OXFORD UNIVERSITY PRESS | International Journal of Epidemiology

Big Data Supplement: Advances in spatio-temporal models for non-communicable disease surveillance

Journal:	*International Journal of Epidemiology*
Manuscript ID	IJE-2018-09-1139.R2
Manuscript Type:	Supplement [approved only]
Date Submitted by the Author:	07-May-2019
Complete List of Authors:	Blangiardo, Marta; Imperial College London, Department of Epidemiology and Biostatistics, School of Medicine Boulieri, Areti; Imperial College London, Department of Epidemiology and Biostatistics, School of Medicine Diggle, P; Lancaster University Faculty of Health and Medicine, Medical School Piel, Frédéric; Imperial College London, Epidemiology & Biostatistics Shaddick, G; Exeter University, Mathematics Elliott, Paul; Imperial College Londofn, Department of Epidemiology & Biostatistics
Key Words:	surveillance, non-communicable diseases, Bayesian hierarchical models, spatio-temporal modelling

Advances in spatio-temporal models for non-communicable disease surveillance

Blangiardo, M.[1,2], Boulieri, A.[2], Diggle, P.[3], Piel, F.B.[1,2], Shaddick, G.[4] & Elliott, P.[1,2]

[1] UK Small Area Health Statistics Unit, Department of Epidemiology & Biostatistics, School of Public Health, Imperial College London, London, UK
[2] MRC-PHE Centre for Environment & Health, Department of Epidemiology & Biostatistics, School of Public Health, Imperial College London, London, UK
[3] Centre for Health Informatics, Computing and Statistics (CHICAS), Lancaster Medical School, Lancaster University, Lancaster, UK
[4] Department of Mathematics, University of Exeter, Exeter, UK

Abstract

Surveillance systems are commonly used to provide early warning detection or to assess an impact of an intervention/policy. Traditionally, the methodological and conceptual frameworks for surveillance have been designed for infectious diseases, but the rising burden of non-communicable diseases (NCDs) worldwide suggests a pressing need for surveillance strategies to detect unusual patterns in the data and to help unveil important risk factors in this setting. Surveillance methods need to be able to detect meaningful departures from expectation and exploit dependencies within such data to produce unbiased estimates of risk as well as future forecasts. This has led to the increasing development of a range of space-time methods specifically designed for NCD surveillance.

We present an overview of recent advances in spatio-temporal disease surveillance for NCDs using hierarchically specified models. This provides a coherent framework for modelling complex data structures, dealing with data sparsity, exploiting dependencies between data sources and propagating the inherent uncertainties present in both the data and the modelling process. We then focus on three commonly used models within the Bayesian Hierarchical Model (BHM) framework and through a simulation study we compare their performance.

We also discuss some challenges faced by researchers when dealing with NCD surveillance, including how to account for false detection and the modifiable areal unit problem. Finally, we consider how to use and interpret the complex models, how model selection may vary depending on the intended user group and how best to communicate results to stakeholders and the general public.

Keywords: surveillance, non-communicable diseases, Bayesian hierarchical models, spatio-temporal modelling.

Key messages

- There is increasing recognition of the importance of surveillance for NCDs.
- Spatio-temporal variation in health outcomes and lifestyle and environmental exposures needs to be explicitly modelled in order to reduce bias and uncertainty.
- Hierarchical modelling provides a coherent framework within which spatio-temporal dependencies can be explicitly modelled with integration of the uncertainties associated both with the data and the modelling process.
- In a simulation study, we found that mixture models designed for detection perform better than standard disease mapping models. However, attention should be paid to the choice of threshold as this impacts the results. It is recommended that a simulation study based on the characteristics of the data in hand is run each time for the selection of suitable threshold values.
- Current research challenges in this area include: the use of data from multiple sources at different spatial and temporal scales and with different sources of bias and uncertainty; computationally intense processes; and control for false positive findings.

1. The importance of non-communicable diseases

According to the World Health Organization surveillance is the "ongoing systematic data collection, analysis, interpretation and dissemination of information in order for action to be taken" [1]. National public health agencies, such as the US Centers for Disease Control and Prevention (CDC) and Public Health England (PHE) routinely carry out surveillance data analysis to provide early warnings of unexplained changes in incidence patterns of diseases as well as to aid policy formation and resource allocation [2]. Specific examples include the international influenza monitoring system which started in 1948 and is now distributed in 82 countries [3], the HIV and AIDS Reporting System (HARS) used by PHE [4] and the National HIV Surveillance System used by CDC [4].

To date, the majority of methods and models commonly used in public health surveillance are designed for monitoring cases of infectious diseases [6]. Due to the rising burden of non-communicable diseases (NCDs) worldwide, there is a pressing need to implement surveillance strategies to detect trends, highlight unusual changes and consequently assist in outlining emerging NCD risk factors. NCD surveillance shares many objectives with infectious disease surveillance, including generating information to guide public health action, detecting the health impact of environmental exposures, or of environmentally driven disease vectors; however, it also presents some different methodological challenges [7, 8].

Health data contain both a time and a space component. Surveillance methods must be able to capture spatial and temporal patterns in both lifestyle/environmental exposures and health outcomes. Here, we present an overview of the approaches developed for spatio-temporal disease surveillance of NCDs. We focus on model-based methods and among these on Bayesian hierarchical models (BHMs), which can naturally accommodate complex data structures, as well as propagate uncertainty due to the data themselves and the modelling process.

In this Section, we first discuss how data availability is one of the key challenges in surveillance studies, before giving a generic overview of test-based approaches for NCD surveillance. We then focus on BHMs (Section 2) and describe disease mapping and mixture-based models for anomaly detection. In Section 3, we introduce the computational aspects of the BHM modelling framework for NCD surveillance, while in Section 4 we run a simulation study to evaluate advantages and drawbacks of the approaches presented in detecting areas deviating from the expected trend. Finally, in Section 5 we conclude with a summary and some remaining discussion points.

1.1 Data availability

One of the major challenges of surveillance studies is the availability of suitable data. This applies to both infectious disease and NCD surveillance. It is a particularly important issue in low-income settings because surveillance studies often need to rely on information from surveys, and the lack of financial resources may make comprehensive coverage of data sources (e.g. mortality/cancer registries) over an entire country infeasible [9].

