate priors, possibly diffuse, on $\boldsymbol{\beta}$ and parameters, termed hyperparameters, that appear in the distribution of $\xi_i(t)$. For example, as part of the first model, we assume independent Gaussian distributions for $\{\psi_i, i = 1, \ldots, m\}$ and $\{\gamma_t, t = 1, \ldots, T\}$. We specify appropriate priors on $\boldsymbol{\beta}$ and hyperparameters appearing in the CAR model for $\boldsymbol{\psi}$ [30] and an AR(1) model (see Ghosh et al. [30] for an application) or a random walk model (see Datta et al. [17,19] for applications) for the γ's.

2.1.2. Modeling $\lambda_i(t)$

We could model $\log(\lambda_i(t))$ in a similar fashion as described for $\log(\mu_i(t))$. One might fit a separate form of $\lambda_i(t)$ for each possible disease, taking account of its virulence, incubation period, and so forth; or one might fit a smoothed impulse function, to account for exposure to non-contagious pathogens, such as an anthrax release.

However, the reality is that in a syndromic surveillance system, the inference should not be sensitive to the shape of the epidemic curve. One just needs $\lambda_i(t)$ to be an increasing function over the first few days of the disease. At that point, the syndromic surveillance system signals, public health officials intervene, and the model for the shape of the epidemic curve becomes largely irrelevant as medical intervention and behavioral changes quickly invalidate any standard mathematical model.

For this reason, modeling $\lambda_i(t)$ is relatively easy, compared to other terms in the CAR model. We believe it is sufficient to take it as a linear function of time with moderate slope. Anything more dramatic than that will be quickly detected by the model, since the signal will seem very strong compared to the gradual increase that the model is tuned to discover.

2.1.3. Modeling $\delta_i(t)$

The δ_i are the basis for the spatial portion of the syndromic surveillance system. Since the $\delta_i(t)$'s are binary variables, as a first step we assume these are independent random variables that are 1 with probability $p_i(t)$. Then, we create dependence among the $\delta_i(t)$ through linear modeling of $\mathrm{logit}(p_i(t))$, by means of an increasing function of the number of neighboring hospitals in which a disease outbreak is present. Weights are likely to depend upon distance, but could reflect demographic information and commuter patterns, although the latter would probably require weights that are functions of time.

Much more generality is possible, and some of it may be useful. In particular, it would be natural to describe the joint distribution of the $\delta_i(t)$'s as a binary Markov random field. The field is defined by an undirected graph where the state of the node indicates whether the syndrome is absent or present. This could extend further to a k-state model in which the level of the state indicates the severity of the syndrome at a node. The neighborhood structure for the field is given by the graph.

For the syndromic surveillance system, each hospital is connected in a graph to nearby hospitals. The hospitals form a connected graph with nodes $\{v_1, \ldots, v_m\}$. Let $A = [a_{ij}]$ be the adjacency matrix of this network, such that $a_{ij} = 1$ if and only if there is an (undirected) edge between v_i and v_j. We assume that every hospital is connected to itself, so $a_{ii} = 1$ for $i = 1, \ldots, m$. In this network, the node-set $K = \{v_{i_1}, \ldots, v_{i_k}\}$ is a *clique* if $a_{ij} = 1$ for all $i, j \in K$, which one can think of as all the hospitals in a given city. Also, let $\mathscr{K}$ denote the set of all cliques in the network.

This network is extended to incorporate time as follows. Let $\{v_1(t), \ldots, v_m(t),\ t = 1, \ldots, T\}$ be a network with adjacency matrix given by

$$A(1, \ldots, T) = \begin{bmatrix} A & I_m & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ I_m & A & I_m & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & I_m & A & I_m & \cdots & \mathbf{0} \\ \vdots & & & & & \\ \mathbf{0} & \cdots & I_m & A & I_m & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & I_m & A & I_m \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & I_m & A \end{bmatrix},$$

where I_m is the identity matrix of size m. That is, the neighbors of node $v_i(t)$ are other nodes at time t, as well as itself in the immediate past and present (with directed edges), or $v_i(t-1)$ and $v_i(t+1)$. The cliques of this extended network are as follows. If $K = \{v_{i_1}, \ldots, v_{i_k}\}$ is a clique then $K(t) = \{v_{i_1}(t), \ldots, v_{i_k}(t)\}$ is also a clique, for $t = 1 \ldots, T$. Also, $\{v_i(t), v_i(t+1)\}$ lie within a clique, for $i = 1, \ldots, m,\ t = 1, \ldots, T-1$. Since the start of an epidemic occurs over a finite time horizon of duration h, we need only consider networks with adjacency matrix $A(t, \ldots, t+h)$.

The syndromic surveillance model assumes that there is a latent indicator variable $\delta_i(t)$ associated with each node. Let $\mathscr{D} = \{0, 1\}^{mT}$, that is, the set of all possible configurations of the latent indicators. Grimmett [34] showed that any joint distribution with positive probability on every point in $\mathscr{D}$ and the Markov random field property has the following probability mass function

$$p(\boldsymbol{\delta}(t),\ t = 1, \ldots, T) = \frac{1}{Z(\boldsymbol{\phi})} \exp\left[\sum_{t=1}^{T} \sum_{K \in \mathscr{K}} \phi_K(t) \prod_{i \in K} \delta_i(t) + \sum_{t=1}^{T-1} \sum_{i=1}^{m} \omega_i(t) \delta_i(t) \delta_i(t+1)\right],$$

where $Z(\boldsymbol{\phi})$ is a normalizing constant that depends upon node interactions. This expression builds on earlier work by Besag [4], Hammersley and Clifford [35], and Spitzer [58]. The joint distribution is referred to as a Gibbs measure for the Markov random field. Note that the summation over the cliques enables a much more tractable expression than if clique structure were not available.

Exploiting the clique structure, the Gibbs measure can be rewritten as follows. Let $\phi_{i_1 \cdots i_k}(t)$ denote $\phi_K(t)$ if $i_j \in K$, $j = 1, \ldots, k$. Then

$$\begin{aligned} p(\boldsymbol{\delta}(t),\ t = 1, \ldots, T) &= \frac{1}{Z(\boldsymbol{\phi})} \exp\Bigg[\sum_{t=1}^{T} \sum_{i=1}^{m} \phi_i(t) \delta_i(t) + \sum_{t=1}^{T} \sum_{\{i,j\} \in \mathscr{K}} \phi_{ij}(t) \delta_i(t) \delta_j(t) \\ &\quad + \sum_{t=1}^{T} \sum_{\{i,j,k\} \in \mathscr{K}} \phi_{ijk}(t) \delta_i(t) \delta_j(t) \delta_k(t) \\ &\quad + \cdots + \sum_{t=1}^{T-1} \sum_{i=1}^{m} \omega_i(t) \delta_i(t) \delta_i(t+1)\Bigg]. \end{aligned} \tag{2}$$

In this representation, $\phi_i(t)$ is the main effect for node i at time t, $\phi_{ij}(t)$ is the 2-way interaction between nodes i and j at time t, ϕ_{ijk} is the 3-way interaction among nodes i, j and k at time t, and so on; $\omega_i(t)$ represents the interaction between node i and itself at times t and $t+1$. The model can be simplified by assuming that the parameters do not change over time, i.e., $\omega_i(t) = \omega_i,\ \phi_i(t) = \phi_i,\ \phi_{ij}(t) = \phi_{ij}$, etc.

