

**Evaluation of fully Bayesian disease mapping  
models in correctly identifying high-risk areas with  
an application to Multiple Sclerosis**

Katia Charland

Degree of Doctor of Philosophy

Department of Epidemiology, Biostatistics and Occupational Health

Faculty of Medicine

McGill University

Montreal, Quebec, Canada

May, 2007

A thesis submitted in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

©Copyright 2007 All rights reserved



Library and  
Archives Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
ISBN: 978-0-494-38569-2

*Our file* *Notre référence*  
ISBN: 978-0-494-38569-2

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.



# Canada



## TABLE OF CONTENTS

TABLE OF CONTENTS .....	iii
LIST OF TABLES .....	v
LIST OF FIGURES .....	vii
STATEMENT OF ORIGINALITY .....	ix
DEDICATION.....	xi
ACKNOWLEDGMENTS .....	xiii
ABSTRACT.....	xv
ABRÉGÉ.....	xvi
CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 BACKGROUND.....	9
2.1 Bayesian methodology for disease mapping .....	9
2.1.1 Introduction to Bayesian inference .....	9
2.1.2 The three-stage hierarchical model for spatial smoothing .....	10
2.1.2.1 The second stage models .....	12
2.1.2.2 The third stage hyperpriors.....	15
2.1.3 Spatio-temporal models .....	15
2.1.4 Implementation of fully Bayesian models .....	16
2.2 Area-level (relative) risk estimates .....	18
2.3 Sensitivity, specificity, Receiver Operating Characteristic curves and predictive values .....	19
2.4 A review of the disease mapping literature .....	22
2.4.1 Empirical Bayesian model and fully Bayesian hierarchical model.....	22
2.4.2 Second level of the three-stage fully Bayesian hierarchical model .....	23
2.4.2.1 Joint modelling of the spatial random effects .....	24
2.4.2.2 Conditional modelling of the spatially structured effect.....	25
2.4.2.3 Semi-parametric models.....	26
2.4.2.4 Spatial Moving Average models.....	27
2.4.2.5 Spatio-temporal models.....	28
2.4.2.6 Studies that evaluate the accuracy of the BYM model in discriminating between high-risk and background-risk areas.....	29
2.4.2.7 The design of the simulation study for assessing accuracy of classification of areas as high-risk or background-risk areas .....	31
CHAPTER 3 SPATIAL SMOOTHING FOR DISEASE MAPPING AND DETECTION OF AREAS WITH ELEVATED RISK.....	35
3.1 Introduction .....	35
3.2 Data and Methods.....	38
3.3 Results .....	46
3.4 Discussion .....	57
CHAPTER 4 COMPARISON OF SPATIAL AND SPATIO-TEMPORAL SMOOTHING MODELS.....	62
4.1 Introduction .....	62
4.2 Data and Methods.....	65
4.2.1 A review of the method of data generation in the assessment of the BYM model .....	65
4.2.2 Generating time series of 'observed' counts for each area .....	67
4.2.2.1 Generating time series to reflect different temporal patterns of risk .....	67
4.2.2.1.1 Scenario 1.....	68
4.2.2.1.2 Scenario 2 .....	69
4.2.2.1.3 Scenario 3 .....	69
4.2.2.1.4 Scenario 4 .....	70
4.2.3 Spatio-temporal models .....	71

4.2.4 Obtaining relative risk estimates for the study period from the results of the spatio-temporal models.....	75
4.2.5 Implementation .....	76
4.2.6 Comparing the accuracy of the models' classification.....	77
4.3 Results .....	77
4.4 Discussion .....	94
<b>CHAPTER 5 APPLICATION TO MULTIPLE SCLEROSIS .....</b>	<b>98</b>
5.1 Introduction .....	98
5.2 Data and Methods .....	100
5.3 Results .....	106
5.4 Discussion .....	122
<b>CHAPTER 6 DISCUSSION .....</b>	<b>126</b>
6.1 Assessing the accuracy of the BYM model.....	127
6.2 Assessing the accuracy of alternative spatial and spatio-temporal models .....	130
6.3 Mapping the incidence of MS in Northern Sardinia .....	132
6.4 Strengths and limitations.....	133
6.4.1 Strengths and limitations of the simulation study .....	133
6.4.2 Strengths and limitations of the spatial analysis of MS incidence .....	135
6.5 Future Research .....	136
<b>APPENDIX A .....</b>	<b>140</b>
A1. Receiver Operating Characteristic curves for the BYM model.....	141
A2. Disease maps of the posterior proportion and the posterior mean for the BYM model.....	142
A3. Disease maps of spatial and spatio-temporal models, Scenario 1.....	154
A4. Confidence Limits for differences in AUCs .....	235
A5. ROC curves for spatial and spatio-temporal models.....	259
A6. Maximum $\hat{R}$ values .....	271
<b>REFERENCES .....</b>	<b>274</b>

## LIST OF TABLES

Table 3.1: Illustration of cross-classification of results from a model's classification of areas using a threshold of 0.65 by true status.....	43
Table 3.2: Area under the ROC curve (AUC), predictive values ( <i>PPV</i> and <i>NPV</i> ), cut-off that minimizes Euclidean distance of (sensitivity, specificity) to the point (1,1) and the corresponding sensitivity and specificity.....	48
Table 3.3: Sensitivity for individual hot spots using a threshold of 0.8.....	55
Table 3.4: Sensitivity for individual hot spots using a RR-sf specific 'optimal' cut-off.....	56
Table 3.5: Maximum $\hat{R}$ values.....	57
Table 4.1: Area under the ROC curve (AUC) for Scenario 1.....	83
Table 4.2: Area under the ROC curve (AUC) for Scenario 2.....	84
Table 4.3: Area under the ROC curve (AUC) for Scenario 3.....	85
Table 4.4: Area under the ROC curve (AUC) for Scenario 4.....	86
Table 4.5: Confidence limits for difference in AUCs for Scenario 1, with hot spot relative risk equal to 1.5 and scale factor equal to 1.....	87
Table 4.6: Confidence limits for difference in AUCs for Scenario 4, with hot spot relative risk equal to 1.5 and scale factor equal to 1.....	88
Table 4.7: Maximum $\hat{R}$ value for a single simulation of Scenario 1.....	92
Table 5.1: Comparison of the distributions of expected counts from the Kentucky cervical cancer simulations and the Northern Sardinia Multiple Sclerosis data.....	107
Table 5.2: Areas identified as hot spots by each model.....	110
Table 5.3: Posterior mean relative risk (RR) and posterior probability (PP) of areas that were classified as 'possible hot spots' by at least one model.....	111
Table 5.4: Posterior mean relative risk (RR) and posterior probability (PP) of areas that were classified as 'probable hot spots' by at least one model.....	112

Table 5.5: Posterior mean relative risk (RR) and posterior probability (PP) of areas that were classified as 'hot spots' by at least one model.....	112
Table 5.6: Deviance Information Criterion (DIC) for spatial smoothing models.....	120
Table 5.7: Deviance Information Criterion (DIC) for spatio-temporal models.....	121

## LIST OF FIGURES

Figure 1.1: John Snow's map of the cholera outbreak in London(1855)....	3
Figure 3.1: Kentucky map with hot spots in blue.....	39
Figure 3.2: <i>Posterior proportion</i> (exceeding 1), averaged over the 50 replications for the case where hot spot RR=1.5 and sf=1.....	47
Figure 3.3: <i>Posterior mean</i> relative risk averaged over 50 replications for the case where the true hot spot RR=1.5, sf=1.....	47
Figure 3.4: Plot of prevalence versus <i>positive predictive value</i> .....	49
Figure 3.5: Graphs of loss functions versus cut-off for the case where the hot spot RR is1.5.....	52
Figure 3.6: Graphs of loss functions versus cut-off for the case where the hot spot RR is 2.....	53
Figure 3.7: Graphs of loss functions versus cut-off for the case where the hot spot RR is 3.....	54
Figure 4.1: Disease maps of the average posterior proportion over 50 replications for Scenario 1 with RR=1.5, sf=1.....	78
Figure 4.2: Disease maps of the average posterior mean over 50 replications, for Scenario 1 with RR=1.5, sf=1.....	80
Figure 4.3: ROC for Scenario 1, for maps with hot spot relative risk equal to 1.5.....	89
Figure 4.4: ROC for Scenario 4, for maps with hot spot relative risk equal to 1.5.....	90
Figure 4.5: Disease maps of PP for log-normal models with optimal threshold.....	91
Figure 5.1: Map of Italy and the Island of Sardinia.....	101
Figure 5.2: Map of the Island of Sardinia with the Province of Sassari in the North.....	101
Figure 5.3: Map of Sassari partitioned into 69 areas.....	103
Figure 5.4: Graphs of the log of the age-specific reference rates versus the log of the area-level age-specific rates for five areas.....	106

Figure 5.5: SMR through time for four areas of Northern Sardinia.....	108
Figure 5.6: Disease map of expected counts.....	109
Figure 5.7: Disease map of crude SMR.....	109
Figure 5.8: Disease map showing all areas that are identified as ‘hot spots’ by at least one model.....	111
Figure 5.9: Disease map of PP for BYM model.....	113
Figure 5.10: Disease map of PP for log-normal model with <i>Gamma</i> (0.001, 0.001) hyperprior for the precision.....	113
Figure 5.11: Disease map of PP for Temporal Linear Trend model.....	114
Figure 5.12: Disease map of PP for Temporal Random Walk model.....	114
Figure 5.13: Disease map of posterior mean relative risk for BYM model.....	115
Figure 5.14: Disease map of posterior mean relative risk for LN1 model.....	115
Figure 5.15: Disease map of posterior mean relative risk for TLT model.....	116
Figure 5.16: Disease map of posterior mean relative risk for TRW model.....	116
Figure 5.17: Disease map of coefficient of linear trend from TLT model.....	117
Figure 5.18: Disease map of PP for BYM model with <i>Gamma</i> (0.001, 0.001) hyperprior for the precision of the unstructured effect.....	118
Figure 5.19: Disease map of PP for log-normal model with <i>Gamma</i> (0.5, 0.0005) hyperprior for the precision of the unstructured effect.....	119
Figure 5.20: Disease map of PP for TLT model with <i>Gamma</i> (0.01, 0.0001) hyperprior for the precision of the unstructured effect.....	119
Figure 5.21: Disease map of PP for TRW model with <i>Gamma</i> (0.01, 0.001) hyperprior for the precision of the random walk and <i>Gamma</i> (0.001, 0.001) hyperprior for the precision of the unstructured effect.....	120

## STATEMENT OF ORIGINALITY

The Besag, York and Mollie (1991) disease mapping model has become the most commonly used Bayesian spatial disease mapping model. Nevertheless, few published studies have assessed its accuracy in classifying areas into high- and background-risk groups. While reporting higher accuracy than many other spatial disease mapping models, all studies were limited due to the small number of replications and/or the inappropriate choice of methods used to assess accuracy. These limitations raise doubts about their conclusions. With the goal of filling in the gaps in the literature, in this thesis, we assess the accuracy of the Besag, York and Mollie (1991) model in identifying high-risk areas with a comprehensive approach using sensitivity, specificity, Receiver Operating Characteristic (ROC) curves and predictive values. Furthermore, these assessments are based on a large number of replications. In light of our findings, we suggest an alternative framework for presenting the results of a disease mapping study for identifying high-risk areas.