In the last 15 years, a number of Health and Demographic Surveillance Systems (HDSS) have been established in low-income settings to provide a reliable source of health data and are now linked together through the International Network for the Continuous Demographic Evaluation of Populations and Their Health (INDEPTH, [10]). Such continuous surveys are an invaluable source of data, but researchers face issues related to population representativeness. A recent study proposed a Bayesian probabilistic clustering method to evaluate the network representativeness in terms of socio-economic and environmental variables in sub-Saharan Africa, identifying areas of poor coverage in the existing network and using predictive probability distributions to suggest the best location for new HDSS sites [11].

Even in high income settings where administrative resources are available for the entire population, there may be issues regarding the population at risk, used as denominator in the risk estimates. In small-area studies on mortality or hospital admissions, the denominator is usually the resident population in each administrative area, typically estimated from national census statistics, but there may be estimation problems for inter-censual years. In addition, it is not straightforward to define the denominator where the interest is for less-defined geographies, such as the catchment areas of clinical centres (e.g. general practices in England).

Furthermore, the availability of administrative or health data may become more limited: for instance, within the UK National Health Service, patients can now decide not to share their medical records for research purposes. This clearly impacts on spatial coverage and could potentially lead to biased statistical inference if the data gaps are clustered in space and/or if they differentially affect specific population groups (e.g. elderly, more deprived) [12].

1.2 Test-based methods

Methods for NCD surveillance have largely been based around the idea of detecting whether the outcome of interest shows a particular behaviour in a defined subset (e.g. an area, period of time, or combination of space and time) when compared to the whole study region. Perhaps the most popular test-based methods used for NCD surveillance are the scan statistics. These were developed originally in the temporal setting only [13]; here, a fixed length “scanning window” is passed over the time-series data with the number of cases in the window being recorded. A log likelihood ratio (LLR) is calculated for each interval, and the test statistic is defined as the maximum LLR over all intervals. This idea was extended to a spatial version of the scan statistic [23, 15], which was later further extended to the spatio-temporal setting [16]. In this case, the scanning window is represented by a cylinder, where the diameter specifies the spatial dimension and the height the temporal dimension. An additional version of spatial scan statistic was proposed to account for correlation across spatial units which was not considered before [17]. Scan statistics have been extensively applied to numerous health care applications. Part of their popularity lies in the availability of free user-friendly SaTScan™ software (https://www.satscan.org/). Recent applications of SaTScan include the identification of signals for colorectal cancer [18], drug activity [19], criminality [20], and bat activity [21].

A further development has been the detection of spatial variations in temporal trends (SVTT). These methods extend the scan statistics to estimate the time trend via a regression-based model specifying either a linear or a quadratic function. The quadratic SVTT method has, for example, been applied to cervical cancer data in women in the US from 1969 to 1995 highlighting areas where the risk was significantly different from the rest [22].

Test-based methods such as scan statistics can only answer questions related to the deviation from the null hypothesis. An alternative approach is to explicitly model the spatio-temporal structure in the data and to assess whether differences between observed data and those predicted from the model provide evidence of anomalies. There are a number of advantages to adopting a model-based approach over a test-based method including the ability to: (i) have more statistical power to handle sparsity in the observed disease counts; (ii) explore more subtle departures from the expectation; (iii) account for the spatial and temporal correlation that is typically evident in health data; (iv) "borrow" information over space and time, therefore increasing the precision of the estimates generated; (v) include covariates that might explain some of the spatio-temporal variability.

2. Hierarchical models and likelihood-based inference

Hierarchical models (HM) are able to deal with complex data structures, to exploit dependencies between data sources and to propagate the inherent uncertainties that are present in both the data and the modelling process. In the current context, an HM combines two elements: a *process model* that describes how disease risk varies over space and time, typically involving both extant covariate effects and a latent spatio-temporal stochastic process; and a *data model* that describes the statistical properties of the available health outcome data conditional on the realisation of the underlying risk process. Both elements are specified up to the values of a set of unknown *parameters*, which can be estimated by Bayesian or non-Bayesian versions of likelihood-based inference, typically implemented using Markov chain Monte Carlo integration and Monte Carlo likelihood maximisation methods, respectively. In addition to estimating parameters, the scientific goals of health surveillance include *prediction* of relevant properties of the unobserved risk surface as it evolves in real-time. Parameters and latent stochastic processes are fundamentally different things, but within the Bayesian paradigm they are both treated as unobserved sets of random variables, and the operational calculus of estimation and prediction coalesces. In what follows, we use BHM (Bayesian Hierarchical Modelling) as a shorthand for Bayesian inference applied to a hierarchically specified model.

2.1 Space-time disease mapping

A class of BHMs that has been extensively used for the analysis of NCD data are the so-called disease mapping models (DM). These are hierarchical models in which the latent component of area-level disease risk is modelled as a spatially discrete Markov random field [23] and, depending on the sampling design, the conditional distribution of area-level case-counts is Poisson or Binomial [24, 25, 26, 27]. While the objectives of these are descriptive, they have been used as the basis for the development of detection models, which are framed in a surveillance perspective. Disease mapping models have been extensively used to estimate and visualise the spatial or spatio-temporal distribution of a disease (see for instance [9, 21, 28]).

Spatial dependence in the latent component of a DM is modelled by specifying neighbourhood relationships amongst the area-level risks, the most widely used definition being that two areas are *neighbours* if they share a common boundary. A common choice for capturing temporal dependence is a random walk prior [29], but extensions to incorporate spatio-temporal interactions among neighbouring areas and time points have also been developed [30]. This framework can also account for factors known to modify spatial and temporal trends that, in the context of NCDs, will include demographic variables (e.g. age/sex/ethnicity) and social economic status. Random effects can be assigned for each factor (with appropriate priors) and for interactions if required. An example is provided by Goicoa *et al.* [31] who proposed a space-time-age model to study prostate cancer incidence across 50 provinces in Spain for 9 age groups over 25 years, accounting for all pair-wise interactions. The authors used ranking of all provinces according to mortality rates to identify high-risk groups.