The formulation has analogies to complex systems of interacting components in reliability theory [26,31]. There the nodes indicate the state of a component. As noted in Gleaton and Lynch [31,32], the fundamental quantity for the component is not the hazard function but the log odds ratio. Here, the conditional log odds ratio $l_i(t_0)$ of $\delta_i(t_0)$, given all the other indicators at time t_0 (i.e., $\boldsymbol{\delta}_{(i)}(t_0)$) and all indicators for time $t = 1, \ldots, t_0 - 1$, is described as follows. For this, let $d_i(t_0)$ denote $(\boldsymbol{\delta}(t), t = 1, \ldots, t_0)$ with $\delta_i(t_0) = 1$; let $\bar{d}_i(t_0) = (\boldsymbol{\delta}(t), t = 1, \ldots, t_0)$ with $\delta_i(t_0) = 0$; and let $\boldsymbol{\delta}_{(i)}(t)$ denote $\boldsymbol{\delta}(t)$ without the ith element. Then

$$\begin{aligned} l_i(t_0) &= \log\left[\frac{P(\delta_i(t_0) = 1 | \boldsymbol{\delta}_{(i)}(t_0), \boldsymbol{\delta}(t),\ t = 1, \ldots, t_0 - 1)}{P(\delta_i(t_0) = 0 | \boldsymbol{\delta}_{(i)}(t_0), \boldsymbol{\delta}(t),\ t = 1, \ldots, t_0 - 1)}\right] \\ &= \log\left[\frac{p(d_i(t_0))/p(\boldsymbol{\delta}_{(i)}(t_0), \boldsymbol{\delta}(t),\ t = 1, \ldots, t_0 - 1)}{p(\bar{d}_i(t_0))/p(\boldsymbol{\delta}_{(i)}(t_0), \boldsymbol{\delta}(t),\ t = 1, \ldots, t_0 - 1)}\right] \\ &= \phi_i + \sum_{j:\{i,j\} \in \mathscr{K}} \phi_{ij} \delta_j(t_0) + \sum_{j,k:\{i,j,k\} \in \mathscr{K}} \phi_{ijk} \delta_j(t_0) \delta_k(t_0) + \cdots + \omega_i \delta_i(t_0 - 1). \end{aligned}$$

Gibbs sampling may be used to generate a sample from this joint distribution based only on these conditional log odds ratios. To carry out fully Bayesian inference we assign a prior distribution on the hyperparameter $\boldsymbol{\phi}$. Gibbs sampling then requires the full conditional distribution of $\boldsymbol{\phi}$ [29], which is proportional to the kernel of $\boldsymbol{\phi}$ determined by the normalizing constant $Z(\boldsymbol{\phi})$ above and the prior distribution on $\boldsymbol{\phi}$. Evaluation of the normalizing constant $Z(\boldsymbol{\phi})$ for a general Markov random field, without which the posterior is intractable, is challenging [37].

One practical simplification is to use empirical Bayes, and substitute the maximum likelihood estimate of $\boldsymbol{\phi}$), although even this is not straightforward [39]. Alternatively, things can be kept at a manageable level by means of the hierarchical network structure explained in Section 2.1.6. Because of the hierarchical structure, the number of cliques in a Markov random field network will be substantially fewer, and for such sparse matrix cases we can compute the normalizing constant $Z(\phi)$ symbolically. In addition, it is reasonable to assume a finite discrete prior distribution on the ϕ's for each node of the network, which greatly reduces the computing time. Depending upon the scale of the application, one could consider richer classes of prior distributions for $\boldsymbol{\phi}$.

2.1.4. Choice of $\boldsymbol{\phi}$

For the Bayesian model one is interested in the posterior of the vectors of latent variables $\boldsymbol{\delta}$ given the vector of counts $\boldsymbol{y}$. In the construction of the prior needed to get the posterior there are a number of practical issues. Since the latent variables are considered alarms or signals, we should not raise any alarms when there is no data. Hence, the prior probability of all the latent variables being zero should be high or close to one. This can be accomplished by putting substantial prior weight on large negative main effects. However, if the main effects are chosen to be $-\infty$, there would be no power to detect an anomaly.

The interaction terms tells us how strong the dependencies are between nodes. If all interactions are zero, then the nodes would be independent. For disease spread, a node being "on" (i.e., 1) would increase the likelihood that its neighbors are "on" or will soon be turned "on". Hence, the 2-way interactions should be positive. How large they are will depend on things such as the volume of car or air traffic between nodes, distance, amount of patient exchange, and so forth.

2.1.5. Posterior distribution

Next, the posterior distribution of $\boldsymbol{\delta}(t)$ given the observed counts $\boldsymbol{y}(t)$ is described. This is the object of primary interest. If the posterior probability that a component of this vector is equal

Please cite this article in press as: D. Banks et al., Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice, Informat. Fusion (2009), doi:10.1016/j.inffus.2009.10.005

to 1 is large, then the corresponding region appears to be undergoing an epidemic outbreak. How large this posterior probability should be in order to trigger attention is a matter of decision theory, and the probabilistic output directly supports such management; one calculates the expected loss from action and inaction, and then makes the best decision.

Our inference on $\boldsymbol{\delta}(t)$ is driven by the posterior probabilities of $\boldsymbol{\delta}(t)$, computed from Gibbs samples. For Gibbs sampling we need the full conditional distribution of $\boldsymbol{\delta}(t)$. For this, note that:

$$\begin{aligned}
&p(\boldsymbol{y}(t)|\boldsymbol{\delta}(t),\, t=1,\dots,T)\\
&\quad=\prod_{i=1}^{m}\frac{(\mu_i(t)+\delta_i(t)\lambda_i(t))^{y_i(t)}e^{-\mu_i(t)-\delta_i(t)\lambda_i(t)}}{y_i(t)!}\\
&\quad=\frac{\exp\left[\sum_{i=1}^{m}y_i(t)\log(\mu_i(t)+\delta_i(t)\lambda_i(t))-\mu_i(t)-\delta_i(t)\lambda_i(t)\right]}{\prod_{i=1}^{m}y_i(t)!}\\
&\quad=\frac{\exp\left[\sum_{i=1}^{m}y_i(t)\left(\delta_i(t)\log(1+\frac{\lambda_i(t)}{\mu_i(t)})+\log(\mu_i(t))\right)-\mu_i(t)-\delta_i(t)\lambda_i(t)\right]}{\prod_{i=1}^{m}y_i(t)!}\\
&\quad=c(\boldsymbol{y}(t),\boldsymbol{\mu}(t))\exp\left[\sum_{i=1}^{m}\left(y_i(t)\log(1+\frac{\lambda_i(t)}{\mu_i(t)})-\lambda_i(t)\right)\delta_i(t)\right].
\end{aligned}$$