Secondly, we assess the accuracy of several alternative spatial and spatio-temporal models to gauge the effects on accuracy of the spatially structured and temporal terms. We consider six models that differ with respect to whether or not they have spatially structured effects, temporal terms and space-time interactions. The spatial models are less complex than the BYM models and were not evaluated in the other simulation studies that assessed the accuracy of classification. For the spatio-temporal models, we propose a new summary statistic that combines, for each area, the smoothed estimates through time, providing an overall area-level measure of relative risk for the study period. The accuracies of classification of the models are compared *via* ROC curves and the area under the ROC curves (AUC). To our knowledge, the classification

accuracies of this set of models have not previously been examined in this way.

Finally, we apply the approaches developed to identify areas of high Multiple Sclerosis risk in the province of Sassari, on the island of Sardinia, Italy. This is the first small-area spatial analysis of Multiple Sclerosis incidence in the province of Sassari.

## DEDICATION

To Alexandra, whose smiles and giggles  
put the seemingly big catastrophes into perspective.



## ACKNOWLEDGMENTS

I'd like to start by thanking my supervisors, Drs. Tina Wolfson and Russell Steele for allowing me enough breathing room to work on what interests me without letting my ideas 'carry me away'. More than once, I've been surprised by the extent of their understanding and support.

I am grateful for the numerous sources of funding that kept me afloat during these years. The McGill Graduate Studies Fellowship, Dr. Tina Wolfson, the Neuroinflammation CIHR Strategic Training Programme Award and the MS Society of Canada research studentship were of great financial help during my five years in the program.

I'd like to acknowledge the Kentucky Cancer Registry for permitting me the use of their data. The Multiple Sclerosis case records and the Northern Sardinian census data were kindly provided by Drs. Maura Pugliatti and Alessandra Sotgiu. I'd like to thank Dr. James Hanley for discussions regarding ROC curves. Donald Dakin and JF Charland must be acknowledged for taking the time to teach me technical tricks that made this large simulation study more manageable. Maria Spina and Marie-Eve Beauchamp were a great help in translating the abstract.

I could have easily abandoned my goals if it weren't for my exceptional support system. My family members, especially my parents, were an invaluable source of encouragement and help. There are no words to express my gratitude. Jacqueline Q. and Marie-Andree G. threw me a life-preserver more times than I can count. Don, my 'huckleberry friend', who simultaneously distracted me from my research and pushed me to meet my objectives, I have to thank him for being goofy and fun when I am way too serious. Most of all, I have to thank Alexandra, the wisest child I have ever known. As a four year old, she taught me that my efforts should be

directed at 'making it better'. This meant, in the context of our conversations, that committing to a plan of action is only worthwhile if it results in a contribution. I've reminded myself of that, more than once, as I have worked on this PhD thesis.

## ABSTRACT

Disease maps are geographical maps that display local estimates of disease risk. When the disease is rare, crude risk estimates can be highly variable, leading to extreme estimates in areas with low population density. Bayesian hierarchical models are commonly used to stabilize the disease map, making them more easily interpretable. By exploiting assumptions about the correlation structure in space and time, the statistical model stabilizes the map by shrinking unstable, extreme risk estimates to the risks in surrounding areas (*local spatial smoothing*) or to the risks at contiguous time points (*temporal smoothing*). Extreme estimates that are based on smaller populations are subject to a greater degree of shrinkage, particularly when the risks in adjacent areas or at contiguous time points do not support the extreme value and are more stable themselves.

A common goal in disease mapping studies is to identify areas of elevated risk. The objective of this thesis is to compare the accuracy of several fully Bayesian hierarchical models in discriminating between high-risk and background-risk areas. These models differ according to the various spatial, temporal and space-time interaction terms that are included in the model, which can greatly affect the smoothing of the risk estimates. This was accomplished with simulations based on the cervical cancer rate of Kentucky and at-risk person-years of the state of Kentucky's 120 counties from 1995 to 2002. High-risk areas were 'planted' in the generated maps that otherwise had background relative risks of one. The various disease mapping models were applied and their accuracy in correctly identifying high- and background-risk areas was compared by means of Receiver Operating Characteristic curve methodology. Using data on Multiple Sclerosis (MS) on the island of Sardinia, Italy we apply the more successful models to identify areas of elevated MS risk.

## ABRÉGÉ

La cartographie d'une maladie montre sur une carte géographique les estimations locales du risque de cette maladie. Lorsque la maladie est rare, les estimations brutes du risque peuvent être variables et mener à des estimations extrêmes dans les zones où la densité de population est faible. Les modèles hiérarchiques bayésiens sont souvent utilisés pour stabiliser les estimations du risque d'une maladie sur une carte géographique, facilitant ainsi l'interprétation.

En exploitant les hypothèses relatives à la structure de corrélation spatiale et temporelle, le modèle statistique stabilise les risques relatifs en rapprochant les estimations instables et extrêmes des estimations des zones géographiques environnantes (lissage spatial local) ou en rapprochant les estimations instables et extrêmes des estimations contigües temporellement (lissage temporel). Les estimations extrêmes qui sont basées sur des populations restreintes sont sujettes à un niveau de rapprochement plus important, particulièrement lorsque les risques dans les zones adjacentes ou les risques contigus temporellement n'appuient pas les valeurs extrêmes et que ces risques sont eux-mêmes plus stables.

Un des buts communs des études de cartographie de maladies est d'identifier les zones ayant un taux élevé de risque. L'objectif de la présente thèse est de comparer la justesse de plusieurs modèles hiérarchiques bayésiens pour distinguer les unités géographiques à haut risque de celles à risque normal. Ces modèles diffèrent par l'inclusion de divers paramètres spaciaux, temporels et d'interaction espace-temps qui peuvent grandement affecter le lissage des estimations de risque. Cette tâche a été effectuée à partir de simulations basées sur le taux de cancer cervical au Kentucky et sur les personnes-années à risque dans les 120

comtés de l'état du Kentucky de 1995 à 2002. Des zones à risque élevé ont été introduites sur les cartes générées dans des zones qui auraient autrement eu, à la base, un risque relatif de un. L'efficacité des divers modèles de cartographie à identifier correctement les zones à risque élevé a été évaluée à l'aide de courbes ROC (« Receiver Operating Characteristic »).

Nous avons utilisé les meilleurs modèles pour identifier les zones caractérisées par un risque élevé de sclérose en plaques en utilisant des données recueillies en Sardaigne, Italie.



## CHAPTER 1 INTRODUCTION

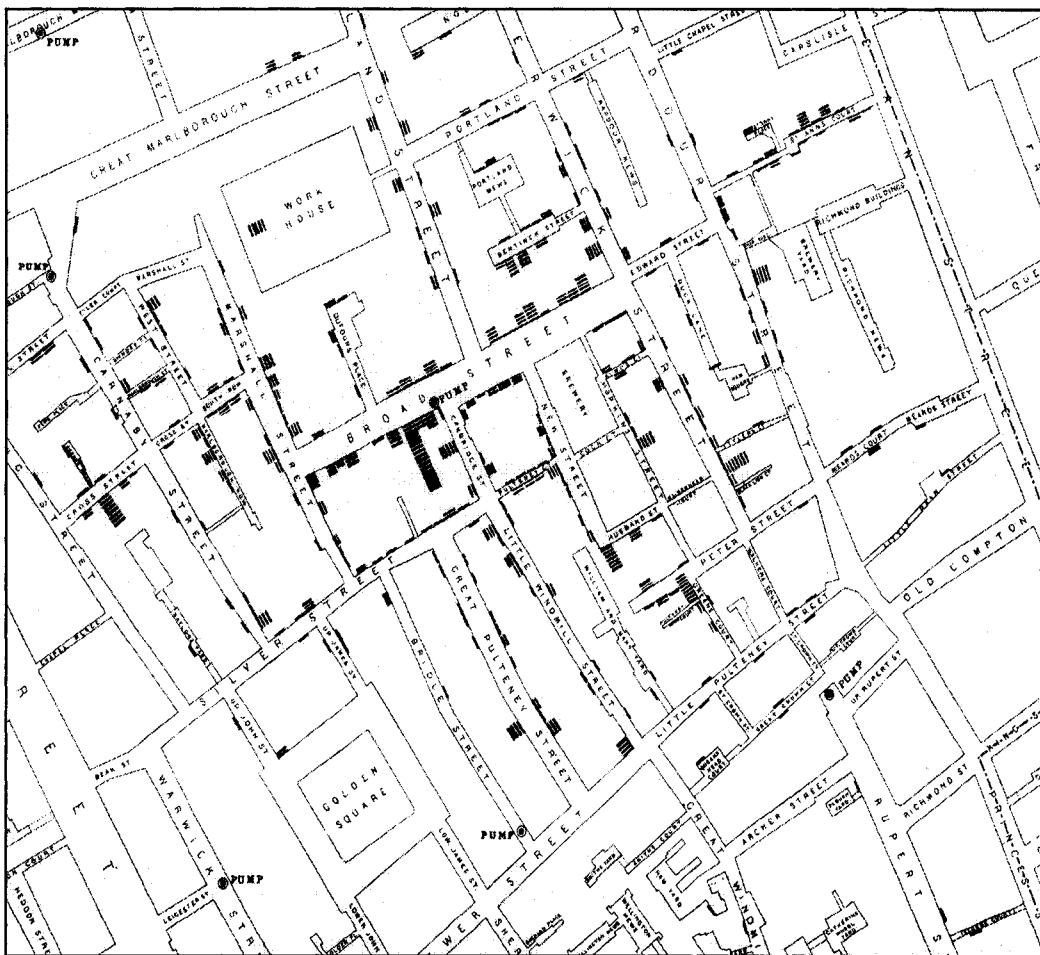
In epidemiologic research, investigators are often interested in the identification of risk factors for the development of disease. Potential risk factors are assessed using observational designs such as case control studies or cohort studies, amongst others, and tentative conclusions are drawn about etiology. However, before such studies can be conducted, potential risk factors must be suggested for investigation. For some diseases, such as Multiple Sclerosis (MS), it is hypothesized that environmental exposures can increase the risk of developing the disease in genetically susceptible individuals (Kahana 2000; Pugliatti et al 2002). In addition, it is further hypothesized that, for these individuals, an environmental trigger initiates the disease process (Granieri, 1997). One method to search for possible risk factors (or triggers) is by examining the spatial variation of disease risk in a large geographical area. In this way, we can gather clues in the search for important environmental risk or protective factors. In particular, if a small area (e.g. county, municipality) appears to correspond to unusually high disease risk (sometimes called a cluster or "hot spot") a closer examination of the inhabitants' characteristics as well as the environment may lead to clues to disease etiology and set the stage for further study.

