Comment:
Exceedence probabilities in the analysis of disease maps where risk anomalies are present

In the area of cluster or anomaly detection, exceedence probabilities have been used to assess hot spot clustering in both space and space-time (S.Richardson et al. 2004; Hossain & Lawson 2006; Lawson & Hossain 2008). A variety of Bayesian models for disease risk were evaluated in those papers and it appeared that exceedence probability was a useful method both for evaluating models and for detection of areas of 'unusual' risk . Bayesian modeling has of course many advantages over testing – based methods for anomaly detection such as scan statistics (Kulldorff & Nagarwalla 1995; Kulldorff & Information Management Services Inc 2002). Bayesian modeling allows incorporation of covariates easily, flexible changes to model, and the incorporation of background heterogeneity is possible. It would seem therefore that using exceedence probabilities might become a panacea, as the computational issues associated with fitting Bayesian models via MCMC are much less challenging, especially with packages such as WinBUGS and links to R.

However there are a number of issues with such use, that should be stressed:
The exceedence probability that is usually defined in classic Poisson likelihood models, is a s follows: for the level one data model

$$y_i \sim Pois(e_i\theta_i)$$

$$\log(\theta_i) = \text{model terms}$$

the posterior sampling output from a converged sampler can provide an estimate of $\Pr(\theta_i > c)$. This is usually estimated simply by counting the number of exceedences in the posterior sample and dividing by the sample size (G): i.e.

$\widehat{\Pr}(\theta_i > c) = \sum_{g=1}^{G} I(\theta_i^g > c)/G$. This can be computed easily on WinBUGS and variants can be derived for other forms of risk estimate. Often it is assumed that a guide level for $\widehat{\Pr}(\theta_i > c)$ should be 0.95. Often c assumed to take the value 1. It is sometimes advocated that mapped values of $\widehat{\Pr}(\theta_i > c)$ can be examined for areas of 'excess' risk and thus hot spot clusters can be detected.

There are some issues with the use of these probabilities that should be noted:
1) there is a trade off between the threshold probability chosen and the threshold level that is assumed, i.e. for relative risk θ_i in a small area (i) : $\Pr(\theta_i > c) > \alpha$ requires the assumption of values for c and α and these can be traded. Hence it is not clear what level should be chosen (as alteration of the other level can compensate). This possible equi-finality is somewhat troubling and means that the α level can be altered. The effect detected at different levels of α with different levels of c could to a degree be different and to a degree this is arbitrary. This arbitrariness also means that power to detect effects (ROC analysis) could vary depending on the α and c level chosen. It has been shown

that different forms of risk anomalies can be detected with different thresholds for spatial effects (Hossain & Lawson 2006) and spatio-temporal effects (Lawson & Hossain 2008).

2) much more seriously, exceedence probabilities are *highly* sensitive to model specification or misspecification. The figure below was produced by fitting two different models to the *same* disease incidence data (South Carolina county level congenital anomaly mortality using statewide expected rate 1990 (Lawson 2008), ch 6)).

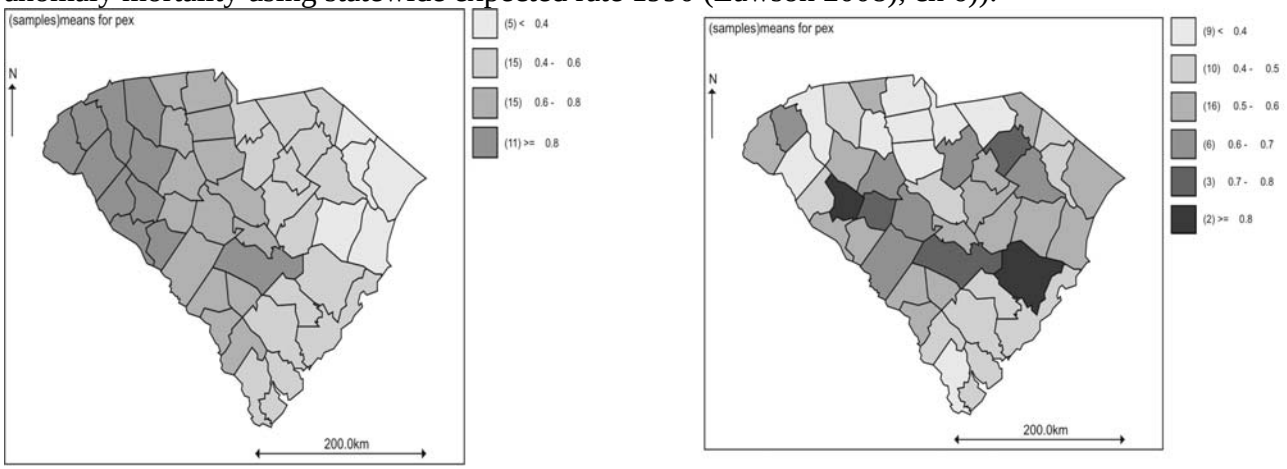

Figure 1 Plots of exceedence probabilities (c=1) generated by the WinBUGS package for the same dataset but different fiitted models: left panel first order spatial trend model, right panel: convolution model with UH and CH components (from Lawson (2008), ch 6).