Using this last expression and (2), and assuming that the counts are conditionally independent given the indicator variables, the distribution of $(\boldsymbol{\delta}(t),\ t=1,\dots,T)$ given $(\boldsymbol{y}(t),\ t=1,\dots,T)$ is

$$\begin{aligned}
&p(\boldsymbol{\delta}(t),\ t=1,\dots,T|\boldsymbol{y}(t),\ t=1\dots,T)\propto p(\boldsymbol{\delta}(t),\boldsymbol{y}(t),\ t=1,\dots,T)\\
&\quad=p(\boldsymbol{\delta}(t),\ t=1,\dots,T)\prod_{t=1}^{T}p(\boldsymbol{y}(t)|\boldsymbol{\delta}(t),\ t=1,\dots,T)\\
&\quad=c(\boldsymbol{y}(t),\boldsymbol{\mu}(t),\ t=1,\dots,T)\\
&\qquad\times\exp\left[\sum_{t=1}^{T}\sum_{i=1}^{m}\left(y_i(t)\log\left(1+\frac{\lambda_i(t)}{\mu_i(t)}\right)-\lambda_i(t)\right)\delta_i(t)\right]\\
&\qquad\times\frac{1}{Z(\boldsymbol{\phi})}\exp\left[\sum_{t=1}^{T}\sum_{i=1}^{m}\phi_i\delta_i(t)+\sum_{t=1}^{T}\sum_{\{i,j\}\in\mathscr{K}}\phi_{ij}\delta_i(t)\delta_j(t)\right.\\
&\qquad\left.+\sum_{t=1}^{T}\sum_{\{i,j,k\}\in\mathscr{K}}\phi_{ijk}\delta_i(t)\delta_j(t)\delta_k(t)+\cdots+\sum_{t=1}^{T-1}\sum_{i=1}^{m}\omega_i\delta_i(t)\delta_i(t+1)\right]\\
&\qquad\propto\exp\left[\sum_{t=1}^{T}\sum_{i=1}^{m}\left(\phi_i+y_i(t)\log\left(1+\frac{\lambda_i(t)}{\mu_i(t)}\right)-\lambda_i(t)\right)\delta_i(t)\right.\\
&\qquad+\sum_{t=1}^{T}\sum_{\{i,j\}\in\mathscr{K}}\phi_{ij}\delta_i(t)\delta_j(t)+\sum_{t=1}^{T}\sum_{\{i,j,k\}\in\mathscr{K}}\phi_{ijk}\delta_i(t)\delta_j(t)\delta_k(t)\\
&\qquad\left.+\cdots+\sum_{t=1}^{T-1}\sum_{i=1}^{m}\omega_i\delta_i(t)\delta_i(t+1)\right];
\end{aligned}$$

that is, the posterior distribution has the same form as the prior, with the same interaction parameters, but with main effect parameters $\phi_i'(t)=\phi_i+y_i(t)\log(1+\lambda_i(t)/\mu_i(t))-\quad\lambda_i(t),\ i=1,\dots,m,\ t=1,\dots,T$.

Notice that $\phi_i'(t)$ increases as the count $y_i(t)$ increases. Also notice that $\phi_i'(t)$ is larger than ϕ_i only if $y_i(t)>\lambda_i(t)/\log[1+\lambda_i(t)/\mu_i(t)]$.

2.1.6. Hierarchical structure of the network

Simplification occurs in the Gibbs measure by considering a hierarchical structure for the graph. For example, consider a hub-feeder network.

This network can be used recursively to build large Markov random fields where there are dependencies between nodes but many nodes are conditionally independent given their predecessors in the hierarchical structure of the network. In particular, a geographical region can be divided into subregions. Major nodes in the various subregions will be designated level 1 hubs while the other nodes are now subdivided into further subregions. Major nodes in these subregions are designated as level 2 hubs but are feeders into the level 1 hubs. Continuing in this fashion we have a series of level 1 to level k hubs where at level k the nodes that have not been designated as hubs are feeders into the various k-level hubs. The model we consider here is one in which the non-hub nodes are feeders into only the level-k hubs. This model is illustrated in Fig. 1.

The advantage of this model is that it is recursive, which simplifies the dependency structure of the joint distribution (the Gibbs measure) for the Markov random field. Thus one can model each level in the same way, using a common hyperprior and block-diagonal matrices for the covariance structure.

Alternative models are possible. The most interesting specification uses dynamic networks. These are more realistic but also more complicated. The ideas in Banks et al. [3] provide an entry point to this line of investigations.

2.2. Extending the basic model

The basic model does not yet satisfy all of the criteria for a surveillance system that are listed in Section 1. In particular, two significant extensions are needed to accommodate multivariate counts and data quality issues. These extensions will be complex in practice, but we believe that they are technically feasible and can be achieved incrementally.

2.2.1. Multivariate counts

Hospitals record counts for symptoms such as fever, cough and rashes, all of which are symptoms of more than one disease. When there is an upswing in a particular disease, these separate counts will increase in a fashion that reflects the dependency among the symptoms characterizing the syndrome that is associated with the disease. We model this scenario by the hierarchical Bayes model described below. For simplicity, we consider a bivariate setup where counts of two symptoms are recoded and both symptoms are indicative of the same disease.

Model:

- Conditional on $\mu_{i1}(t),\ \mu_{i2}(t),\ \delta_i(t),\lambda_{i1}(t),$ and $\lambda_{i2}(t)$, the $Y_{i1}(t),\ Y_{i2}(t)$ are independent Poisson random variables with means $\mu_{i1}(t)+\delta_i(t)\lambda_{i1}(t)$ and $\mu_{i2}(t)+\delta_i(t)\lambda_{i2}(t)$, respectively;
- Each component mean $\mu_{i1}(t),\ \mu_{i2}(t)$ is modeled in the same way as described for $\mu_i(t)$ in Section 2.1;
- $\delta_i(t)$ is modeled as in Section 2.1;
- Each of $\lambda_{i1}(t),\ \lambda_{i2}(t)$ is modeled as described for $\lambda_i(t)$ in Section 2.1.

One completes the model specification by suggesting prior distributions on the hyperparameters that appear in the distributions of $\mu_{i1}(t),\ \mu_{i2}(t),\ \lambda_{i1}(t),\ \lambda_{i2}(t)$ and $\delta_i(t)$. The inference is based on the posterior probability of $\delta_i(t)=1$.

In practice, numerous syndromes, say $A_{i1},A_{i2},\dots,A_{ig}$, are monitored and counts of these syndromes, $Y_{i1}(t),Y_{i2}(t),\dots,Y_{ig}(t)$, are recorded at hospital i. For simplicity we assume that these syndromes are disjoint from one another. (If not, the syndromes can be combined and labeled to account for this.) We assume that these counts are independent Poisson random variables with means $\mu_{ij}(t)+\delta_{ij}(t)\lambda_{ij}(t)$.

2.2.2. Data quality

In any real-world application, the quality of the data reported from the hospitals will be suspect. We describe a two-tier system of analysis that maps known characterizations of data quality into final estimates of uncertainties.

Please cite this article in press as: D. Banks et al., Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice, Informat. Fusion (2009), doi:10.1016/j.inffus.2009.10.005

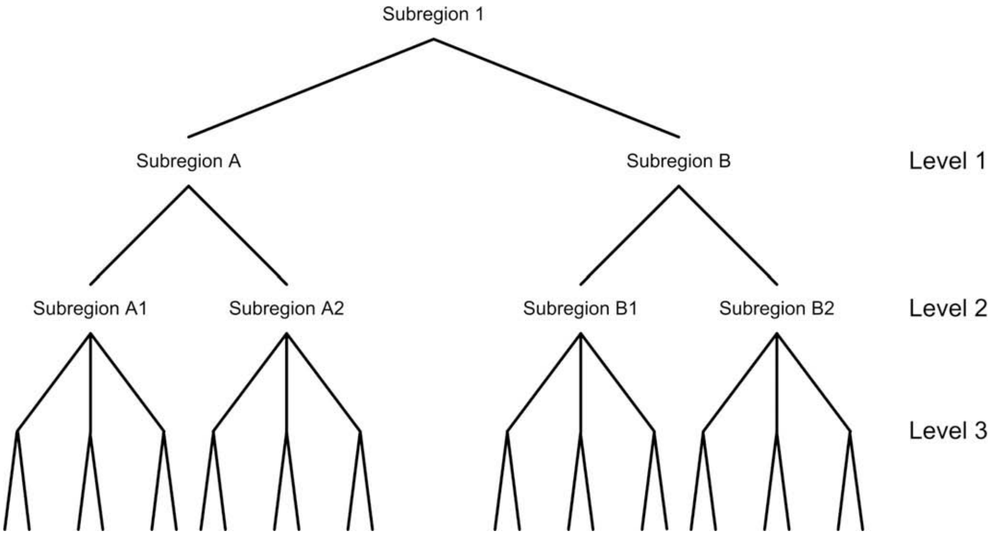

Fig. 1. A hub-feeder network that permits a relatively simple hierarchical model for syndromic surveillance data.