A key characteristic of BHMs is the ready availability of joint posterior/predictive distributions for parameters/latent processes and whatever of their properties are relevant to the public health questions of interest. In the context of disease mapping this leads to a spatial, temporal and spatio-temporal risk distribution that researchers can map, both in terms of point estimates but also of associated measures of uncertainty. For the latter, a common choice is the predictive probability that the relative risk exceeds a pre-specified threshold [32, 33]. Exceedance probabilities can be used to flag areas and/or time points characterised by increased risk that may then be further investigated. In this way, disease mapping, though not formally a surveillance method, can be used as a descriptive tool for the identification of areas and/or time periods with marked deviation from expectation. It is important to note that the strong smoothing effect of disease mapping models leads to conservative risk estimates, hence to a small number of false positive findings, at the expense of a low power for detecting high risk areas with low signal. To minimise that, an extensive simulation was run to find the best threshold on the exceedance probability scale to classify an area as high risk [33]. The authors showed that a good trade-off between false-positive and false-negative rates is achieved with a probability above 0.8 for a relative risk to be higher than 1, however this largely depends on the number of expected counts, the number of areas and time points, and the spatial risk [34].

As an example of the typical disease mapping output, Figure 1 shows the incidence of malignant melanoma in males, at the census ward level in England and Wales over the period 1985-2009 from the Environmental and Health Atlas produced by the UK Small Area Health Statistics Unit (SAHSU) [35]; the map on the left presents the spatial distribution of the posterior relative risk mean estimates, while the map on the right plots the posterior probability that the corresponding relative risk is above 1, using the categorisation suggested in [33].

Figure 1 here

Disease mapping models can be extended to two or more outcomes that might share spatial (and temporal) patterns, for instance due to common risk factors. A joint model allows information to be borrowed across the outcomes, thus helping stabilise estimates, particularly when the outcomes are rare. The shared component model [36], originally developed for two diseases, includes a common component (likely to reflect common risk factors), and a disease-specific one, which can point towards specific risk factors otherwise masked in a single disease model. It was applied to male and female lung cancer [37] and later extended to jointly model multiple diseases [30, 38], with an application on oral cavity, oesophagus, larynx and lung cancers in males in the 544 districts of Germany from 1986 to 1990. Recently, it was further extended to jointly model age- and gender-specific diseases [39].

An alternative multivariate specification considers spatial and temporal terms explicitly, modelling the correlation among the outcomes in space/time. As an example, road traffic accidents characterised by different severity were analysed over the period 2005-2011 at the ward level in England while detection of high-risk areas was performed using exceedance probabilities of the area ranks based on accident rates [40].

2.2 Space-time anomaly detection

The standard disease mapping approach has been used informally to detect anomalies (unusual observations) in space and time, i.e. areas and/or time points with trends different to the expected ones, adding a space-time interaction parameter into the latent process [37]. The detection of anomalies may indicate the presence of an emerged localised risk factor, the impact of an intervention, or differences in the quality of data, such as misdiagnosis of a disease, and under- or over-reporting of cases.

Mixture models have been proposed as a formal approach to anomaly detection. In particular, Abellan *et al.* [41], developed a BHM model (termed STmix) where a mixture of two normal distributions characterised by different variances is specified for the space-time interaction. Then, the interaction is used to classify areas as *common* and *unusual*. The authors performed a simulation study to compare the method against the standard disease mapping approach. The results of the simulation study showed that the standard approach was not able to capture the variability in the spatio-temporal interactions and therefore it was not able to distinguish between common and unusual areas. This is due to the excessive smoothing following the assumption of a common variance across all the areas and time points. STmix was applied to mammography screening data in Brisbane, Australia, at the statistical local area (SLA) level, from 1997 to 2008 in order to identify SLAs whose temporal trend exhibited volatility [42]. A well-known drawback of this approach is its limitation to incorporate specific time patterns, for example step changes that could signal the emergence of a new risk factor.

Another mixture model, proposed by Li *et al.* [43], accommodates this issue. Here, the mixture specification of the method is defined directly on the relative risks in space and time, to allow for detection of areas with unusual time trends rather than space-time deviations. In particular, two alternative models are considered: the first one assumes a global time trend for all areas (*common trend*), while the second estimates a time trend for each area independently (*area-specific trend*). Through a simulation study the authors showed better performance in terms of both sensitivity and proportion of false positives compared to SaTScan on a wide range of scenarios. This approach, named BaySTDetect, was applied to detect unusual trends for asthma and chronic obstructive pulmonary disease at Clinical Commissioning Group (CCG) level in England (211 in total) on monthly data between August 2010 and March 2011 across mortality, hospital admissions, and general practice drug prescriptions [44]. To illustrate the typical output obtained from this model, Figure 2 shows the area-specific time trends of the CCGs that were detected as unusual plotted against the national trend. Other applications of this method include burglary data [45], gray whale abundance [46] and mammography data [42].

Figure 2 here

This method was further extended to increase its flexibility by accounting for different space-time patterns in the unusual observations, as well as by allowing for longer time series to be analysed. This improved method, termed FlexDetect, had a better performance when compared to the original method through an extensive simulation study [47].

Multiple testing

As surveillance studies involve evaluating trends for different health outcomes, many areas and different time periods at the same time, false detections are likely to occur by chance. Bonferroni correction has been extensively used in epidemiology to correct for multiple testing, particularly in omics studies [48, 49], but it is well known that this approach leads to conservative results. Benjamini and Hochberg first introduced an alternative index, the false discovery rate (FDR) [50], as the expected value of the rate of false positive findings among all rejected hypotheses and used it in a frequentist approach. The same method was suggested in the context of descriptive spatial epidemiology, to obtain areas characterised by a Standardised Mortality Ratio different from 1 [51]. Even in the Bayesian setting, FDR rules were suggested by many authors [52, 53, 54, 55, 56]. The mixture model proposed by Li et al. uses the specification suggested by Newton et al. [53] later used by Ventrucci et al. [57] in order to account for multiple testing. The authors base the FDR statistic on the posterior model probability, which represents the likelihood of the space-time unit investigated to follow the common trend model, i.e. to exhibit a risk pattern not deviating from the expected one [43].

While the importance of controlling for multiple testing is clear in classical significance testing, the analogous problem in predictive setting is less of a concern [58]. One reason for this is that local predictions from hierarchical models are naturally smoothed towards the global mean, making these consequently less prone to false positive findings than unsmoothed area-by-area interval estimates. Another is that the Monte Carlo sampling method allows the computation of whatever joint probability statements are required. For example, if the public health question is whether current risk exceeds an agreed acceptable level in all areas that do, and in no areas that do not, meet a particular criterion such as adherence to a particular advisory policy, the correct predictive probability to attach to this statement can be calculated.