Detecting hot spots, however, is not a simple task particularly when dealing with a rare disease or sparse populations (i.e. areas in which the population density is low). In these situations, areas that *appear* as hot spots may be areas of true elevated risk or they may simply be a result of random variability. In order to extract the information from the data that is needed to differentiate the *truth* from the random variability, sophisticated statistical techniques are required. The focus of this research is the

evaluation of several statistical models in relation to their accuracy in identifying areas of elevated risk as compared to the background.

To examine the spatial variation in disease risk in geographical regions, researchers use disease maps. Disease maps are geographical maps that show “small-area” risk estimates. For example, a disease map may display the area-level *period prevalence* or the *relative risk (RR)* of developing the disease. The *relative risk* associated with living in a locality is the ratio of the risk of developing the disease given that the individual lives in the area to the risk of developing the disease given that he lives in some *reference* region. The *reference* region may be the study region as a whole or a region distinct from the study region. Disease maps highlight the spatial variation in disease risk and have been used to explore the spatial variation in disease occurrence, generate etiologic hypotheses, and plan resource allocation (Wakefield, 2000a). The disease mapping literature goes back to the 1800’s where street-level maps of yellow fever in the US and cholera in Europe were used to identify risk factors (Walter, 2000). One of the more famous disease maps from that time is that created by John Snow in 1855 which showed the residence-level frequency of cholera deaths in London. This disease map is presented below and each bar on the map represents a cholera death. This map was critical in targeting the source of the water supply as an important risk factor for the occurrence of cholera.

Figure 1.1: John Snow's map of the cholera outbreak in London (1855).  
Each bar corresponds to a death from cholera.



The statistical methods that are used to map diseases use either point level data or areal data. Point level data give the exact geographical coordinates of cases of disease as well as non-cases. John Snow's cholera map was constructed from point level data. Point level data require detailed information on the location of each case and non-case in the area and such data are often not readily available and require extensive resources for collection. We do not discuss methods for point level data in this research, but rather focus on methods for areal data. Areal data arise when the study region is partitioned into areas. In maps using areal data, the data consist of *counts* of cases and at-risk person-

time for each area in the study region. In contrast to point level data, areal data are more commonly available from routinely collected data (e.g. cancer registries, census, etc.). In addition, since the exact locations of the cases are not needed, the privacy of the cases is protected (Olson et al, 2006). Typically, areal data are available for several partitions of a geographic region. For example, disease maps of the United States can show subdivisions by state or by counties within states. Ideally, the choice of map partition should minimize *ecological bias*, where associations observed at the aggregate level do not apply at the individual level due to within-area heterogeneity in risk. To ensure that the disease risk is relatively homogeneous throughout the area (i.e. that area-level risk estimates approximate individual-level risk); it is recommended that the map with the finest subdivision or highest resolution be used (Elliot and Wakefield, 2000). Indeed, choosing a coarse grid may lead to estimates that are, in fact, averages of high and low risk patches.

Although a finer grid is preferred, the resulting risk estimates are subject to increased random variation, particularly in the context of a rare disease or sparse populations (Wakefield et al, 2000a). As a result it may be difficult to decide whether extreme risk estimates indicate an area of true increased (or decreased) risk or if they are a result of extra-variation due to small numbers. For example, consider disease maps showing local *standardized morbidity/mortality ratios (SMR)*. The SMR for area  $i$  is the observed number of cases of the disease in area  $i$  divided by the expected number of cases, say  $E_i$ . The variance of the SMR is equal to  $Y_i / E_i^2$ . Since  $E_i$  is the product of a reference rate of disease and the at-risk person-time in the area, when the disease is rare or the area's population is sparse,  $E_i$  will be small and so the variance of the SMR will be large. This makes it difficult to identify true hot spots since it is not

possible to know if an extreme SMR is due to random variability or a true elevated risk (Wakefield et al, 2000a).

Through the use of more sophisticated statistical methods the effect of random variation can be dealt with to some extent. This is typically done by introducing assumptions of similarity in area-level risk estimates according to proximity in space and time. For example, we may assume that adjacent areas are more alike in their risk or that an area's risk at a point in time is similar to the risk at contiguous time points. The general class of random effects models can incorporate these assumptions. Specifically, Bayesian hierarchical models (or Bayesian "smoothing" models) have been used extensively for disease mapping as a means to increase the precision of the risk estimates by incorporating the risk information from other areas and time points.

Bayesian smoothing models are statistical models designed to remove some of the extra-binomial variation from a disease map. In using these models one can stabilize a disease map by allowing each area to 'borrow strength' from other areas of the map or from the same area at different time points. For example, by assuming that disease risk in a given area is similar to the risk in neighbouring areas, the so-called (*local*) spatial smoothing model shrinks extreme values to a weighted average of the neighbours' risks. *Global* smoothing allows borrowing of strength across non-contiguous areas under the assumption that the area-level risks differ from some mean level without spatial pattern. Although fully Bayesian disease mapping models usually employ spatial smoothing alone, smoothing over time (i.e. temporal smoothing) is also possible. For example, temporal smoothing may involve shrinking an extreme risk estimate to a weighted average of the risk at contiguous time points (i.e. to its temporal neighbours). While it may be advantageous to perform

spatio-temporal analyses in disease mapping studies, spatial analyses are much more common (Nobre et al, 2005).

To stabilize the disease map using spatial smoothing, for each area, we count the total number of cases that were observed in the study period (the observed count) and compute the expected number of cases. The spatial model then produces, for each area, a smoothed relative risk estimate (or smoothed count) for the study period. Spatio-temporal models use observed and expected counts for each time point of the study period to produce a relative risk estimate (or fitted count) for each area at each time point. A clear advantage of applying the spatio-temporal model is that we can examine the evolution of risk through time for each area. There may be another important advantage. For each area of the map, we may estimate the overall relative risk for the study period by summing smoothed counts through time and dividing by the expected count for the study period. This is an estimate of the area's SMR for the study period. This would give a disease map that is like the spatially smoothed map in that each area has a relative risk estimate for the study period. However, there may be increases in precision due to the added (temporal) smoothing, particularly when we suspect that the temporal patterns vary from area to area and we model the space-time interaction.

The objectives of disease mapping studies are varied. For example, the goal may be to accurately estimate the relative risks throughout the map. Another common goal is to correctly identify areas of true elevated risk. Several simulation studies have been conducted to assess the appropriateness of different disease mapping models for specific objectives. Fully Bayesian disease mapping models that perform (*local*) spatial smoothing, such as the Besag, York and Mollie (1991) model have done relatively well in studies using goodness-of-fit criteria (Lawson et al, 2000; Best et al, 2005). However, (*local*) spatial smoothing for the goal of

cluster detection is controversial. A model that performs *local spatial smoothing* may not capture discontinuities in risk because it assumes a smooth risk surface and by doing so, may mask true clusters (Gagnon and Clayton, 2003; Green and Richardson, 2002). Models that smooth spatially may remove some of the extra-variation and produce better estimates in areas where the assumption of a smooth surface holds. However, this increase in stability may result in a loss of accuracy in risk estimation, particularly in isolated hot spots (i.e. high-risk areas that are surrounded by background-risk areas). This concern is reasonable because isolated hot spots result from abrupt changes in the risk surface; a phenomena that is at odds with the assumption underlying *local spatial smoothing*.

Spatio-temporal models do not need to have a spatially structured term for *local spatial smoothing*. Those that do not perform *local spatial smoothing* but, instead, smooth temporally or introduce other assumptions about risk behaviour through time may facilitate hot spot detection while stabilizing the disease map. It is possible that a spatio-temporal analysis is not feasible if for many areas there are too few cases that occur in the study period. Although estimating the area-level relative risks through spatio-temporal modelling may not always be feasible, in some contexts this approach may give a more informative picture of risk behaviour through space and time.

In summary, disease maps with the finest subdivision are preferred because they provide some reassurance that the area-level risk approximates the individual-level risk. However, the map of crude risk estimates can be too unstable when the disease is rare or when the populations are small. Bayesian disease mapping models with temporal and/or spatial smoothing components can be used to reinforce the estimates by removing some of the extra-variation. While some believe

that spatially smoothed disease maps may offer a more cohesive picture of clustering behaviour (Wakefield et al, 2000b), others claim that their underlying assumptions may hinder the identification of hot spots (Gangnon and Clayton, 2003; Green and Richardson, 2002). Spatio-temporal models without the *local* spatial smoothing component may stabilize the disease map while maintaining some of the between-area differentiation.

The purpose of this research is three-fold. First, we evaluate the most popular Bayesian spatial disease mapping model, the Besag, York and Mollie (1991) model in its ability to correctly identify areas of elevated risk using various criteria (i.e. in addition to sensitivity and specificity estimates). Second, in order to assess changes in accuracy of hot spot detection that result from performing *local* spatial smoothing and temporal smoothing we evaluate several competing disease mapping models, which have different spatial and temporal components. Lastly, using data on Multiple Sclerosis on the island of Sardinia, Italy we apply the more successful models to map the relative risk of Multiple Sclerosis.