The first tier treats the data as if they were correct, and performs the MCMC updating to find the posterior probability that one or more hospitals is experiencing a public health emergency.

When a hospital or a set of hospitals appears to signal, the second tier of analysis is used. This tier incorporates, for example, a heavy-tailed error model for the data, built to reflect data entry error, misdiagnosis or over-diagnosis, conflated symptom categories, and so forth. The point of the second tier analysis is to determine whether there is a single report (or small number of reports) that account for nearly all the signal. If that is the case, as opposed to seeing an alarm that is robustly dependent upon separate reports, then the decision-maker should be more conservative in managing the response.

This kind of assessment might be relatively informal; perhaps a phone call with the hospital administrator would obviate the need for mathematically athletic analysis. But we note that, if full calculation is done, it could be done locally, and would not pose the scalability issues that arise in the more elaborate model that was previously discussed. Tools and approaches outlined in Karr et al. [40] may also be used to incorporate external and domain knowledge of data quality into the quality assessment.

2.2.3. Other extensions

Additional extensions necessary for practical applications to disease surveillance using hospital reports include:

- Use of covariate information and Bayesian updating to handle slow drift in the local baseline rates, such as might be due to immigration or pollution trends.
- Use of contextual information to downweight spurious signals from innocent causes, as might happen if one hospital closes and thus the reporting rates at neighboring hospitals jump.
- Modeling of abrupt non-contagious change (anthrax aerosol) versus slow/fast spread of a contagious disease—this will affect how quickly the signal emerges.
- Handling of extra-Poisson variation, which would naturally arise if the background rate were spiked by occasional situations in which a large family all became sick, or a group of co-workers were exposed to an unusually infectious strain; see Dey and Ravishanker [21] and Martin [45] for Bayesian treatments of overdispersion in generalized linear models.
- Multiple geographies. While it is sensible as an entry point to the research, as in Section 2.1.6, to posit a graph structure for hospitals that reflects geographical proximity, this may fail to capture important dependencies in the data. For instance, a workplace-induced outbreak may be most visible in the hospitals associated with the employer's health benefits plan, which are distributed rather than adjacent in space.
- At this point, it is simply unclear to what extent data confidentiality issues will impact the use of syndromic surveillance systems. While there is clear social benefit from the early detection of disease outbreaks, there are counter-balancing considerations of both individual and organizational privacy that may influence the kinds of analysis that can be performed.

In order to actually fit a model of this kind in practice, model-checking steps would be needed to identify which systematic inaccuracies of this kind that require further adjustments. Some of these checks are indicated in the following section.

3. Drug abuse surveillance: an application

Data on abuse of prescription opioid analgesics offers an opportunity for comparing syndromic surveillance methods [14]. Prescription drug abuse has similarities to infectious disease due to the inherent geographical effects [11], multiple reporting systems which vary in coverage and data quality, and the fact that it is a major concern for public health management. Prescription opioid analgesic abuse costs the US an estimated $8.6 billion in 2001 due to increased health care, workplace, and criminal justice costs [7].

This section uses data from the Researched Abuse, Diversion, and Addiction-Related Surveillance (RADARS®) system. In particular, we use the Opioid Treatment Programs (OTP) study, which collects questionnaires on a quarterly basis from abusers enrolled in Methadone Maintenance Treatment Programs (MMTPs) and thus captures a sentinel population of sophisticated abusers. We further focus on the opioid analgesic OxyContin® (oxycodone HCI, controlled-release) Tablets, since it has been the target of abuse over several years [9]. Using this data, we fit a CAR model of the kind described in Section 2, and then compare the performance of that model to two other methods used in syndromic surveillance: CUSUM charts and paired-difference studies.

The comparison is based on the power of the three procedures for detecting change (a decrease) in abuse rates. The power simulation uses the OTP data as the baseline (to ensure realistic complexity in the correlation structure of the reporting centers and heteroscedasticity in the center rates), and fits a generative CAR model to these data. That model is then used to simulate OTP data

Please cite this article in press as: D. Banks et al., Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice, Informat. Fusion (2009), doi:10.1016/j.inffus.2009.10.005

Table 1
Percentile points of the posterior distributions on the fixed-effect terms in the CAR model.

	2.5%	50%	97.5%
Intercept	−2.2	−1.7	−1.2
Gender	0.31	0.46	0.58
Race	0.89	1.15	1.47
Age	−0.051	−0.043	−0.034
CAR precision	0.19	0.36	0.64

corresponding to prescribed percentage reductions in the overall abuse rate. The methods are compared in terms of their power to detect small changes.

3.1. A CAR model for drug abuse surveillance

In this application we modify the model in (1) to reflect the emphasis upon changing proportions of opioid drug abusers at the MMTPs and the covariates available in the OTP data. Specifically, we use a logit (log odds) model instead of the Poisson response model, which stays within the Generalized Linear Model family ([46]); this linearizes the relationship between the proportion of MMTP clients reporting opioid abuse and the covariate terms. Among the covariates available from the OTP, our model incorporates age, gender, and race.

CAR models include spatial dependence via a neighborhood structure, so that reporting units that are near each other have correlated responses. In our application, there are known hot spots of opioid drug abuse in Appalachia and Maine, which are fairly stable over the timespan of OTP data (the third quarter of 2004 to the second quarter of 2006). In this analysis, we aggregated the geographic information to the state level; it is coarse, but our results show that it is effective.

Inference is done through Markov chain Monte Carlo [60]. Despite the relative simplicity of this application compared to disease surveillance, there are still computational challenges. Using coarse geography enables our (unoptimized) code to execute within a few hours. Finer resolution could require several days, unless one took systematic advantage of relatively sophisticated methods to accelerate the calculation. For similar reasons we do not attempt to fit a full spatio-temporal model, as described in Section 2. However, in the following subsection on power comparisons we incorporate time dependence through a fixed-effect term.

The CAR model used in this example is:

$$\begin{aligned} &Y_{ik}(t) \sim \text{Bernoulli}(p_{ik}(t)) \\ &\text{logit}(p_{ik}(t)) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + b_k \end{aligned} \tag{3}$$

where $Y_{ik}(t)$ is the outcome for individual i living in state k (i.e., it is 1 if the individual has used an oxycodone product in the past month and 0 otherwise). The covariates describe gender, race, and age, respectively, where race has been dichotomized to White or non-White and age has been broken down into 17 ordered categories. Flat priors were used for all coefficient parameters. The spatial random effects for US states in the CAR model are assumed to have prior distribution

$$\pi(b_1, \ldots, b_K) \propto \exp\left(-\frac{\tau}{2} \sum_{i \neq j} w_{ij} (b_i - b_j)^2\right)$$

where τ is the precision parameter, given a $\Gamma(1,1)$ prior, and w_{ij} is obtained from the matrix of binary weights that indicate whether two states share a border. We examined the use of a quadratic term in age, but it was not significant and thus was excluded from this model.

Initial work also fit linear and quadratic terms in time, to discover trends in the abuse rate. The coefficients on these terms were not (Bayesianly) significant in any of the models (neither the model with both terms, nor the models with each term separately). For this reason, the trend component was excluded in fitting the model described above. However, this same fitted model, but with a piecewise linear trend, is used in the next subsection to generate simulated data with decreasing abuse rates over time, so that three different surveillance methods may be compared with respect to their power to detect change.