3. Computational aspects

One of the biggest challenges researchers face when analysing large and complex space-time datasets is their computational burden. This applies particularly in the small-area context, where the number of space-time units investigated can vary substantially depending on the chosen spatial and temporal resolution, from few hundreds to hundreds of thousand units, particularly when several outcomes are jointly analysed (for instance Foreman et al. [59] considered jointly deaths/age/sex specific space-time trends in the U.S).

The degree of complexity of the model (e.g. the number of parameters) also impacts on the computational burden, for instance in terms of convergence time when running MCMC simulations. Under the Bayesian paradigm, the choice of the prior will also influence convergence; an informative prior, assuming that there is no conflict with the data, will normally speed up convergence, while a vague prior will most likely lead to longer time to reach convergence. Finally, the choice of software used for the analysis will affect the model running time. The user-friendly software BUGS (Bayesian inference Using Gibbs Sampling) [60] has been traditionally used for Bayesian inference using MCMC methods, however it can be slow when high dimensional data and/or complex models are used. Other MCMC-based methods, such as Stan [61] and NIMBLE [62], are currently attracting attention due to their active development community. An alternative way of dealing with computational limitations is to use approximative methods; for instance INLA (integrated nested Laplace approximations) [63] has been successfully used for running space-time disease mapping models (e.g. [31, 64]); however, this method is somewhat less flexible than the aforementioned ones and as it relies on Normality of the latent process is not able to deal with mixture distributions.

Computationally intensive BHMs benefit from high-performance computing clusters to speed up computation times, however these are not necessarily required. For instance in [40] road traffic accidents data of different severity in England were analysed simultaneously at the ward level (~8000), over 9 years (for a total of around 150,000 units and 32,000 parameters); the analysis was run in OpenBUGS and took 20 to 27hr on an Intel Core processor at 3.40 GHz with 8 Gbytes of random-access memory. On a much bigger scale, a US small area study considered more than half a million units and nearly 6,000 parameters [59]; the analysis was implemented in Stan using higher performance computing (HPC) clusters for faster calculations.

4. Simulation-based example

Considering the multitude of space-time methods available, as described above, it is important to formally evaluate their respective detection performance. In this paper we carried out a simulation study to formally evaluate the detection performance and compare DM, STmix and FlexDetect (see description of the models in Section 2.2). Following the design initially proposed by [43], and later used by [47], we used real asthma hospital episode statistics (HES) data to generate 50 simulated datasets. The asthma dataset was obtained from SAHSU, Imperial College London and consisted of disease counts across 211 Clinical Commissioning Groups (CCG) in England for 15 months, from January 2010 to March 2011. The 50 simulated datasets were generated to closely resemble the patterns seen in the real dataset. A standard spatio-temporal model [27] was first fitted to the real data and the obtained parameters were selected for the generation of the simulated data. We selected 15 areas to deviate from the overall time trend over the last 5 time points. For these 15 areas, we selected the signal to be increased by log(2) for time points 3 and 10, and decreased by log(2) for time points 6, 12 and 15 out of the total 15 time points. In this way, we ensured that a realistic scenario was used (for more details see [47] Supplementary materials, Scenario 1). The R code used for the data simulation, together with the three models written in BUGS can be found on https://github.com/aretib/bayesSTmodels.git.

The results are presented in Table 1 in terms of four different performance measures. We defined TP as the number of true positives, FP as false positives, TN as true negatives and FN as false negatives respectively. Sensitivity measures the ability of the model to correctly classify an unusual observation as such, defined as TP/(TP+FN), and similarly specificity measures the ability of the model to correctly classify a common observation as such (TN/(TN+FP)). In addition, it is crucial to control the proportions of observations that are falsely classified as unusual (false discovery rate (FDR)) and common (false omission rate (FOR)) respectively; these should typically not exceed a value of 0.05, specified based on the standard p-value threshold. We consider (i) two different thresholds for DM: 0.8 as commonly used and previously described (DM1); a more conservative threshold of 0.9 (DM2), under the assumption that false positives are more important to minimise than false negatives, and (ii) two different rules for STmix as presented in the original paper: an area is modified if at least for one time point the space-time interaction has a probability greater than 0.8 to be above 1 (STmix1); an area is modified if for at least three time points the space-time interactions have an average probability greater that 0.8 to be above 1 (STmix2).

Table 1 here

As can be seen, the disease mapping approach using the standard threshold of 0.8 on the posterior probability scale (DM1) shows the worst performance; as expected the method is able to detect nearly all unusual areas, with a sensitivity of 0.979; however, roughly 79% of the detected findings are not actually unusual (FDR = 0.785) (Table 1). Fixing a 0.9 threshold (DM2), FDR decreases, despite still being above the standard threshold of 0.05, while at the same time sensitivity also decreases (0.660) (Table 1).

The two mixture models returned more comparable performances. STmix1 gave no false positive results (FDR = 0) and a sensitivity of 0.773, while for STmix2 sensitivity increased to 0.969, but at the same time a much higher proportion of false positive was detected (FDR = 0.220). The results of FlexDetect provided a balance between the two extremes, giving a proportion of FDR equal to 0.019 and a sensitivity of 0.796. In terms of specificity and FOR, both STmix and FlexDetect behave similarly. Differences across the competing models were observed in terms of computation time, an important factor in assessing their performance. All models were run in an Intel Xeon Core processor 3.40GHz with 125GB RAM. Each of the 50 simulations took on average 33.4 mins for models DM1 and DM2, 39.2 mins for models STmix 1 and STmix2 and 66.8 mins for FlexDetect. The simulation results suggest that using disease mapping (DM) for surveillance purposes is not appropriate and that one of the mixture models designed for detection should be used instead. Between STmix and FlexDetect it is worth mentioning that STmix can only identify areas where anomalies are present, and not the time points when these occur. In addition to this, its detection mechanism does not consider specific patterns in the time trends. These can be accommodated by FlexDetect, which however is more computationally intensive. Also note that mixture models notoriously have problems converging, suffering from issues such as label switching, which lead to multimodal posterior distributions. Both the detection methods deal with this through the modelling specification, such as [41] constraining the variance of the modified areas to be larger than that of unmodified areas, or through informative priors on the variances of the two components [43].