The two models were a naïve first order spatial trend model with no random effects, (left panel) and a convolution model (right panel). It is clear from this rather extreme example that the exceedence probabilities display quite markedly different patterns and levels and so it is clear also that this can lead to *major* misinterpretation. In this example the trend model gave a marginally *lower* DIC: 171.81 with pD = 2.85 whereas the random effect (convolution) model gave a DIC of 174.46 with pD = 11.57 (after burnin of 10000 with checked convergence and sample size of 10000). This in fact hints at a deeply troubling issue with the models chosen. While prior sensitivity must be addressed in any Bayesian modeling, the fact remains that researchers could happen to use a trend model (based on a goodness-of-fit criterion) and get a completely different set of exceedences compared to random effect models. While this may seem to be a deliberately extreme example, it is clear that model dependence is a very serious issue and means that sensitivity to model assumptions and to the *choice* of models is crucially important, if the results are to be trusted. This example also casts into doubt the use of random effect models, on their own, as a panacea in spatial modeling. It is even more troubling when spatially-referenced covariates are included additively with CAR model components (Ma et al. 2007). Of course, model dependence is also true when other diagnostics are used such as residuals (Kleinman et al. 2004; Vidal-Rodeiro & Lawson 2006).

3) Finally, given sensitivity to models chosen, it might be a concern as to what models are chosen within a study. Tthere is a large range of potential models that could be assumed for different model components (and in particular spatial components). If the selection made is motivated by what can be fitted on WinBUGS rather than any general criteria of relevance then this is a limitation. Even using WinBUGS there are a large

range of possible models that could be assumed for the spatial components (e.g. convolution models, zero-inflatedCAR models, mixture, geostatistical models, to name but a few). In the example above the trend model gave lower DIC and pD than the convolution model and both of these are available on WinBUGS.

In another study of clustering detection capability, Hossain & Lawson (2006) examined three different models for their ability to detect spatial anomalies using exceedence measures, including variants of a local likelihood model and a mixture model (L&C model). They all performed reasonably well in comparison with a standard convolution model under simulation. Some of these models can be fitted on WinBUGS (L&C and convolution) and some not (local likelihood).

However, if I were to fit a different model during a model fitting exercise I could in fact get a different exceedence probability answer, and I would not be able to use the rules as cited.

The fact that exceedence probabilities are highly model dependent is a troubling issue. It means that rules for evaluating these measures cannot be general and must be tailored to specific models and their reliability must be doubted given that models with similar GOF measures can yield markedly different exceedence patterns.

References

Hossain,M. & Lawson ,A.B. 2006. Cluster Detection diagnostics for small area health data. Statistics in Medicine 25: 771-786.

Kleinman,K., Lazarus,R. & Platt,R. 2004. A generalised linear Mixed Models Approach for Detecting Incident Clustrs of Disease in Small Areas with an Application to Biological terrorism. Am. J. Epidemiol. 159: 217-228.

Kulldorff, M. & Information Management Services Inc . SaTScan v. 3.0: Software for the spatial and space-time scan statistics. Bethesda, MD, National Cancer Institute. URL http://srab.cancer.gov/satscan. 2002.
Ref Type: Unpublished Work

Kulldorff,M. & Nagarwalla,N. 1995. Spatial Disease Clusters:Detection and Inference. Statistics in Medicine 14: 799-.

Lawson,A.B. 2008. Bayesian Disease Mapping: Hierarchical Modeling in spatial epidemiology. CRC Press, New York.

Lawson,A.B. & Hossain,M. 2008. Space-time cluster diagnostics with Bayesian small area health models. submitted.

Ma, B., Lawson, A. B. & Liu, Y. Evaluation of Bayesian Models for Focused Clustering in Health Data. Environmetrics 18, 1-16. 2007.
Ref Type: Journal (Full)

S.Richardson, A.Thomson, N.Best & P.Elliott 2004. Interpreting posterior relative risk estimates in disease mapping studies. Environmental Health Perspectives 112: 1016-1025.

Vidal-Rodeiro,C.L. & Lawson ,A.B. 2006. Monitoring Changes in Spatio-temporal Maps of Disease. Biometrical Journal 48: 1-18.