This thesis is organized as follows: in chapter 2 we summarize some basic Bayesian concepts, describe three-stage Bayesian hierarchical models and Receiver Operating Characteristic curve methodology and review the Bayesian disease mapping literature. In chapter 3 we report the results of a simulation study designed to assess the accuracy of the Besag, York and Mollie (1991) model in correctly identifying high-risk areas and, in chapter 4, we assess the changes in accuracy that result from including different spatial and temporal terms. The results of applying disease mapping models to the problem of identifying hot spots of Multiple Sclerosis in Northern Sardinia, Italy are presented in chapter 5. Finally, in chapter 6, we discuss the conclusions of the research, strengths and limitations and future directions.

## CHAPTER 2 BACKGROUND

In this chapter, we introduce some of the fundamental concepts of Bayesian statistics and describe how they are applied to disease mapping studies through spatial and spatio-temporal smoothing methods. In addition, we briefly review several measures of diagnostic accuracy, in particular, Receiver Operating Characteristic (ROC) curves, one of the means by which model accuracy is compared in this thesis. The final sections of chapter 2 review the various disease mapping methods and the simulation studies that have examined the accuracy of the Besag, York and Mollie (1991) (BYM) model.

### 2.1 Bayesian methodology for disease mapping

#### 2.1.1 Introduction to Bayesian inference

Let  $\theta$  be a, possibly vector valued, parameter of interest and assume that we have  $n$  observations  $y_1, y_2, \dots, y_n$ . The *prior distribution* for  $\theta$  is described by a probability density (or mass function) that summarizes our knowledge about the parameter before the study has begun (i.e. before we have examined the data). This knowledge can be truly subjective or may be based on past experience. Let  $p(y_1, y_2, \dots, y_n | \theta)$  be the likelihood function for  $\theta$ ,  $p(\theta)$  be the prior density of  $\theta$  and

$p(y_1, y_2, \dots, y_n) = \int p(y_1, y_2, \dots, y_n | \theta) p(\theta) d\theta$  be the marginal density at  $y_1, y_2, \dots, y_n$ . Then by Bayes' Theorem we obtain an expression for the *posterior distribution* of  $\theta$ :

$$p(\theta | y_1, y_2, \dots, y_n) = \frac{p(\theta) p(y_1, y_2, \dots, y_n | \theta)}{p(y_1, y_2, \dots, y_n)} .$$

The *posterior distribution* summarizes the information about  $\theta$  that is provided by both the *prior distribution* and the data. The parameters of the prior distribution are called *hyperparameters* and in a fully Bayesian formulation each of the *hyperparameters* is assigned a distribution called a *hyperprior* distribution.

### 2.1.2 The three-stage hierarchical model for spatial smoothing

Throughout this research we focus on statistical models that stabilize disease maps of the relative risk (RR). The relative risk is generally defined as the ratio of the cumulative incidence rate of disease in the exposed group to the cumulative incidence rate in the unexposed group. Although, the term relative risk is used extensively in relation to the models we are investigating in this research, the quantities we are estimating are also referred to as risk ratios or standardized morbidity ratios (i.e. SMRs) (MacNab and Dean, 2002). We are essentially comparing the disease rate in an area to the disease rate of some reference region. The ratio of these rates is an estimate of the SMR associated with living in a particular area, which is the number of cases observed in the time period of interest divided by the expected number of cases, using as a reference rate the incidence rate of the disease of the study region or a region distinct from the study region. In the discussion that follows, we use the terminology that is commonly used in the literature, but it should be kept in mind that any reference made to the ‘relative risk’ can be thought of as the SMR.

Let  $Y_i$  represent the observed number of cases in area  $i$ ,  $RR_i$  the relative risk for area  $i$  and  $E_i$  the expected count in area  $i$ . A commonly used three-stage Bayesian hierarchical model assumes that, for the first stage

or level, the observed count  $Y_i$  has a Poisson distribution with mean  $RR_i \cdot E_i$ . Note that the Poisson distribution is used as an approximation to the Binomial distribution and is appropriate when the disease is rare and the population is large. If this is not the case the Binomial distribution must be used (Hogg and Tanis, 1988). When the disease is rare, using the Poisson distribution allows us to collapse over strata to model the overall number of cases per area. Smoothing models are typically used to stabilize risk estimates when the disease is rare and the population is sparse so we focus on the Poisson distribution throughout the thesis. The  $\log(RR_i)$  is modeled as the sum of terms representing spatially structured and unstructured random effects plus terms for important covariates. The spatially structured term reflects an assumption of smooth changes in risk as we move between adjacent areas. Thus, at the second stage of the hierarchy, the spatially structured term is typically assigned a prior distribution that allows for spatial dependence amongst neighbouring small-area log relative risk values, so that areas that are closer to one another are more alike in their estimates. This prior allows for *local* spatial smoothing of risks. This term can be interpreted as the effect (i.e. on  $\log(RR_i)$ ) of unmeasured covariates that display spatial structure. The spatially unstructured term assumes heterogeneity (or variability) that does not display spatial structure. The unstructured term is assigned a prior distribution that allows for *global* smoothing.

The specification of the model is as follows:

$$\begin{aligned}
Y_i &\sim \text{Poisson}(RR_i \cdot E_i) \\
\log(RR_i) &= \alpha + U_i + V_i + \beta X_i \\
U_i \mid U_{j \neq i} &\sim f_U(\cdot \mid \tau_U) \\
V_i &\sim f_V(\cdot \mid \tau_V) \\
i &= 1, \dots, n
\end{aligned}$$

where,  $U_i$  represents the spatially structured effect,  $V_i$  the spatially unstructured effect, and the  $\tau$ .s are *hyperparameters* that represent measures of dispersion, such as the variance or precision. The intercept,  $\alpha$ , represents the overall log relative risk of disease in the study region as compared to the reference rate. The  $X_i$  is a covariate and  $n$  is the number of areas in the study region.

### 2.1.2.1 The second stage models

A commonly used prior for the spatially *unstructured* effects ( $V_1, V_2, \dots, V_n$ ) is the Multivariate Normal Distribution (MVN) with mean  $\theta_n$  and covariance matrix  $\sigma_v^2 I$ . If  $V$  is the  $n \times 1$  vector of spatially unstructured terms then

$$V \sim \text{MVN}(\theta_n, \sigma_v^2 I)$$

where  $\theta_n$  is an  $n \times 1$  vector of zeros,  $\sigma_v^2 > 0$  and  $I$  is the  $n \times n$  identity matrix. Here  $\sigma_v^2$  controls the variation in the log relative-risks throughout the study region. Its effect is to control the amount of *global* smoothing that takes place. One common alternative to the MVN distribution is the Gamma distribution.

To model the dependence in the spatially structured terms  $U_i$ , either the joint distribution of the  $U_1, U_2, \dots, U_n$  or the set of conditional distributions of  $U_i$  given  $U_j = u_j$ ,  $j \neq i$ ,  $i, j = 1, 2, \dots, n$  may be specified. The conditional formulation described here is referred to as the class of Conditional Autoregressive (CAR) Models.

Suppose that  $U$ , the  $n \times 1$  vector of  $U_i$ 's, is modeled by a multivariate normal distribution with  $n \times 1$  mean vector  $\mu$  and covariance matrix,  $\sigma_U^2 \Sigma$ , where  $\Sigma$  is a  $n \times n$  positive definite, between-area correlation matrix. The factor  $\sigma_U^2$  controls the amount of *local smoothing*. The correlation matrix  $\Sigma$  can be re-written as  $(I - \gamma \cdot C)^{-1} M$ , where  $M$  is a  $n \times n$  diagonal matrix whose entries,  $M_{ii}$ , are proportional to the conditional variance of the  $U_i$  given  $U_j$ . The  $C$  matrix has dimension  $n \times n$  and its entries,  $C_{ij}$ , are a measure of the spatial dependence between areas  $i$  and  $j$ , while  $\gamma$  determines the strength of spatial dependence. This joint specification leads to the following set of conditional distributions for  $U_i$  given  $U_j$ :

$$U_i | U_j = u_j \sim N(\mu_i + \Sigma_j \gamma \cdot C_{ij}(u_j - \mu_i), \phi \cdot M_{ii})$$

where  $\phi$  corresponds to the amount of the total variability that is due to spatially structured variation (Thomas et al, 2004, Besag and Kooperberg, 1995).

To use the joint formulation a symmetric positive definite matrix  $\Sigma$  must be selected. For the conditional specification a choice of parameter,  $\gamma$ , and matrices,  $M$  and  $C$  is required. Constraints on  $M$ ,  $C$  and  $\gamma$  must be respected to ensure that the corresponding correlation matrix,  $\Sigma$ , is symmetric and positive definite. The fact that the entries of  $M$  represent variances implies that  $M_{ii}$  must be positive. The symmetry of  $\Sigma$  requires that  $C_{ij}M_{jj} = C_{ji}M_{ii}$  while the positive definite criterion for  $\Sigma$  constrains the

$\gamma$  to lie between the inverses of the largest eigenvalues of  $M^{-1/2}CM^{1/2}$  (Thomas et al, 2004).

A commonly used second stage CAR prior, called the Intrinsic CAR prior, was developed by Besag, York and Mollie (1991). Let  $n_i$  be the number of areas sharing a border with area  $i$  (i.e. a neighbour). For the Intrinsic CAR prior we set  $C_{ij}$  equal to  $1/n_i$  if areas  $i$  and  $j$  are neighbours. If areas  $i$  and  $j$  are not neighbours or if  $j = i$  then  $C_{ij} = 0$ . For the  $M$  matrix,  $M_{ii}$  is equal to  $1/n_i$  while the spatial dependence parameter  $\gamma$  is set to the largest possible value of 1. The conditional distributions that correspond to this choice of  $M$  and  $C$  matrix are:

$$U_i \mid U_j = u_j, j \neq i \sim N\left(\bar{u}_i, \frac{\omega_u^2}{n_i}\right).$$

The conditional mean  $\bar{u}_i$  is the mean of the  $u_j$  of neighbouring areas. It is important to note that the choice of  $M$  and  $C$  matrix for the Intrinsic CAR leads to a joint specification with a correlation matrix,  $\Sigma$ , that is not positive definite. The variance  $\omega_u^2$  is not proportional to  $\sigma_u^2$  and, thus, is only interpreted as a conditional variance. In addition, this model does not have a fixed overall mean level for  $U_i$ ; however, identifiability can be achieved in the model with an intercept by constraining the sum of the  $U_i$  to zero and assigning a uniform prior on the real line for the intercept. Alternatively, a model without intercept may be used (Wakefield et al, 2000a; Besag and Kooperberg, 1995).