5. Discussion

In this paper we have presented an overview of the main statistical methods for disease surveillance in the context of NCDs, both from a test-based and model-based perspective and with a particular focus on the BHM approach, which provides a flexible framework to allow for complex data dependencies present in surveillance studies. Through a simulation study we showed that disease mapping is not satisfactory when looking for data anomalies, while the two methods based on mixture models provide a better compromise between detecting areas characterised by a deviation from the expected trend and limiting false positives. Note that our perspective is on methods that detect single areas, rather than clusters of adjacent spatial units. If the interest lays on detection in the presence of spatial proximity, recent methods have been developed to combine clustering with spatial smoothing, see for example [65] and [66]. An interesting aspect of the general hierarchical framework presented is that it can easily incorporate forecasting of the disease risk, which is relevant in the context of epidemiological surveillance to evaluate the need for resources/policies/costs in specific areas and future time points. Some work in this area includes Foreman et al., [59] who, using annual vital statistics for 1974-2011 at the US state spatial resolution, forecasted mortality up to 2024, while Ugarte et al. [67] used P-splines to forecast cancer mortality counts in Spanish regions for 2009-2011 using data from 1975-2008. Most of the work presented is based on routinely collected data for retrospective studies. However, there is increased importance of early warning detection, so that unusual behaviour can be detected at the earliest possible time. Syndromic data, such as primary care data, drug prescriptions, nurse calls and home visits, which are indicative of a potential anomaly, may provide an additional level of information leading to a detection event before the data aberration occurs [68]. Diggle et al. [69, 70] analysed NHS non-emergency telephone calls reporting symptoms of gastrointestinal diseases. The authors specified a spatio-temporal point process on the location and time of the individual calls and modelled the spatial and temporal dependency on the intensity of the process. They used exceedance probabilities to define maps of potential outbreaks. Another example can be found in Morrison et al. [71] who forecasted multiple measures of healthcare utilization (including physician visits and prescriptions of asthma medication) within British Columbia, Canada, where seasonal wildfires produce high levels of air pollution, significantly impacting population health. Here, the focus was on efficient, near real-time, computation which was achieved using INLA to perform approximate Bayesian inference.

Potentially syndromic information can also be linked with routine data such as Hospital Episode Statistics and provide predictors in order to obtain a better description of the data and more accurate one-step-ahead forecasts. Lately work has been done to take advantage of the rich data from social media in a surveillance perspective. For instance, Dai et al. [72] linked tweets with the American Community Survey and the Behavioral Risk Factor Surveillance System to study asthma prevalence at the State level in the US. The authors claimed that the inclusion of social media data could be a cost-effective real time health detection system. However, there may be challenges in future due to selective data availability following perceived concerns about data security and confidentiality, as demonstrated by the newly implemented NHS National Data Optout Programme. This will potentially lead to bias in population representativeness due to non-random missingness [12] which will need to be addressed using advanced statistical methods, for instance through the integration of data from appropriate surveys / cohorts, as proposed in the context of residual confounding [73].

An important issue with surveillance studies is that of the spatial resolution and the type of geographical areas considered; modifying these might lead to different results, as the spatial distribution of the outcome will depend on these choices. For instance, if within-area variability is substantial, results from statistical inference might suffer from false negative observations, as potentially high-risk places are aggregated with low-risk ones. The more spatial variability is present in the data, the more profound the potential impact of the modifiable areal unit (MAUP) [74, 75]. As MAUP depends on the level of aggregation, this issue has been linked to ecological bias [76] and the general suggestion in the scientific literature is to consider the finest spatial scale available. This can be particularly challenging for rare diseases where the number of cases at small-area level are very low. Furthermore, the choice of spatial resolution is mostly dependent on data availability and sparsity. BHMs have been suggested as a way to, at least partially, deal with MAUP. As there is an explicit relationship among areas globally and/or locally, through structured random effects, places belonging to a particularly small area can influence results for other areas, hence alleviating the MAUP problem [77].

A key aspect of surveillance studies concerns how to communicate information to public health researchers and policy makers. This is particularly challenging as the statistical modelling of surveillance data becomes more sophisticated. In this context it is essential to develop user friendly tools such as atlases, web applications and reporting services that allow for data visualisations and easy implementation of the advanced methodologies. The Environment and Health Atlas for England and Wales [35] (typical output from the Atlas was presented in Figure 1) is an example of work in this direction, providing stakeholders and the general public with a collection of maps to inform on the spatial distribution of environmental factors and diseases. Through the exceedance probabilities, these maps give a perception of the uncertainty around the area level relative risks estimates.

Web applications allow the ready implementation of statistical methods and perform complex data analyses, often through interactive data visualisations. These can be particularly useful for practitioners less skilful in statistical modelling and programming. As an example, the Rapid Inquiry Facility (RIF) which is currently being redeveloped within SAHSU, is designed to facilitate disease mapping and risk analysis studies and has been employed by more than 45 institutions in a number of countries [78]. A more recent example is the SpatialEpiApp that integrates two methods for disease mapping and cluster detection [79].

To conclude, in this paper we presented a range of BHMs, which have proved to be useful for non-communicable disease surveillance. The choice of model should depend on various factors and most importantly on the objective of the study, characteristics of the data, and computational resources. It is commonly recommended to perform simulation studies based on the data in hand, to inform the model and to select detection rules that are most appropriate in each case.

We believe that epidemiological surveillance will be at the centre of future methodological research to match the continuous increase in data availability, e.g. through social media; this will also open up issues related to data integration, selection bias and spatio-temporal misalignment. At the same time there will be the need to reduce the computational burden of increasingly complex models applied to large datasets, in order to provide timely results for decision making.

Acknowledgments

The UK Small Area Health Statistics Unit (SAHSU) is part of the MRC-PHE Centre for Environment and Health, which is supported by the Medical Research Council (MR/L01341X/1) and Public Health England (PHE). Part of this work was supported by an Early Career MRC Fellowship awarded to AB and a Wellcome Trust Seed Award in Science awarded to FBP (204535/Z/16/Z). PE is Director of SAHSU and Director of the MRC-PHE Centre for Environment and Health.

All authors declare no conflicts of interest.