Table 1 shows posterior credible regions (the Bayesian analogue of confidence intervals) on the coefficients for gender, race, and

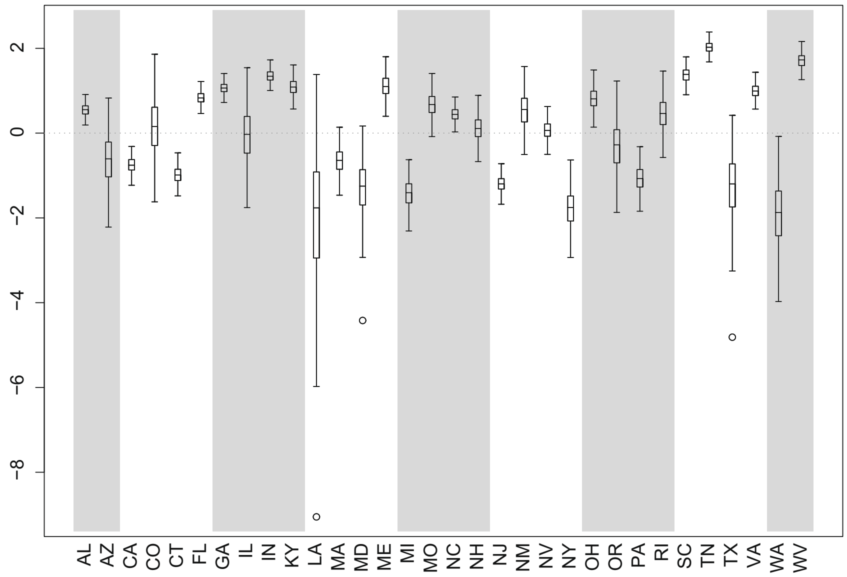

Fig. 2. A display of the 95% credible regions on the location effect. If the span of the line segment for a state straddles zero on the y-axis, then there is more than a 5% chance that that state has no effect on the abuse rate, after accounting for gender, race and age. The central boxes contain the middle 50% of the probability mass for the magnitude of the state effect, and the midline within the box is the point estimate of the magnitude of the state effect.

Please cite this article in press as: D. Banks et al., Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice, Informat. Fusion (2009), doi:10.1016/j.inffus.2009.10.005

age. The estimate for the gender coefficient is 0.46 and the credible region excludes 0, so women are less likely to abuse opioid drugs. Also, the coefficient of 1.15 for race means that Whites are more likely to abuse opioids than non-Whites. And the age coefficient is negative, so older people are less likely to abuse.

The location terms b_k are random state effects, and correlated. Since data are aggregated by state, the correlation structure is coarse: two states directly interact if they are contiguous; otherwise, states are conditionally independent given their neighbors. (Alaska is the only state in the OTP data that has no neighbors.) The model for the correlation is multivariate Gaussian with unknown but common correlation for states that share boundaries.

Fig. 2 shows estimates of the state effects in the CAR model. Three states with large positive effects were Tennessee, West Virginia, and Virginia. This accords with previous reports of high opioid abuse rates in Appalachia. Some states, such as California and Connecticut, have lower than expected rates of opioid abuse. The wide interval for Louisiana surely reflects uncertainty in the data due to Hurricane Katrina; Texas also has a wide interval, perhaps from a combination of evacuee spillover and the impact of Hurricane Rita.

The CAR model used in this analysis is substantially simpler than the one needed for general disease surveillance. In particular, the geography is coarse and there is no temporal neighborhood structure. However, it succeeds in discovering known patterns of opioid drug abuse and known relationships with such covariates as age, gender, and race.

3.2. Power comparisons

We now assess power in terms of change detection for a one-sided alternative which specifies that the drug abuse rate has decreased over time. This case is simpler than two-sided alternatives, reflects federal interest in measuring the effectiveness of drug prevention programs, and the results extend directly to the symmetric hypothesis that drug abuse has increased.

Using the data from the OTP component of RADARS®, we explore power by simulation. For each of three different surveillance procedures, we examine power as a function of simulated levels of abuse reduction. The simulations were performed by bootstrapping [27] from the original data sets, after adjustment to achieve specified reduction levels. Specifically,

1. a generative model is assumed and fit for the data;
2. simulations with artificial signal are generated from this model; and
3. the CAR, CUSUM, and paired-difference techniques are applied to the resulting data.

The generative model for the simulated OTP data is a CAR model that includes covariates as estimated previously, but modified to create user-specified levels of decrease in drug abuse, thereby enabling power comparisons for known levels of signal.

The three surveillance procedures that we compare are:

- A paired-difference two-sample test, which looks for differences in abuse rates over time at each reporting MMTP site.
- A sequential process control procedure, using the CUSUM chart, similar to that used by the CDC [55].
- A CAR model which incorporates covariates as well as a geographic correlation.

These methods are compared with respect to their power in detecting simulated signal using historical abuse data.

3.2.1. Paired-difference and CUSUM tests

A two-sample test looks for a change between two time points. We use the traditional one-sided test for a difference in binomial proportions. The test statistic is:

$$z = (\hat{p}_1 - \hat{p}_2) / \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where $\hat{p}_1$ is the observed proportion of opioid abusers in the previous quarter and $\hat{p}_2$ is the observed proportion in the current quarter. This test statistic is referred to a standard normal table.

When the same sites report each time period, the two-sample test can be improved by pairing the previous and current reports. Thus, if there are k such sites, then one can perform the two-sample test separately at each, and pool the resulting P-values according to Fisher's rule [28]. Let p_i, for $i = 1, \ldots, k$, be the P-value for site i; then

$$\chi^2_{2k} = -2 \sum_{i=1}^{k} \ln p_i$$

which is referred to a chi-squared table. Pairing allows many small reductions to be pooled to give stronger evidence of an overall pattern of reduction. Repeated use requires adjustment for multiple testing. To indicate geographical variation in the reductions, one can map P-values by region.

Control charts take a different approach to syndromic surveillance. To decide whether a succession of observations has drifted away from a baseline value, a CUSUM chart plots the sum of the differences between the previous quarters' proportions and a baseline proportion. As described in [48], when this sum falls below a lower control line, the result is statistically significant.

CUSUM charts assume that the baseline is fixed and known. This may sometimes be reasonable in manufacturing, but for drug surveillance, we do not know the baseline abuse rate; we can only estimate this, with uncertainty, from historical data. The assumption of no trend in this historical data is critical.

3.2.2. Simulation procedures

The simulations in this study used data that were generated using the model in (3). The data in the first year was simulated using model parameters estimated from the ten quarters of OTP data available in RADARS®. The data in the second year was produced by multiplying those means by the appropriate fraction to produce, on average, a linear decline in abuse during that year, so that a user-specified reduction was achieved at the end of the year. The data in the third year was simulated using the parameter values for the last quarter of the second year. The baseline parameters used in these simulations were drawn from the posterior distributions for the parameters estimated previously from the CAR model, shown in Table 1. We focus on power for two significance levels: $\alpha = .05$ and $\alpha = .05^2 = .0025$. These bracket loose and stringent levels for Type I error. Each plotted point is based upon 200 simulations with a specific, simulated decrease in opioid abuse from the historical record for OTP.

3.2.3. Results of the power analyses

Fig. 3 shows the estimated power of the paired-difference two-sample test using OTP data with Fisher's test for change at MMTP clinics that appear in both time periods. This "blocking" of an MMTP with itself automatically controls for many biases and reduces the variance in comparisons.