### **2.1.2.2 The third stage hyperpriors**

In the fully Bayesian formulation, parameters from the second stage models are assigned distributions. For the model with both local and global spatial smoothing, the conditional variance for the  $U_i$  and the marginal variance of the  $V_i$ ,  $\omega_U^2$  and  $\sigma_V^2$  (or the corresponding precisions  $\tau_U$  and  $\tau_V$ ), respectively, are assigned hyperpriors. Gamma distributions are typically used for the precisions. If  $Gamma(a, b)$  represents a Gamma distribution with mean  $a/b$  and variance  $a/b^2$ , a common choice has been to assign small values to  $a$  and  $b$ , for example,  $a=b=0.001$  (Wakefield et al, 2000a; Pascutto et al, 2000). This choice of prior does not place enough prior mass on random effect standard deviation values close to zero and, thus, may not be appropriate when there is little variation amongst the random effects (Wakefield et al, 2000a). As an alternative, Kelsall and Wakefield (1999) recommend a  $Gamma(0.5, 0.0005)$  as it results in a reasonable range of relative risk values.

### **2.1.3 Spatio-temporal models**

In the spatial smoothing model, spatially structured effects and spatially unstructured effects lead to *local* and *global* smoothing of crude risk estimates, respectively. Terms representing temporal effects and space-time interaction can be added to model each area's risk through time. Two such models have been presented by Knorr-Held and Besag (1998) and Bernardinelli et al (1995a).

Knorr-Held and Besag (1998) model the log-odds as the sum of spatially structured and unstructured effects and temporal effects, plus the effect of covariates. The temporal effects are assumed to follow a random walk

model. The spatially structured term is assigned an Intrinsic CAR prior and the unstructured effects are modeled as independent normal random variables each with mean zero and fixed variance. The authors state a preference for random walk priors for the temporal terms because they allow greater flexibility as compared to a linear trend model or a model with autoregressive priors such as the AR(1), where the state at time  $t$  is regressed on the state at time  $t-1$ . Such autoregressive priors require stationarity which generally means that there are no systematic changes in mean level or variance and no periodic fluctuations.

Bernardinelli et al (1995a) model the log relative risk as a linear trend model with either spatially structured or spatially unstructured effects for the intercept and coefficient of the linear trend. When employing spatially structured effects, the intercept of the model is the sum of a term representing the baseline log relative risk and a spatially structured term which is assigned an Intrinsic CAR prior. The coefficient of the linear trend is the sum of a term representing the overall trend in the study region and a term for the area-specific *differential trend*. The term of the coefficient that represents the *differential trend* is typically assigned an Intrinsic CAR prior. However, Bernardinelli et al (1995a) recommend that the spatially structured random effects be replaced with spatially unstructured random effects if we have reason to believe *a priori* that cluster size (i.e. the size of the hot spot) will be no larger than the area size. Although the model is restrictive due to its assumption of linearity, it does allow for space-time interaction.

#### **2.1.4 Implementation of fully Bayesian models**

For many specifications of the Bayesian disease mapping models an analytically tractable solution does not exist, thus, an empirical estimate of the *posterior distribution* of the parameters is typically obtained through

Markov Chain Monte Carlo (MCMC) techniques. A Markov Chain is a sequence of random variables indexed by time, say  $\{X_t, t = 1, 2, 3, \dots\}$ , such that  $X_{t+1}$  is conditionally independent of  $X_{t-1}, X_{t-2}, \dots, X_1$  given  $X_t$ . Monte Carlo integration is used to estimate the mean of a random variable by independently sampling from the distribution of the random variable and averaging over the sampled values. Markov Chain Monte Carlo techniques are simulation-based techniques that draw dependent samples from a probability distribution, thus producing a Markov Chain that serves as a sample from the distribution of interest. In the Bayesian framework we are often interested in sampling from the posterior distribution of the parameter of interest to produce an empirical posterior distribution from the chain of estimates that are sampled. As this is an iterative procedure, initial values for the parameters are required to start the process and sometimes many thousands of iterations are needed before the process converges. Once the process has converged (i.e. has become independent of the choice of initial values), the iterations before convergence are discarded leaving a collection of estimates for each parameter. This collection forms a sample from the *posterior distribution* of the parameter called the *posterior sample*. It is recommended that several MCMC chains be produced from different starting values in order to better assess dependence on the initial values. The segments of the chains, produced after convergence, are aggregated to form a single *posterior sample*. Often MCMC techniques are implemented using a publicly available software package called WinBUGS which uses a variety of MCMC techniques for Bayesian applications (Spiegelhalter DJ et al, 2002). WinBUGS can be used in conjunction with the flexible R software (R Development Core Team, 2005) in order to conduct Bayesian analyses.

As described earlier, to ensure that the *posterior sample* is independent of the initial values used to start the process, the *posterior sample* includes only those estimates that are produced after convergence of the MCMC chain has been achieved. There are several measures available to assess convergence. One such measure is the  $\hat{R}$  statistic (Gelman et al, 2003).  $\hat{R}$  compares the variances within each MCMC chain to the variance between the chains. It basically assesses whether the empirical distributions from each chain are approximately the same as the empirical distribution that is formed by combining the estimates from all chains.  $\hat{R}$  approaches the value of 1 when the first half of the chain has converged leaving the remainder to form the *posterior sample* (Gelman et al, 2003).  $\hat{R}$  values are automatically included in the numerical output when running hierarchical models in WinBUGS through R (Gelman et al, 2003).

## 2.2 Area-level (relative) risk estimates

As described in section 2.1.4, the MCMC technique produces a *posterior sample* for the risk measure of interest. From this, several different maps can be produced based on empirical (summary) measures of the *posterior sample* but, in general, there are two types of maps that are presented. The first is a map showing the mean of the *posterior distribution* (i.e. *posterior mean* or *median*) of the area's relative risk. The second is a map showing the proportion of relative risk estimates in our *posterior sample* that exceed a reference value. Often the reference value for the relative risk will be the null value 1. This proportion is called the *posterior proportion* and is abbreviated as PP.

## **2.3 Sensitivity, specificity, Receiver Operating Characteristic curves and predictive values**

There are many measures that are used to judge the accuracy of a classification procedure. In this section we discuss sensitivity and specificity, Receiver Operating Characteristic (ROC) curves, and predictive values.

The sensitivity of a procedure refers to the probability of correctly classifying a true positive as positive. The specificity is the probability of correctly classifying a true negative as such. In evaluating the accuracy of a diagnostic test or procedure we estimate the sensitivity by the proportion of true positives that are classified as positive by the test and we estimate the specificity by the proportion of true negatives that are classified as negative by the test. When the result of a test or procedure is dichotomous i.e. the result is 'positive' or 'negative', the sensitivity and specificity can be directly estimated. When the result is on a continuous scale, a cut-off that distinguishes a 'positive' from a 'negative' must be established. Once we use the cut-off to determine which results are 'positive' and which are 'negative', we can find the corresponding sensitivity and specificity.

Clearly, the choice of cut-off for classifying results as either positive or negative affects the procedure's classification accuracy, in that, if we allow the cut-off to vary, the proportion of true positives and true negatives will also vary. Thus, the sensitivity and specificity are cut-off dependent. Another means of assessing the accuracy of a classification procedure that considers several cut-off values is the Receiver Operating Characteristic (ROC) curve. A ROC curve is the graph of the proportion of true positives against the proportion of false positives that is formed when we vary the cut-off that distinguishes a positive from a negative

classification. If the test is completely non-informative about the true classification, then varying the cut-off will yield a curve that looks like the graph of  $y=x$ . This is statistically equivalent to tossing a fair coin to determine the test result. If the procedure is more successful, the curve will again start at the origin  $(0,0)$  but, as the cut-off changes, the graph will rise above the line of  $y=x$  and continue to the point  $(1,1)$ . For a highly successful procedure the ROC curve is ‘tucked up’ in the upper left corner of the graph. The area under the ROC curve (AUC) is used as a summary measure of the accuracy of the procedure (Hanley and McNeil, 1982). The more accurate the procedure the closer the AUC will be to the maximum value of one. A less accurate procedure will have AUC closer to 0.5 (i.e. the area underneath the line  $y=x$ ).

To compare the accuracies of two procedures, a test statistic was developed by Hanley and McNeil (1982) that tests the significance of differences in AUCs. If the classification procedures are applied to different sets of objects (e.g. different sets of patients), then we have that

$$z = (AUC_1 - AUC_2) / \sqrt{se^2(AUC_1) + se^2(AUC_2)}$$

approximately follows a normal distribution with mean 0 and variance 1, where  $AUC_1$  and  $AUC_2$  are the AUCs of the procedures, and  $se(AUC)$  is the standard error of the AUC.

If the same set of objects is used to test the two procedures then the denominator is replaced by

$$\sqrt{se^2(AUC_1) + se^2(AUC) - 2r \cdot se(AUC_1) \cdot se(AUC_2)}$$

where  $r$  represents the correlation induced by using the same objects for both procedures. Formulae for  $r$  and  $se(AUC)$ , and further details can be found in Hanley and MacNeil (1982), Hanley and McNeil (1983) and the appendix to Hanley and McNeil (1983).

While the sensitivity and specificity provide valuable information on how well the classification procedure performs on known positives and negatives they cannot provide a complete picture of the classification procedure's utility. It is important to know how likely it is that a positive classification corresponds to a true positive or that a negative classification corresponds to a true negative. That is, we want to know the probability of a true positive given a positive result of the classification procedure and the probability of a true negative given a negative result of the classification procedure. These probabilities describe the positive and negative predictive values of the procedure, respectively. In the context of a diagnostic test classifying a patient as having a disease or not, the positive predictive value is the probability of disease given a positive test result, while the negative predictive value is the probability of not having the disease given a negative test result. In the disease mapping context, we are interested in the probability that an area that is classified as high-risk by the disease mapping model corresponds to a true high-risk area and the probability that an area that is classified as background-risk corresponds to a true background-risk area.