References

1. Thacker SB, Stroup DF. Future directions for comprehensive public health surveillance and health information systems in the United States. Am J Epidemiol 1994; 140(5):383-97.
2. Lawson AB, Kleinman K, editors. Spatial and syndromic surveillance for public health. John Wiley & Sons; Chichester, England 2005.
3. World Health Organization. Global Influenza Surveillance and Response System (GISRS) [Internet]. 2018 [cited 24 April 2018]. Available from: http://www.who.int/influenza/gisrs_laboratory/en/
4. Public Health England. HIV surveillance systems [Internet]. 2008 [cited 24 April 2018]. Available from: https://npin.cdc.gov/funding/national-hiv-surveillance-system-nhss
5. Centers for Disease Control and Prevention. National HIV Surveillance System (NHSS) [Internet]. 2015 [cited 24 April 2018]. Available from: https://www.gov.uk/guidance/hiv-surveillance-systems
6. Lombardo JS, Buckeridge DL. Disease surveillance: a public health informatics approach. John Wiley & Sons; Chichester, England 2012.
7. Thacker SB, Stroup DF, Parrish RG, Anderson HA. Surveillance in environmental public health: issues, systems, and sources. Am J Public Health 1996; 86(5):633-8.
8. Frumkin H, Hess J, Luber G, Malilay J, McGeehin M. Climate change: the public health response. Am J Public Health 2008; 98(3):435-45.
9. Diggle PJ, Giorgi E. Model-based geostatistics for prevalence mapping in low-resource settings. J Am Stat Assoc 2016; 111(515):1096-120.
10. Ye Y, Wamukoya M, Ezeh A, Emina JB, Sankoh O. Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Sahara Africa?. BMC public health 2012; 12(1):741.
11. Utazi CE, Sahu SK, Atkinson PM, Tejedor N, Tatem AJ. A probabilistic predictive Bayesian approach for determining the representativeness of health and demographic surveillance networks. Spat Stat-Neth 2016; 17:161-78.
12. Piel FB, Parkes BL, Daby H, Hansell AL, Elliott P. The challenge of opt-outs from NHS data: a small-area perspective. J Public Health 2018 (to appear).
13. Naus JL. Clustering of random points in two dimensions. Biometrika 1965; 52(1-2):263-6.
14. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. Stat Med 1995; 14(8):799-810.
15. Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods 1997; 26(6):1481-96.
16. Kulldorff M, Heffernan R, Hartman J, Assunçao R, Mostashari F. A space–time permutation scan statistic for disease outbreak detection. PLoS Med 2005; 2(3):e59.
17. Loh JM, Zhu Z. Accounting for spatial correlation in the scan statistic. Ann Appl Stat 2007; 1(2):560-84.
18. Sherman RL, Henry KA, Tannenbaum SL, Feaster DJ, Kobetz E, Lee DJ. Peer Reviewed: Applying Spatial Analysis Tools in Public Health: An Example Using SaTScan to Detect Geographic Targets for Colorectal Cancer Screening Interventions. Prev Chron Dis 2014; 11.

19. Linton SL, Jennings JM, Latkin CA, Gomez MB, Mehta SH. Application of space-time scan statistics to describe geographic and temporal clustering of visible drug activity. J Urban Health 2014; 91(5):940-56.
20. Runadi T, Widyaningsih Y. Application of hotspot detection using spatial scan statistic: Study of criminality in Indonesia. In AIP Conference Proceedings 2017 (Vol. 1827, No. 1, p. 020011). AIP Publishing.
21. Adams AM, Fenton MB. Identifying peaks in bat activity: a new application of SaTScan's space–time scan statistic. Wildl Res 2017; 44(5):392-9.
22. Moraga P, Kulldorff M. Detection of spatial variations in temporal trends with a quadratic function. Stat Methods Med Res 2016; 25(4):1422-37.
23. Rue H, Held L. Gaussian Markov random fields: theory and applications. CRC press; Boca Raton, US 2005.
24. Leroux BG, Lei X, Breslow N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In Statistical models in epidemiology, the environment, and clinical trials (pp. 179-191). Springer, New York, US 2000.
25. MacNab YC, Dean CB. Parametric bootstrap and penalized quasi-likelihood inference in conditional autoregressive models. Stat Med 2000; 19(17-18):2421-35.
26. Rodrigues EC, Assunção R. Bayesian spatial models with a mixture neighborhood structure. J Multivar Anal 2012; 109:88-102.
27. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 1991; 43(1):1-20.
28. World Health Organization. Global action plan for the prevention and control of non communicable diseases 2013–2020. Geneva: WHO; 2013. http://www.who.int/nmh/publications/ncd-action-plan/en/
29. Fahrmeir L, Lang S. Bayesian inference for generalized additive mixed models based on Markov random field priors. J R Stat Soc Ser C Appl Stat 2001; 50(2):201-20.
30. Held L, Natário I, Fenton SE, Rue H, Becker N. Towards joint disease mapping. Stat Methods Med Res 2005; 14(1):61-82.
31. Goicoa T, Ugarte MD, Etxeberria J, Militino AF. Age–space–time CAR models in Bayesian disease mapping. Stat Med 2016; 35(14):2391-405.
32. Clayton DG. Bayesian methods for mapping disease risk. Geographical and environmental epidemiology: methods for small-area studies. 1992:205-220.
33. Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. Environ Health Perspect 2004; 112(9):1016.
34. Ugarte MD, Goicoa T, Militino AF. Empirical Bayes and Fully Bayes procedures to detect high-risk areas in disease mapping. Comput Stat Data An 2009; 53(8):2938-49.
35. Hansell AL, Ghosh RE, Fecht D, Fortunato L. The environment and health atlas for England and Wales. Oxford University Press, Oxford, England 2014.
36. Knorr-Held L, Best NG. A shared component model for detecting joint and selective clustering of two diseases. J R Stat Soc Ser A Stat Soc 2001; 164(1):73-85.
37. Richardson S, Abellan JJ, Best N. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK). Stat Methods Med Res 2006; 15(4):385-407.
38. Mahaki B, Mehrabi Y, Kavousi A, Schmid VJ. A Spatio-Temporal Multivariate Shared Component Model with an Application in Iran Cancer Data. arXiv preprint arXiv:1707.06075. 2017 Jul 19.