For control charts, one cannot plot power for all possible values of reduction (percentage decrease). So Fig. 4 plots the probability of rejection for four different reduction levels: 20%, 15%, 10% and 5%, reading the curves from left to right. Extensive pre-simulation runs were made to determine the lower control line values for these charts; this was not trivial, and creating CUSUM charts that attain specific false positive rates for observed baseline data is a drawback to this surveillance approach. A CUSUM run was considered to signal if, at any of the 12 quarters of simulated data, the cumulative sum fell below the lower control line.

Interpreting Fig. 4 requires some care. Note that the two-sample tests compare a year's worth of initial data to a year's worth of final data (and ignores the transitional second year), whereas the CUSUM test employs the original OTP data to establish the lower control limit and then uses 3 years of data, with all drift occurring in the second year (i.e., this is a piecewise linear trend). Looking at the power in the last quarter gives a basis for comparison, but recall that control charts do not adjust for multiple testing and that the CUSUM chart is exploiting partial signal available in the second year.

To test for a drop in abuse rate, the CAR model in Eq. (3) is modified to include a term for time. If the time coefficient is significantly less than zero, this indicates one-sided change (reduction). The magnitude of the effect can be estimated from the coefficient.

The CAR model was fit using Gibbs sampling run through WinBugs with an *R* interface. The OTP data takes a relatively long time to run (about 10 min per simulation run) so the power is only calculated for reductions of 1%, 2%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, and 12%, as shown in Fig. 5.

For OTP data, on a 3-year horizon, the CUSUM appears more powerful than the CAR test, which is more powerful than the two-sample test. But the CUSUM power is misleading—it is not di-

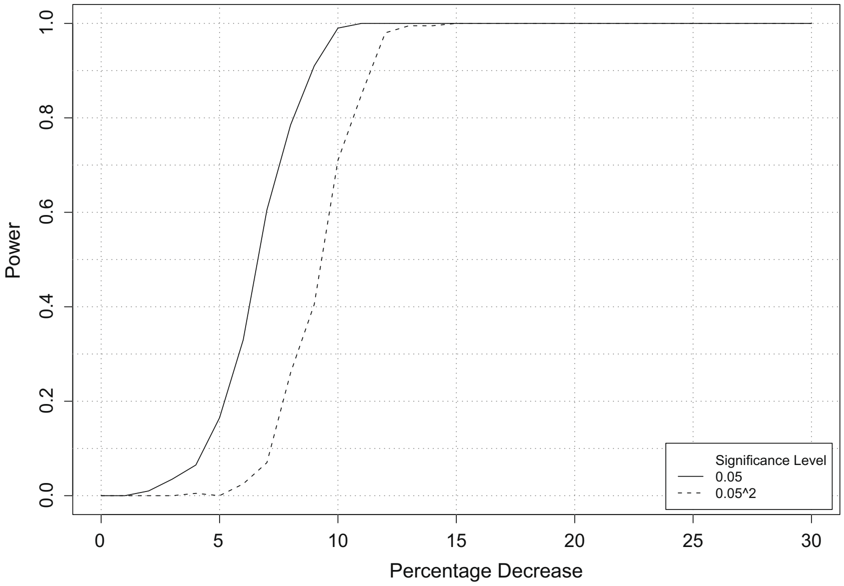

Fig. 3. Power curves for a two-sample OTP test. The most recent four quarters are combined to give the pre-sample abuse level.

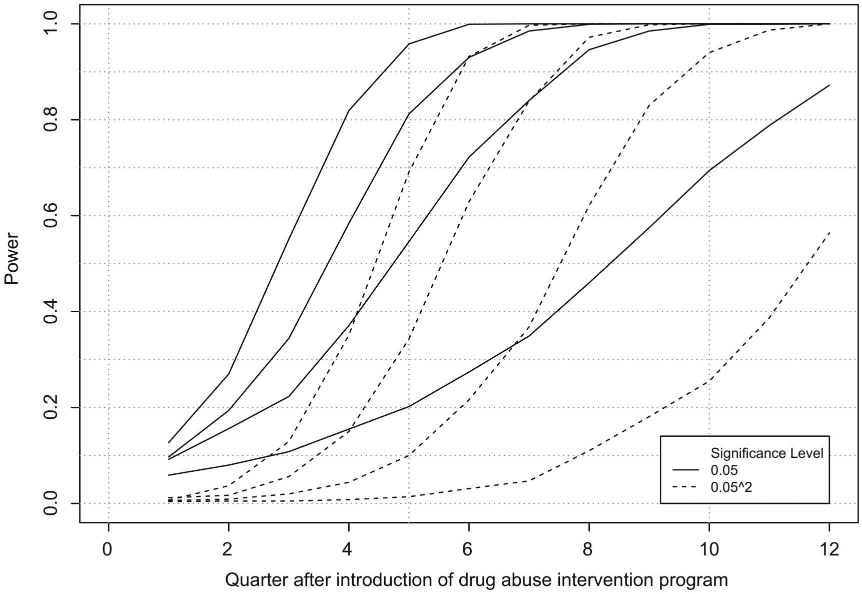

Fig. 4. Power curves for CUSUM testing with OTP data. The solid (broken) lines correspond to .05 (.0025) level tests. The lines, reading from the bottom up, correspond to 5%, 10%, 15% and 20% reductions in abuse rate, pro-rated over 3 years.

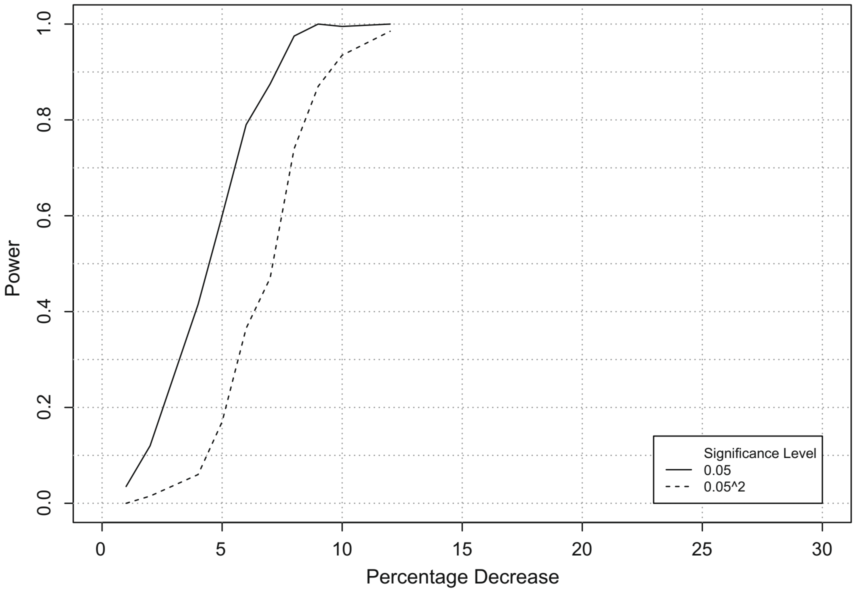

Fig. 5. Power curves for a CAR model test of abuse reduction using OTP data. The solid line is for an alpha level of .05; the dashed line is for an alpha level of .0025.

Table 2
Estimated power for the Two-Sample test, the CUSUM test, and the CAR test, for different levels of reduction in the OTP abuse rate.