The positive and negative predictive values depend on the sensitivity and specificity of the test as well as the prevalence (or proportion) of 'positives' in the population. Let  $PPV$  represent the positive predictive value,  $sens$  the sensitivity,  $spec$  the specificity and  $prev$  the prevalence. An expression relating the  $PPV$  to  $sens$ ,  $spec$  and  $prev$  is as follows:

$$\text{logit}(PPV) = \log\left(\frac{sens}{1 - spec}\right) + \text{logit}(prev). \quad (2.1)$$

We see from this expression that when the prevalence is small,  $\text{logit}(prev)$  is negative and will dominate  $\log\left(\frac{sens}{1 - spec}\right)$ , resulting in a negative value for  $\text{logit}(PPV)$  (van Belle, 2002). When the logit of the  $PPV$  is negative,  $PPV$  is less than 50%. This means that when the prevalence of positives is small, the probability that a positive test indicates a true positive state is low. Thus, the *predictive value* of a test is highly dependent on the proportion of true positives or in the disease mapping context, the proportion of areas that are ‘high-risk’. Similarly, when the prevalence of negatives is low the  $NPV$  will be low but when the prevalence of negatives is high the  $NPV$  will be higher.

## 2.4 A review of the disease mapping literature

Since the mid 1980’s, there has been a great deal of research oriented towards improving the methodology used in disease mapping in order to produce more reliable and interpretable disease maps. In this section several classes of disease mapping methods are reviewed.

### 2.4.1 Empirical Bayesian model and fully Bayesian hierarchical model

As discussed earlier, high-resolution disease maps result in areas with small populations and, in the case of a rare disease, conventional standardized morbidity ratios (SMRs) will be subject to increased

variability (Wakefield et al, 2000a). This effect of sample size on the variance of the SMRs makes comparisons between areas difficult.

In order to address this problem, Clayton DG and Kaldor (1987) suggested an Empirical Bayes formulation to shrink extreme crude relative risk estimates (i.e. SMR) towards the average RR estimates of neighbouring areas. Empirical Bayes methodology uses the data at hand to compute point estimates of the parameters of the prior distributions of the random effects in the expression of the log(RR). By failing to account for the uncertainty about these parameters, the overall uncertainty associated with the RR estimates can be underestimated. Subsequently, Besag, York and Mollie (1991) and Clayton DG and Bernardinelli (1992) proposed a fully Bayesian three-stage hierarchical model. In this model, the parameters of the prior distributions of the random effects (i.e. the *hyperparameters*) are assigned a (*hyperprior*) distribution to reflect the uncertainty in the parameter values.

#### **2.4.2 Second level of the three-stage fully Bayesian hierarchical model**

At the second-level of the fully Bayesian hierarchy, several classes of models have been suggested. These models are discussed in an excellent review paper by Best et al (2005) and readers are referred to that paper for a more complete description. The following discussion is largely based on their paper.

#### 2.4.2.1 Joint modelling of the spatial random effects

As described earlier, the spatially structured random effects may be jointly modeled with a multivariate normal distribution. Typically, the corresponding between-area correlation matrix is modeled as a function of the distance between pairs of areas. For example, letting  $d_{ij}$  be the distance between the centroids of area  $i$  and area  $j$ , the exponential decay function  $\exp(-\varphi \cdot d_{ij})$ , for  $\varphi > 0$ , expresses the correlation between two area-level risks. Here,  $\varphi$  represents the rate of decrease in correlation with increasing distance. This form of correlation matrix leads to a positive definite covariance matrix. One disadvantage of using this joint specification is the computationally expensive requirement of matrix inversion at each step of the MCMC algorithm. As reported by Best et al (2005) in their evaluation study, another disadvantage of this model is its tendency to oversmooth the risk estimates. *Oversmoothing* occurs when too much between-area differentiation is lost due to the amount of shrinkage that takes place. An alternative to modelling the covariance matrix in the joint specification is modelling the precision matrix (i.e. the inverse of the covariance matrix) (MacNab YC, 2003). This specification does not require matrix inversion resulting in a savings in computational resources.

Yet another alternative has been proposed by Kelsall and Wakefield (2002) where the log of the relative risk is modeled as a Gaussian Random Field, a formulation that has been commonly used in the field of Geostatistics. This approach models the covariance between the spatially structured random effects for areas  $i$  and  $j$  as the product of the marginal variance and the average correlation between two randomly chosen points from areas  $i$  and  $j$ . Essentially, this formulation results in a continuously

varying risk surface which may be a more appropriate approach in some contexts.

#### 2.4.2.2 Conditional modelling of the spatially structured effect

The Conditional Autoregressive (CAR) models are another class of models that conditionally specify the distribution of the spatially structured terms. The Intrinsic Conditional Autoregressive Model put forth by Besag, York and Mollie (1991) and the Proper Conditional Autoregressive model by Cressie and Chan (1989) are two such models. If the correlation matrix of the joint distribution is expressed as  $(I - \gamma \cdot C)^{-1} M$ , the conditional distribution of the spatially structured term for area  $i$  is

$$U_i | U_j = u_j, j \neq i \sim N(\mu_i + \Sigma_j \gamma \cdot C_{ij}(u_j - \mu_j), \phi \cdot M_{ii}).$$

As described in section 2.1.2.1, the Intrinsic CAR model corresponds to

$$C_{ij} = \begin{cases} 1 / n_i & \text{if areas } i \text{ and } j \text{ are neighbours } (i \neq j) \\ 0 & \text{otherwise} \end{cases}$$

and  $M_{ii} = 1 / n_i$  where  $n_i$  is the number of areas sharing a border with area  $i$ . The spatial dependence parameter  $\gamma$  is set to 1, the largest possible value. In contrast to the Intrinsic CAR, the Proper CAR results in a joint distribution with a positive definite correlation matrix. The  $C$  and  $M$  matrices are composed of entries:

$$C_{ij} = \begin{cases} (E_j/E_i)^{1/2} & \text{if areas } i \text{ and } j \text{ are neighbours } (i \neq j) \\ 0 & \text{otherwise} \end{cases}$$

$$M_{ii} = 1 / E_i$$

where  $E_i$  is the expected number of cases in area  $i$ .

As an alternative to an Intrinsic CAR prior based on the normal distribution, Besag et al (1991) propose a double exponential distribution as it is thought to result in less smoothing. Richardson et al (2004), in their simulation study, found that this was true under certain conditions but that in general the performance of the two models was similar (Richardson et al, 2004). Pascutto et al (2000) compared the two and found the results virtually identical.

#### 2.4.2.3 Semi-parametric models

*Semi-parametric* models were developed to address the problem of oversmoothing encountered when using the above parametric models. These models allow both *local* smoothing and the borrowing of strength from non-adjacent areas that have been judged, by the model, to have similar relative risk. Estimates of the spatially structured effects are found by taking a weighted average over several cluster configurations where the weights are the *posterior probabilities* corresponding to the cluster configuration. Green and Richardson (2002), Knorr-Held and Raßer (2000), Denison and Holmes (2001) and Gangnon and Clayton (2000) have developed semi-parametric spatial models.

Although the semi-parametric models were developed to resist *oversmoothing*, the simulation study by Richardson et al (2004) concluded that, overall, the BYM model and the semi-parametric model proposed by Green and Richardson (2002) were both highly specific but not sensitive in their ability to correctly classify areas as background or hot spot. Similarly, Best et al (2005) note that in their simulation study, the Knorr-Held and

Raßler (2000) model and the BYM model had a greater tendency to oversmooth compared to the Green and Richardson semi-parametric model, yet the Green and Richardson model was not as accurate in correctly classifying areas as hot spot or background.

#### **2.4.2.4 Spatial Moving Average models**

The Spatial Moving Average models are yet another class of Bayesian models originally developed for continuous processes in the field of Geostatistics (Best et al, 2005). As they are meant for continuous processes, they are not commonly used for disease mapping. Best et al (2000) and Wolpert and Ickstadt (1998), however, have implemented a discrete version for their respective identity-link Poisson regression models. Their modification is based on a gamma moving average process. These models work by integrating a white noise process in two-dimensional space *via* a smoothing kernel that depends on distance and sometimes location. The result is a smooth risk surface. One of the advantages of this model over the modelling of the covariance matrix (section 2.4.2.1) is its savings in computational costs due to the fact that matrix inversion is unnecessary. Another feature is the flexibility of the model in that we can choose from a variety of kernel smoothers to capture important characteristics of the spatially structured variation (Best et al, 2005). Best et al (2005) found that the Spatial Moving Average model did not fit the data as well as the BYM model and the semi-parametric models of Knorr-Held and Raßler (2000) and Green and Richardson (2002). They note, however, that this may be partly due to the fact that the data were simulated under a multiplicative model which is an underlying assumption for the other models but the Spatial Moving Average model assumes an additive structure for the area-level risks.

#### **2.4.2.5 Spatio-temporal models**

Spatio-temporal models estimate the relative risk of disease associated with living in a specific area at a certain time. An application of these models results in a series of disease maps over time, although it is possible to produce a single disease map for the study period from the time-specific relative risk estimates. Hierarchical models are typically used for space-time modelling of relative risks. Many build on the BYM model by adding terms for temporal effects. Some model the spatial and temporal effects additively while others incorporate space-time interaction terms. Some have restrictive assumptions of stationarity or linear trend through time while others do little to model temporal change.

Bernardinelli, Clayton et al (1995a) assume the area-specific  $\log(RR)$  displays linear trend through time. The intercept and the coefficient of the linear trend term each include a random effect assigned an Intrinsic CAR prior to capture the spatial structure in the intercept and trend, respectively. The model allows interaction between space and time but is somewhat restrictive in its modeling of the temporal effect as a linear trend. An alternative formulation uses spatially unstructured rather than spatially structured effects. Sun et al (2000) constructed a similar model but estimated the age-specific rate for each area through time. Waller et al (1997) apply the BYM model at each time point separately hence there is virtually no modeling of the temporal effect. A modification of this model by Waller, Carlin and Xia (1997) uses an AR(1) model to describe temporal variation in risk. In their mapping studies Pickle (2000) and McNab and Dean (2002) use splines for age and non-linear temporal effects in their generalized mixed models. Knorr-Held and Besag (1998) propose an extension of the BYM model that includes time-dependent covariates and a temporally structured term that follows a random walk

prior. Unlike other time series models (e.g. ARIMA), the random walk model does not assume stationarity. Knorr-Held (2000) modified the Knorr-Held and Besag formulation to allow for space-time interaction. He considers four interaction types and models interaction behavior by taking the Kronecker product of the precision matrices assigned to the spatial and temporal terms of interest.