39. Etxeberria J, Goicoa T, Ugarte MD. Joint modelling of brain cancer incidence and mortality using Bayesian age-and gender-specific shared component models. Stoch Environ Res Risk Assess 2018; 1-9.
40. Boulieri A, Liverani S, Hoogh K, Blangiardo M. A space–time multivariate Bayesian model to analyse road traffic accidents by severity. J R Stat Soc Ser A Stat Soc 2017; 180(1):119-39.
41. Abellan JJ, Richardson S, Best N. Use of space–time models to investigate the stability of patterns of disease. Environ Health Perspect 2008;116(8):1111.
42. Duncan EW, White NM, Mengersen K. Bayesian spatiotemporal modelling for identifying unusual and unstable trends in mammography utilisation. BMJ open 2016; 6(5):e010253.
43. Li G, Best N, Hansell AL, Ahmed I, Richardson S. BaySTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice. Biostatistics 2012; 13(4):695-710.
44. Boulieri A, Hansell A, Blangiardo M. Investigating trends in asthma and COPD through multiple data sources: a small area study. Spat Spatiotemporal Epidemiol 2016; 19:28-36.
45. Li G, Haining R, Richardson S, Best N. Space–time variability in burglary risk: a Bayesian spatio-temporal modelling approach. Spat Stat 2014; 9:180-91.
46. Durban JW, Weller DW, Lang AR, Perryman WL. Estimating gray whale abundance from shore-based counts using a multilevel Bayesian model. J Cetacean Res Manage 2015; 15:61-8.
47. Boulieri A, Bennett J, Blangiardo M. Bayesian mixture modelling approach for public health surveillance, Biostatistics 2018 (to appear).
48. Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Kawamoto M, Fujimoto M, Takehara K, Sato S, Hara M, Tsuchiya N. Association of the FAM167A–BLK region with systemic sclerosis. Arthritis Rheumatol 2010; 62(3):890-5.
49. Morin A, Laviolette M, Pastinen T, Boulet LP, Laprise C. Combining omics data to identify genes associated with allergic rhinitis. Clin Epigenetics 2017; 9(1):3.
50. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995; 289-300.
51. Catelan D, Biggeri A. Multiple testing in disease mapping and descriptive epidemiology. Geospat Health 2010; 4(2):219-29.
52. Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. Ann Stat 2003; 31(6):2013-35.
53. Newton MA, Noueiry A, Sarkar D, Ahlquist P. Detecting differential gene expression with a semiparametric hierarchical mixture method. Biostatistics 2004; 5(2):155-76.
54. Muller P, Parmigiani G, Rice K. FDR and Bayesian multiple comparisons rules. In Proceedings Valencia. 2006.
55. Scott JG, Berger JO. An exploration of aspects of Bayesian multiple testing. J Stat Plan Inference 2006; 136(7):2144-62.
56. Whittemore AS. A Bayesian false discovery rate for multiple testing. J Appl Stat 2007; 34(1):1-9.
57. Ventrucci M, Scott EM, Cocchi D. Multiple testing on standardized mortality ratios: a Bayesian hierarchical model for FDR estimation. Biostatistics 2010; 12(1):51-67.
58. Gelman A, Hill J, Yajima M. Why we (usually) don't have to worry about multiple comparisons. J Res Educ Eff 2012; 5(2):189-211.

59. Foreman KJ, Li G, Best N, Ezzati M. Small area forecasts of cause-specific mortality: application of a Bayesian hierarchical model to US vital registration data. J R Stat Soc Ser C Appl Stat 2017; 66(1):121-39.
60. Spiegelhalter DJ, Thomas A, Best NG, Gilks W, Lunn D. BUGS: Bayesian inference using Gibbs sampling. Version 0.5,(version ii) http://www. mrc-bsu. cam. ac. uk/bugs. 1996; 19.
61. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. Stan: A probabilistic programming language. J Statis Softw 2017; 76(1).
62. de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R. Programming with models: writing statistical algorithms for general model structures with NIMBLE. J Comput Graph Stat 2017; Apr 3;26(2):403-13.
63. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Series B Stat Methodol 2009; 71(2):319-92.
64. Ugarte MD, Adin A, Goicoa T. One-dimensional, two-dimensional, and three dimensional B-splines to specify space–time interactions in Bayesian disease mapping: Model fitting and model identifiability. Spat Stat 2017; 22:451-68.
65. Anderson C, Lee D, Dean N. Identifying clusters in Bayesian disease mapping. Biostatistics 2014; 15(3):457-69.
66. Adin A, Lee D, Goicoa T, Ugarte MD. A two-stage approach to estimate spatial and spatio-temporal disease risks in the presence of local discontinuities and clusters. Stat Methods Med Res 2018; 0962280218767975.
67. Ugarte MD, Goicoa T, Etxeberria J, Militino AF. Projections of cancer mortality risks using spatio-temporal P-spline models. Stat Methods Med Res 2012; 21(5):545-60.
68. Corberán-Vallet A, Lawson AB. Prospective analysis of infectious disease surveillance data using syndromic information. Stat Methods Med Res 2014; Dec;23(6):572-90.
69. Diggle P, Rowlingson B, Su TL. Point process methodology for on-line spatio-temporal disease surveillance. Environmetrics 2005; 16:423–434.
70. Diggle P, Moraga P, Rowlingson B, Taylor B. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. Stat Sci 2013; 28(4):542-563.
71. Morrison KT, Shaddick G, Henderson SB, Buckeridge DL. A latent process model for forecasting multiple time series in environmental public health surveillance. Stat Med 2016; 35(18):3085-100.
72. Dai H, Lee BR, Hao J. Predicting asthma prevalence by linking social media data and traditional surveys. Ann Am Acad Pol Soc Sci 2017; 669(1):75-92.
73. Wang Y, Pirani M, Hansell AL, Richardson S, Blangiardo M. Using ecological propensity score to adjust for missing confounders in small area studies. Biostatistics 2017; 1-16.
74. Gehlke CE, Biehl K. Certain effects of grouping upon the size of the correlation coefficient in census tract material. J Am Stat Assoc 1934; 29(185A):169-70.
75. Openshaw S. The modifiable areal unit problem. Concepts and techniques in modern geography. Geobooks 38, Norwich, England 1984;
76. Yiannakoulias N, Svenson LW, Schopflocher DP. An integrated framework for the geographic surveillance of chronic disease. Int J Health Geogr 2009; 8(1):69.
77. Wakefield J, Lyons H. Spatial aggregation and the ecological fallacy. In Handbook of Spatial Statistics. London, CRC Press, 2010: 541-58.

78. Jarup L. Health and environment information systems for exposure and disease mapping, and risk assessment. Environ Health Perspect 2004; 112(9):995.
79. Moraga P. SpatialEpiApp: A Shiny web application for the analysis of spatial and spatio-temporal disease data. Spat Spatiotemporal Epidemiol 2017; 23:47-57.