Percent reduction	Two-sample		CUSUM		CAR	
	$\alpha = .05$	$\alpha = .0025$	$\alpha = .05$	$\alpha = .0025$	$\alpha = .05$	$\alpha = .0025$
5	.16	.02	.87	.55	.52	.16
10	.98	.67	.99	.99	.98	.93
15	.99	.99	.99	.99	.99	.99
20	.99	.99	.99	.99	.99	.99

rectly comparable with the other procedures. The 3-year time frame gives it a larger effective sample size than the CAR or two-sample tests; both of these make a single test, comparing the historical data in the first year to the simulated data in the third year, instead of the 12 tests implicit in the CUSUM chart. With these caveats, Table 2 summarizes the estimated power of the three procedures at the two different α-levels for the simulated data.

There are other comparison issues besides the power. The two-sample test has the desirable property of blocking on the reporting site, which controls for a potentially important source of variability. The CUSUM test requires a very precise estimate of the baseline, and there is considerable difficulty in accounting for multiple testing. And the CAR test allows investigation of regional and covariate effects.

4. Conclusions

This paper has developed a high-end Bayesian approach to the problem of syndromic surveillance, and illustrated some of the practical issues by applying the strategy to data on drug abuse surveillance. The results support the technical feasibility of the approach, with caveats as indicated below. But a larger issue is political feasibility.

Bayesian methods are unpopular with federal agencies. There is a widespread misperception that the use of prior information undercuts the scientific objectivity of the analysis—and it is true that a partisan analyst could select a prejudicial prior, but that analysis would not withstand the brief scrutiny of the professional community. On the other hand, Bayesian methods provide many solid advantages: they follow from sensible axioms, unlike the hodgepodge of ad hoc frequentist rules; they build on what one has learned, rather than approaching each question from the expensive posture of amnesia; and they provide clear probability statements, rather than the clumsy locutions that define significance probabilities and confidence intervals. It is beyond the scope of our paper to revisit this debate, but we emphasize that the statistical world has changed since current administrators took their last course in basic statistics, and Bayesian methods are now entirely uncontroversial and almost universally preferred.

But Bayesian methods require honest work. The application to drug abuse surveillance shows that analysis on the scale needed for disease monitoring requires serious computational resources. To accommodate the entire US hospital system, our methods must be able to scale to at least 7500 reporting units. The direct strategy is to partition the graph of hospitals into sets of cliques. Updating within a clique may be done in alternation with updating across cliques. The schedule for the alternation is key to success, but there are general heuristics (e.g., rapid alternation at first, then slowing over time) which are useful. Since multiple MCMC chains can be started on different processors, a great deal of scalability can be achieved by parallel processing. This has the disadvantage that burn-in time increases linearly with the number of processors. On the other hand, recent solutions should provide very accurate approximations to the current posterior, and researchers are developing smart ways to take advantage of this.

Besides computation, modeling requires care. The parameters of the prior distribution described in (2) were chosen to be fixed (see Section 2.1.4). However, in practice these are unknown and a source of uncertainty. A full hierarchical Bayesian approach would put priors on these parameters. However, the normalizing constant $Z(\boldsymbol{\phi})$ can be intractable, since putting a prior distribution on $\boldsymbol{\phi}$ requires knowledge of $Z(\boldsymbol{\phi})$, which in turn requires evaluation of the numerator of (2) for all 2^{mT} possible values of $\delta(t)$, $t = 1, \ldots, T$. In disease surveillance, we believe that a practical solution can be built by assuming that all nodes within a city or subregion have the same main effect and interaction parameters. This multiplicity in the parameters makes it easier to calculate an expression for the normalizing constant.

A positive point for the Bayesian approach is that decision theory approach can be built into the syndromic surveillance system.

Please cite this article in press as: D. Banks et al., Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice, Informat. Fusion (2009), doi:10.1016/j.inffus.2009.10.005

A basic requirement for a useful surveillance system is that its false alarm rate should be low and commensurate with the resources available for post-alarm investigation. Unlike the scan statistic and other frequentist methods, in the Bayesian approach it is possible to explicitly incorporate the resource matrix for all the hospitals and/or cities in the decision process. While the consequence of a false alarm is not the same for all the locations, the available resources for investigation or follow up of an alarm may not also be the same. We recommend a decision-theoretic formulation following Scott and Berger [57], Duncan [25] and Waller and Duncan [61].

In short, the Bayesian approach has many desirable properties, in terms of power, interpretability, and practicality. There are some technical issues that need resolution in order to monitor the large datastreams expected in disease surveillance, but there are strategies to address these and none seem insurmountable. There larger barriers are institutional—those will require imaginative leadership.

Acknowledgments

The authors thank the Statistical and Applied Mathematical Sciences Institute for organizing the research program that initiated this research. Additionally, Gauri Datta received support from NSF Grant MSPF-07G-082, and Jim Lynch received support from NSF Grants DMS-0243594 and DMS-0805809.