#### **2.4.2.6 Studies that evaluate the accuracy of the BYM model in discriminating between high-risk and background-risk areas**

The objectives of disease mapping studies are varied. For example, these may include the accurate estimation of area-level risk, the ranking of areas according to the estimated risk and/or the identification of high-risk areas. Many studies have compared the BYM model to other models with respect to goodness-of-fit (for example, Green and Richardson, 2002; Lawson and Clark, 2002; Hosseini and Larson, 2005; Best et al, 2005; Lawson et al, 2000; Bernardinelli et al, 1995b). Fewer have examined the accuracy of the BYM model in discriminating between high-risk and background risk areas. To our knowledge, five published studies have considered this aspect of the BYM model's accuracy but all five studies have limitations that make it difficult to draw general conclusions regarding the accuracy of the BYM model (Bernardinelli et al, 1995b; Jarup et al, 2002; Richardson et al, 2004; Best et al, 2005; Aamodt et al, 2006). The earliest by Bernardinelli et al (1995b) is a small simulation study with six replications designed to compare the accuracy of relative risk estimation under different choices of hyperprior and to compare the fully Bayesian model to the method of Maximum Likelihood where classification into high and background-risk groups is based on p-values. As the examination of accuracy in detecting high-risk areas is only a small portion of the study, they present ROC curves for the different priors and for the method of

maximum likelihood but there is very little discussion dedicated to this comparison.

Another small simulation study, by Jarup et al (2002), is embedded in a study that considers the geographic distribution of prostate cancer in Great Britain. The Bayesian hierarchical disease mapping model used to smooth the map is not specified. The simulation study considers maps with one hot spot, consisting of several areas. Areas with a PP over 0.8 are classified as high-risk areas. An examination of the sensitivity and specificity for varying true hot spot RR values indicates high overall accuracy (i.e. sensitivity and specificity) when the RR is 3 but low sensitivity combined with high specificity when the relative risk is 1.5. The simulation study by Richardson et al (2004) is based on the same prostate cancer data but accuracy is examined for a variety of clustering patterns. One hundred replications are generated for every pattern. They consider an area with PP over 0.8 as a high-risk area and examine the sensitivity and specificity under varying RR values, and conclude similarly, that the Bayesian disease mapping models tend to have low sensitivity but high specificity when the RR in hot spot areas is below 3. Best et al (2005) compare model fit and accuracy of several Bayesian disease mapping models in a simulation exercise with 5 replications. They fix the cut-off for the PP at 0.8 and consider the resulting sensitivity for various true hot spot RR values. They conclude that the BYM model tends to oversmooth the estimates, shrinking the high-risk areas' RR to 1, but the BYM model is still better than other models at discriminating between high- and background-risk areas. The most recent study, by Aamodt et al (2006), compares the accuracy of the BYM model to two other methods based on sensitivity, specificity and percentage of correctly classified areas. They do not provide the cut-off for the PP that they use for calculating the sensitivities and specificities. Furthermore, their findings are difficult to interpret. For example, one method, they say, has highest sensitivity but lower

specificity than the other methods (at the un-disclosed value of the cut-off for the BYM model). Perhaps, a more appropriate cut-off would result in comparable sensitivity but fair specificity.

Apart from the study by Bernardinelli et al (1995b), the methods used for assessing the accuracy are similar in all studies. All authors find the PP for each area, fix a cut-off for the PP (e.g. 0.8) and consider the sensitivity and sometimes the specificity. In using ROC curves, the Bernardinelli et al (1995b) method considers accuracy for a variety of thresholds, thus making the assessment of accuracy independent of the choice of threshold.

#### **2.4.2.7 The design of the simulation study for assessing accuracy of classification of areas as high-risk or background-risk areas**

The strategy of these simulation studies is to generate maps where the true RR's are known, apply the smoothing models, classify the areas according to the PP's and then compare the set of declared hot spots to the set of true hot spots through sensitivity, specificity and ROC curves. The above five simulation studies are similar in their setup. The steps in the implementation are as follows:

1. Choose the type of map to investigate by considering one or more clustering patterns that will appear in the generated map. Select one or more areas or groups of contiguous areas as hot spots.
2. Choose RR values (or disease rates) for all areas of the map. For example, choose RR=1 for background-risk areas and RR=1.5 for high-risk areas.

3. Generate observed counts from the true relative risk values (or rates). Obtain several replications to guard against the effects of random variability in the data generation.
4. Apply the smoothing model(s).
5. Find the PP for all areas.
6. Either (a) establish a cut-off for the PP that classifies an area as belonging to high- or background-risk groups and find sensitivity and specificity or (b) use ROC curves.

As the general methods for generating data are similar across the above simulation studies, we describe in detail the methods from the Richardson et al (2004) study.

Richardson et al (2004) set out to assess the accuracy of the Bayesian spatial disease mapping models in estimating the relative risk throughout the map, particularly in areas with true elevated risk. In summary, for each area of a map they generate an observed count from known RR values (equal to 1.5, 2 or 3 for high-risk areas and 1 for background-risk areas), then they apply three Bayesian disease mapping models to the generated counts and they evaluate the discriminating power of the models.

Richardson et al (2004) use, as a skeleton, the map of the 532 wards of the county of Yorkshire in England. They use the expected counts of prostate cancer from 1975 to 1991 from the Jarup et al (2002) study. In general, expected counts are found by taking the sum over strata-specific regional incidence rates multiplied by the at-risk person years. Three types of spatial patterns for the true map of the relative risks are produced. The first type of map has five hot spots where each hot spot consists of a single area with elevated risk. The expected counts for these five isolated hot spots range from 0.8 to 7.3. The second map has one hot spot with expected count totalling 1% of the total expected counts and consists of

four adjacent areas. The third type of map has 20 non-overlapping 1% clusters of areas. The observed count for area  $i$  is generated from a Poisson distribution with mean  $RR_i \cdot E_i$ , where  $RR_i$  and  $E_i$  represent the relative risk and expected count, respectively, for area  $i$ . As the amount of smoothing that takes place depends on the size of the expected counts, they generate sets of observed counts from expected counts that are increased by a scale factor (sf) equal to 2, 4, or 10, depending on the map. Thus, they consider three types of spatial patterns, three relative risk values for hot spots and four sets of expected counts corresponding to the four scale factors (1, 2, 4, or 10). This results in 36 types of maps (i.e.  $3 \times 3 \times 4$ ). To control for the effects of random variability in map generation, they produce 100 realizations for each of the 36 types.

Three models are used to smooth the disease map. For each area, the model produces a *posterior sample* of relative risk estimates. To judge whether an area corresponds to a high- or background-risk area, they consider the proportion of relative risks of the *posterior sample* that are greater than 1 (i.e. an estimate of  $P(RR_i > 1)$ ). This means that for each replication, they have a PP for each area on the map. To classify an area as having high-risk or background-risk they need to specify a cut-off for the PP. For this, they consider three loss functions involving the false positive ( $fp$ ) and false negative ( $fn$ ) rates. The goal is to find the cut-off that minimizes the loss. The first loss function is the average of the false positive and false negative rates (i.e.  $loss = \frac{fn + fp}{2}$ ). The second loss function regards a false negative as being twice as costly as a false positive (i.e.  $loss = \frac{2 \cdot fn + fp}{3}$ ), while the third loss function considers that a false positive is twice the cost of a false negative (i.e.  $loss = \frac{fn + 2 \cdot fp}{3}$ ). They report that the cut-off that minimizes the loss is

on average 0.8, though they mention that a cut-off below 0.7 is more appropriate when the scale factor is equal to 1 (i.e. when the expected counts in the true high-risk areas range from 0.8 to 7.3). Thus, a PP over 0.8 classifies the area as a high-risk area. They average the sensitivities and specificities over the 100 replications and conclude that the BYM model has high specificity and low sensitivity in classifying areas as hot spot or background unless the true relative risk of the hot spot areas is large (i.e. over 2) or the expected counts are large (i.e. over 50).

In this chapter, we discussed Bayesian hierarchical models and their importance for disease mapping studies. We also described methods for assessing the accuracy of a classification procedure (i.e. sensitivity and specificity, ROC curves and predictive values). The discussion concluded with a review of studies that evaluate the accuracy of disease mapping models in classifying areas as high-risk or background-risk areas.

# CHAPTER 3 SPATIAL SMOOTHING FOR DISEASE MAPPING AND DETECTION OF AREAS WITH ELEVATED RISK

In this chapter, we assess the accuracy of the Besag, York and Mollie (1991) (BYM) model, in identifying several isolated high-risk areas.

Receiver Operating Characteristic (ROC) curves are used to measure the accuracy of the model's classification, while predictive values are used to assess the utility of the model in detecting high-risk areas, in practice. In addition, we demonstrate that assessing accuracy with sensitivity and specificity measurements for a single cut-off is inadequate for evaluating the performance of the model.

## 3.1 Introduction

The Besag, York and Mollie (1991) (BYM) model has become one of the most widely used Bayesian models for disease mapping, due in part to the ease of implementation via WinBUGS software. In fact, a non-exhaustive review of the fully Bayesian disease mapping literature by Best et al (2005) showed that the BYM model was the only Bayesian spatial disease mapping model used in the non-statistical literature. It appears that little consideration is given to alternative Bayesian disease mapping models that may be no more difficult to implement than the BYM model but may result in a more accurate classification of areas into high-risk and background-risk groups. In addition, despite its popularity, to our knowledge, only five studies (described in chapter 2) have assessed the accuracy of the BYM model, and of these, only two had the assessment of accuracy of the model's classification as a main objective (Richardson et al, 2004; Aamodt et al, 2006). Two of the studies included only a small

number of replications (5 or 6 replications) (Bernardinelli et al, 1995; Best et al, 2005) and a third study was described as small, by the authors, but the number of replications was not specified (Jarup et al, 2002).