	FDR	FOR	Sensitivity	Specificity
DM1	0.785 (0.773, 0.800)	0.002 (0.000, 0.006)	0.979 (0.933, 1.000)	0.722 (0.695, 0.744)
DM2	0.191 (0.100, 0.267)	0.026 (0.020, 0.030)	0.660 (0.600, 0.733)	0.987 (0.981, 0.995)
STmix1	0.000 (0.000, 0.000)	0.017 (0.010, 0.024)	0.773 (0.683, 0.867)	1.000 (1.000, 1.000)
STmix2	0.220 (0.167, 0.300)	0.002 (0.000, 0.005)	0.969 (0.933, 1.000)	0.978 (0.969, 0.985)
FlexDetect	0.019 (0.015, 0.031)	0.005 (0.004, 0.006)	0.796 (0.763, 0.827)	1.000 (0.999, 1.000)

Table1: Results of the simulation study to compare the detection performance of disease mapping (DM), the mixture model on the spatio-temporal interaction (STmix1, STmix2) and the mixture model on the spatio-temporal rates (FlexDetect).

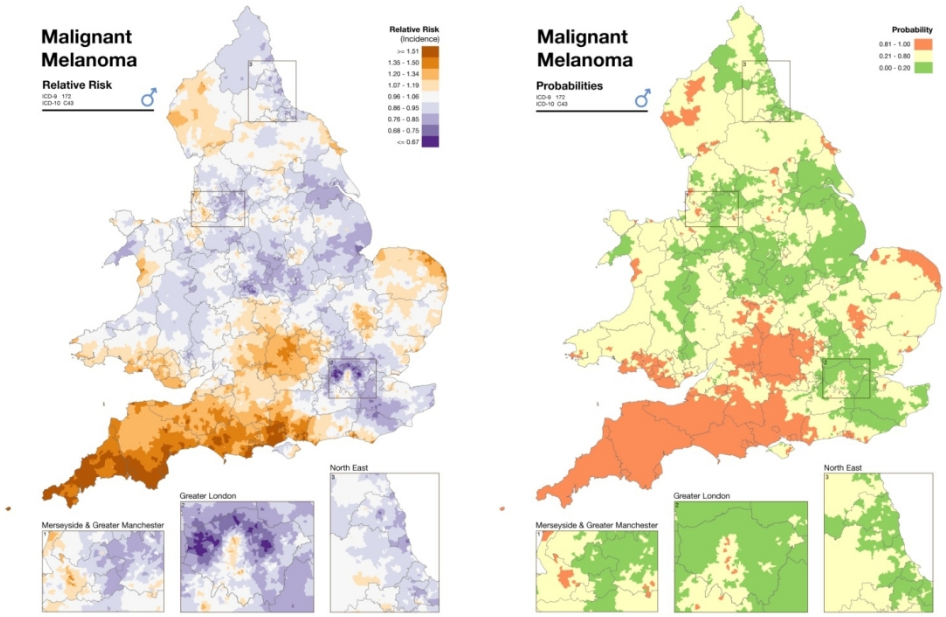

Figure 1: Area specific posterior mean relative risk of malignant melanoma (left) and posterior probability that an area is characterised by a relative risk above 1 (right). Source: Environment and Health Atlas [42].

548x386mm (72 x 72 DPI)

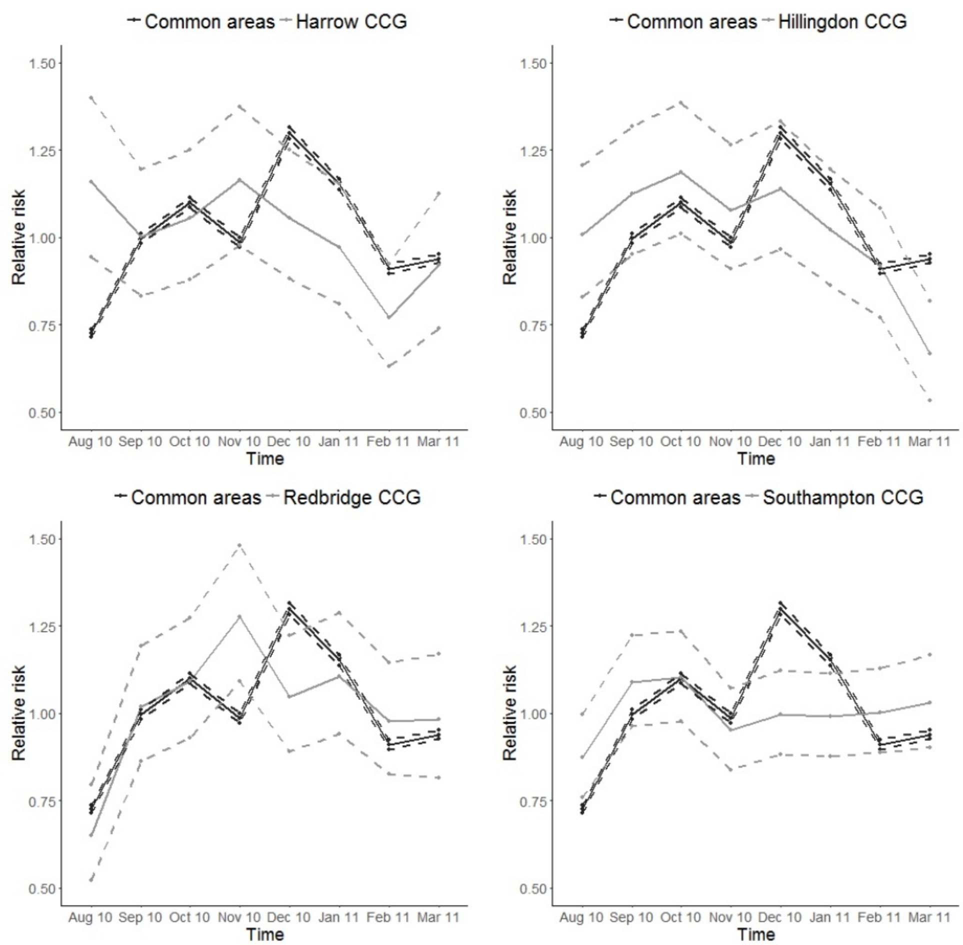

Figure 2: Relative risks and 95% credible intervals of hospital admissions for asthma and COPD for the national (common) temporal trend and for four areas classified as unusual.

377x371mm (72 x 72 DPI)