References

[1] R. Assuncao, M. Costa, A. Tavares, S. Ferreira, Fast detection of arbitrarily shaped disease clusters, Statistics in Medicine 25 (2006) 723–742.
[2] S. Banerjee, B.P. Carlin, A. Gelfand, Hierarchical Modeling and Analysis for Spatial Data, Chapman & Hall/CRC, Boca Raton, FL, 2004.
[3] H.T. Banks, A.F. Karr, H.K. Nguyen, J.R. Samuels Jr., Sensitivity to noise variance in a social network dynamics model, Quarterly of Applied Mathematics 66 (2008) 233–247.
[4] J. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussion), Journal of the Royal Statistical Society, Series B 36 (1974) 192–236.
[5] J. Besag, Statistical analysis of non-lattice data, The Statistician 24 (1975) 179–195.
[6] J. Besag, On the statistical analysis of dirty pictures (with discussion), Journal of the Royal Statistical Society, Series B 48 (1986) 259–302.
[7] H. Birnbaum, A. White, J. Reynolds, P. Greenberg, Z. Mingliang, S. Vallow, J. Schein, N. Katz, Estimated costs of prescription opioid analgesic abuse in the United States in 2001: a societal perspective, Clinical Journal of Pain 22 (2006) 667–676.
[8] H. Burkom, Development, adaptation, and assessment of alerting algorithms for biosurveillance, Johns Hopkins APL Technical Digest 24 (4) (2003) 335–342.
[9] D. Carise, K. Dugosh, A. McLellan, A. Camilleri, G. Woody, K. Lynch, Prescription oxycontin abuse among patients entering addiction treatment, American Journal of Psychiatry 164 (2007) 1750–1756.
[10] B.P. Carlin, S. Banerjee, Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion), in: J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, M. West (Eds.), Bayesian Statistics, vol. 7, Oxford University Press, Oxford, UK, 2003, pp. 45–63.
[11] T. Cicero, J. Inciardi, A. Munoz, Trends in the abuse of oxycontin and other opioid analgesics in the United States: 2002–2004, Journal of Pain 6 (2005) 662–672.
[12] D. Clayton, L. Bernardinelli, Bayesian methods for mapping disease risk, in: P. Elliott, J. Cizick, D. English, R. Stern (Eds.), Geographical and Environmental Epidemiology: Methods for Small-Area Studies, Oxford University Press, Oxford, UK, 1992, pp. 205–220.
[13] D. Clayton, J. Kaldor, Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, Biometrics 43 (1987) 671–681.
[14] W. Compton, N. Volkow, Major increases in opioid analgesic abuse in the United States: concerns and strategies, Drug and Alcohol Dependency 81 (2006) 103–107.
[15] M. Costa, R. Assuncao, A fair comparison between the spatial scan and the besag-newell disease clustering tests, Environmental and Ecological Statistics 12 (2005) 301–319.
[16] G.S. Datta, M. Ghosh, Bayesian prediction in linear models: applications to small area estimation, The Annals of Statistics 19 (1991) 1748–1770.
[17] G.S. Datta, P. Lahiri, T. Maiti, K.L. Lu, Hierarchical Bayes estimation of unemployment rates for the states of the U.S, Journal of the American Statistical Association 94 (1999) 1074–1082.
[18] G.S. Datta, M. Ghosh, L. Waller, Hierarchical and empirical Bayes methods for environmental risk assessment, in: P.K. Sen, C.R. Rao (Eds.), Handbook of Statistics: Bioenvironmental and Public Health Statistics, vol. 18, North-Holland, Amsterdam, 2000, pp. 223–245.
[19] G.S. Datta, P. Lahiri, T. Maiti, Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data, Journal of Statistical Planning and Inference 102 (2002) 83–97.
[20] G.S. Datta, J.N.K. Rao, D.D. Smith, On measuring the variability of small area estimators under a basic area level model, Biometrika 92 (2005) 183–196.
[21] D. Dey, N. Ravishanker, Bayesian approaches for overdispersion in generalized linear models, in: S.K. Ghosh, B.K. Mallick, D.K. Dey (Eds.), Generalized Linear Models: A Bayesian Perspective, Marcel Dekker, New York, 2000, pp. 73–88.
[22] F. Dominici, D. Peng, M. Bell, M. Pham, A. McDermott, S.L. Zeger, J.M. Samet, Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases, Journal of the American Medical Association 295 (2006) 1127–1134.
[23] L. Duczmal, R. Assuncao, A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters, Computational Statistics and Data Analysis 45 (2004) 269–286.
[24] L. Duczmal, M. Kulldorff, L. Huang, Evaluation of spatial scan statistics for irregularly shaped clusters, Journal of Computational and Graphical Statistics 15 (2006) 428–442.
[25] D.B. Duncan, A Bayesian approach to multiple comparisons, Technometrics 7 (1965) 171–222.
[26] S.D. Durham, J.D. Lynch, A threshold representation for the strength distribution of a complex load sharing system, Journal of Statistical Planning and Inference 83 (2000) 25–46.
[27] B. Efron, Bootstrap methods: another look at the jackknife, Annals of Statistics (1979) 1–26.
[28] R. Fisher, Combining independent tests of significance, The American Statistician 2 (1948) 30.
[29] A.E. Gelfand, A.F.M. Smith, Sampling-based approaches to calculating marginal densities, Journal of the American Statistical Association 85 (1990) 398–409.
[30] M. Ghosh, N. Nangia, D.H. Kim, Estimation of median income of four-person families: a Bayesian time series approach, Journal of the American Statistical Association 91 (1996) 1423–1431.
[31] J.U. Gleaton, J.D. Lynch, On the distribution of the breaking strain of a bundle of brittle elastic fibers, Advances in Applied Probability 36 (2004) 98–115.
[32] J.U. Gleaton, J.D. Lynch, Properties of generalized log-logistic families of lifetime distributions, Journal of Probability and Statistical Science 4 (2006) 51–64.
[33] A. Goldenberg, G. Shmueli, R.A. Caruana, S.E. Fienberg, Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales, Proceedings of the National Academy of Sciences 99 (8) (2002) 5237–5240.
[34] G.R. Grimmett, A theorem about random fields, Bulletin of the London Mathematical Society 5 (1973) 81–84.
[35] Hammersley, J.M. and Clifford, P. (1971). Markov fields on finite graphs and lattices. Unpublished manuscript.
[36] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.
[37] J.A. Hoeting, M. Leecaster, D. Bowden, An improved model for spatially correlated binary responses, Journal of Agricultural, Biological, and Environmental Statistics 5 (2000) 102–114.
[38] H. Hogmander, J. Moller, Estimating distribution maps from atlas data using methods of statistical image analysis, Biometrics 51 (1995) 393–404.
[39] H. Hogmander, A. Sarkka, Multitype spatial point patterns with hierachical interactions, Biometrics 55 (1999) 1051–1058.
[40] A.F. Karr, A.P. Sanil, D.L. Banks, Data quality: a statistical perspective, Statistical Methodology 3 (2) (2006) 137–173.
[41] M. Kulldorff, A spatial scan statistic, Communications in Statistics—Theory and Methods 26 (6) (1997) 1481–1496.
[42] M. Kulldorff, W. Athas, E. Feuer, B. Miller, C. Key, Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico, American Journal of Public Health 88 (1998) 1377–1380.
[43] M. Kulldorff, T. Tango, P. Park, Power comparisons for disease clustering tests, Computational Statistics and Data Analysis 42 (2003) 665–684.
[44] M. Kulldorff, L. Huang, L. Duczmal, An elliptic spatial scan statistic, Statistics in Medicine 25 (2006) 3929–3943.
[45] A. Martin, Bayesian inference for heterogeneous event counts, Sociological Methods and Research 32 (2003) 30–63.
[46] P. McCullagh, J. Nelder, Generalized Linear Models, Chapman and Hall, London, 1989.
[47] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, Journal of Chemical Physics 21 (1953) 1087–1092.
[48] D. Montgomery, Introduction to Statistical Quality Control, John Wiley & Sons, Inc., New York, NY, 2001.
[49] Neill, D.B., Moore, A.W., Sabhnani, M., and Daniel, K. (2005). Detection of emerging space-time clusters. In Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
[50] Neill, D.B., Moore, A.W., and Cooper, G.F. (2006). A Bayesian spatial scan statistic. Preprint, School of Computer Science, Carnegie Mellon University.
[51] D. Pfeffermann, L. Burck, Robust small area estimation combining time series and cross-sectional data, Survey Methodology 16 (1990) 217–237.

Please cite this article in press as: D. Banks et al., Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice, Informat. Fusion (2009), doi:10.1016/j.inffus.2009.10.005

[52] N.G.N. Prasad, J.N.K. Rao, The estimation of the mean squared errors of small area estimators, Journal of the American Statistical Association 85 (1990) 163–171.

[53] J.N.K. Rao, Small Area Estimation, Wiley, New York, 2003.

[54] J.N.K. Rao, M. Yu, Small area estimation combining time series and cross-sectional data, Canadian Journal of Statistics 22 (1994) 511–528.

[55] H. Rolka, H. Burkom, G. Cooper, M. Kulldorff, D. Madigan, W.-K. Wong, Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: research needs, Statistics in Medicine 26 (2007) 1834–1856.

[56] S.K. Sahu, A.E. Gelfand, D.M. Holland, Spatio-temporal modeling of fine particulate matter, Journal of Agricultural, Biological, and Environmental Statistics 11 (2006) 61–86.

[57] J.G. Scott, J.O. Berger, An exploration of aspects of Bayesian multiple testing, Journal of Statistical Planning and Inference 136 (2006) 2144–2162.

[58] F. Spitzer, Markov random fields and Gibbs ensembles, American Mathematical Monthly 78 (1971) 142–154.

[59] R.K. Tsutakawa, Mixed model for analyzing geographic variability in mortality rates, Journal of the American Statistical Association 83 (1988) 37–42.

[60] L. Waller, B. Carlin, H. Xia, A. Gelfand, Hierarchical spatio-temporal mapping of disease rates, Journal of the American Statistical Association 92 (1997) 607–617.

[61] R.A. Waller, D.B. Duncan, A Bayes rule for the symmetric multiple comparisons problem, Journal of the American Statistical Association 64 (1969) 1484–1503.

Please cite this article in press as: D. Banks et al., Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice, Informat. Fusion (2009), doi:10.1016/j.inffus.2009.10.005