An additional limitation of some of these studies was the means by which they assessed accuracy. Four studies selected a single cut-off for the posterior probability (PP) that would separate high-risk from background-risk areas (Jarup et al, 2002; Richardson et al, 2004; Best et al, 2005; Aamodt et al, 2006). Using this classification they estimated the sensitivity and, in some studies, the specificity of the models. This type of assessment can be misleading since sensitivity and specificity depend on the choice of cut-off. A cut-off for the PP that is too high will yield low sensitivity but high specificity and a cut-off that is too low will yield high sensitivity and low specificity. Two of the simulation studies used the same cut-off of 0.8 for all true hot spot relative risk values, another used a cut-off of 0.75 and the fourth did not report their choice of cut-off. Thus, the resulting sensitivities and specificities measured the accuracy of the smoothing model *with a threshold of 0.8 (or 0.75)*, rather than the accuracy of the smoothing model alone. Methods that assess accuracy independently of the choice of threshold, such as ROC curve methodology, are more appropriate as there is no universally accepted threshold or cut-off separating a high-risk from a background risk area. In practice, many researchers will not feel comfortable interpreting the PP as the ‘true’ probability, so they will not be able to evaluate whether the 0.75 or 0.8 cut-off is appropriate for their problem.

Finally, none of the studies included an assessment of the predictive value of the smoothing models. Even when sensitivity and specificity are high, there may be low probability that a ‘detected hot spot’ corresponds to a true hot spot. The probability that an area classified as high-risk corresponds to a true high-risk area is given by the positive predictive

value. The positive predictive value of a procedure depends on the sensitivity, the specificity and the prevalence, where the prevalence, in this context, refers to the proportion of the areas that are true high-risk areas.

In this chapter we evaluate the accuracy of the BYM model through a simulation study. We focus on the spatial clustering pattern in which there are several small isolated hot spots, where a 'small, isolated hot spot' is defined as a single high-risk area surrounded by background-risk areas. This research uses, as a basis, the design of the other simulation studies that have addressed this question (see section 2.4.2.7) but avoids some of their limitations by performing the analysis on a large number of replications and assessing the accuracy more completely with ROC curves and predictive values. Finally, we find the 'optimal' cut-off that jointly maximizes the sensitivity and specificity (discussed in section 2.3). We judge the impact of using a universal threshold of 0.8 as in Richardson et al (2004) by applying their threshold to our PP estimates and comparing the resulting sensitivity and specificity to those obtained by applying the 'optimal' cut-off.

The simulated data are based on the county structure for the state of Kentucky, the at-risk person-years from 1995 to 2002 and the Kentucky cervical cancer rate from 1995 to 2002. These data are publicly available from the Kentucky Cancer Registry website ([www.kcr.uky.edu](http://www.kcr.uky.edu)). For each simulated disease map, we create several isolated hot spots. We smooth the disease maps using the BYM model and then evaluate the accuracy of the BYM model through ROC curves and predictive values.

### 3.2 Data and Methods

In order to assess the accuracy of spatial smoothing models we generate sets of observed counts (or maps of observed counts) based on our choices of RR for each area of the map. The resulting maps are smoothed using the BYM model giving a *posterior sample* of relative risks for each area. The proportion of the *posterior sample* that exceeds the null value of 1 (i.e. the PP) is used as a measure of the likelihood that the area is a hot spot. To form a ROC curve, the sensitivity is plotted against one minus the specificity for several cut-off values for the PP. The area under the curve (AUC) is one measure of the accuracy of the classification procedure (i.e. the model) (Hanley and McNeil, 1982).

We proceed by describing the data from which we generate maps, the selection of areas that are simulated as high-risk areas, the specification of the model, the calculation of positive and negative predictive values and the construction of the ROC curve.

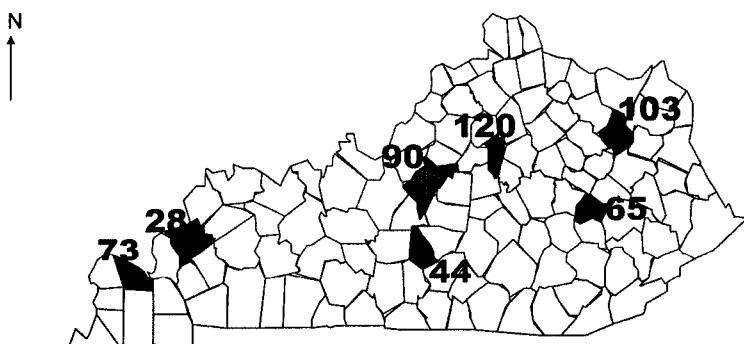
The state of Kentucky is divided into 120 counties and for the study period 1995 to 2002, the at-risk person-years for cervical cancer and the cervical cancer rate for Kentucky are used to calculate expected counts for each county. Let  $E_i$  represent the expected count for county  $i$ ,  $ref.rate$  the rate of cervical cancer in Kentucky and  $at.risk_i$  the at-risk person-years in area  $i$ , then the expected counts are calculated as follows:

$$E_i = ref.rate \cdot at.risk_i$$

In choosing the number of hot spots we want to strike a balance between the number of opportunities the model will have to detect a hot spot with a number of hot spots that is realistic for a disease map. Thus, we settle on seven hot spots, recognizing that our results could certainly be sensitive to

this choice. We want to choose seven hot spots that vary with respect to the size of their expected counts because the amount of shrinkage that takes place depends in part on the expected count. Areas with extreme RR values that have larger expected counts are not as influenced by the RR's of the neighboring areas, particularly when the neighboring expected counts are smaller. Three counties are chosen from the bottom third of the distribution of expected counts, two from the middle third and two from the top third. After selecting each county, the neighboring counties are removed from selection to ensure that each hot spot is isolated (i.e. does not have a neighbor with high risk). Figure 3.1 shows the map of Kentucky with high-risk counties in blue and background-risk counties in white. The counties are coded from 1 to 120 in the dataset so we refer to the counties by these codes. Counties 65, 28 and 44 are in the bottom third of the distribution of expected counts with expected counts of 3.9, 4.9, and 5.9, respectively. Counties 103 and 120 are in the middle third with expected counts of 11.2 and 11.7 and counties 90 and 73 are in the upper third of the distribution with expected counts of 18.8 and 36.3, respectively.

Figure 3.1 Kentucky map with hot spots in blue. Hot spots are labeled by area numbers.



At this point, we have calculated our expected counts for each county and we have selected the counties we want as hot spots. As described in chapter 2, we assume the observed count for county  $i$ ,  $Y_i$ , is Poisson distributed with mean  $RR_i \cdot E_i$ , where  $RR_i$  is the relative risk of county  $i$ . Thus, for each county, we generate the observed count by drawing from a Poisson distribution with mean  $RR_i \cdot E_i$ . We have calculated the expected counts ( $E_i$ ), but we need to choose relative risk values ( $RR_i$ ). All background-risk counties are assigned a relative risk equal to 1, meaning that the risk associated with living in a background county is no different from the risk associated with living in the study region as a whole, i.e. Kentucky. For the high-risk counties we want to examine the accuracy under different, realistic hot spot RR values and we want to compare our assessment of accuracy to those of Richardson et al (2004) so we assign relative risks of 1.5, 2 or 3 to the high-risk counties, as is done in their simulation studies. Thus, we generate observed counts from independent Poisson distributions with mean  $RR_i \cdot E_i$  using the assigned relative risks and the expected counts that we compute from the cervical cancer rate of Kentucky and the at-risk person-time of the 120 counties. Three separate sets of simulations will utilize different true values for the hot spot relative risk. One set has hot spot RR equal to 1.5 for all seven planted hot spots, another has hot spot RR equal to 2 and the third set has hot spot RR equal to 3.

Recall that the amount of smoothing depends on the degree of instability (i.e. lack of precision) of the crude SMR estimates or equivalently the size of the expected counts. To study the accuracy of the spatial smoothing model under differing degrees of stability, we generate the observed counts from expected counts ( $E_i$ ) that are reduced by a factor of 0.5, 0.3, and 0.1. These scale factors produce a realistic range of expected counts that result in sufficient instability of the RR estimates to warrant the use of

a smoothing model. For each scale factor ( $sf$ ) and hot spot relative risk ( $RR$ ) combination we generate 50 realizations of the 120 observed counts. This means that 50 sets of observed counts are generated for each of the 12  $RR - sf$  combinations ( $RR = 1.5, 2, \text{ or } 3$ , and  $sf = 1, 0.5, 0.3, \text{ or } 0.1$ ).

$$Y_i \sim Poisson(E_i^* \cdot RR_i) \quad \text{where } E_i^* = E_i \cdot sf, \text{ for } i = 1, \dots, 120.$$

After we generate observed counts, we apply the BYM model to produce smoothed disease maps. The specification of the model is as follows:

$$\begin{aligned} Y_i &\sim Poisson(E_i^* \cdot RR_i) \\ \log(RR_i) &= \alpha + U_i + V_i \\ U_i \mid U_{j \neq i} &\sim CAR(\tau_U) \\ V_i &\sim N(0, \tau_V) \\ \tau_U &\sim Gamma(0.5, 0.0005) \\ \tau_V &\sim Gamma(0.5, 0.0005) \end{aligned} \tag{3.1}$$

where  $\alpha$  is the intercept that is required to achieve identifiability when employing the Intrinsic CAR prior (Thomas et al, 2004), the  $U_i$  represent the effect of unmeasured covariates that display spatial structure and the  $V_i$  represent the effect of unmeasured covariates that do not display spatial structure.  $CAR$  refers to the Intrinsic CAR prior based on the normal distribution. The precisions of  $U_i$  and  $V_i$  are  $\tau_U$  and  $\tau_V$ . We specify distributions (hyperpriors) for the precisions. Each precision is assigned a Gamma distribution with shape parameter equal to 0.5 and

inverse scale equal to 0.0005, following the recommendation of Kelsall and Wakefield (1999) (Wakefield et al, 2000a). The intercept  $\alpha$  is assigned a flat, non-informative prior.

For a single replication, we use WinBUGS1.4 (Spiegelhalter et al, 2002) and R software (R Development Core Team, 2005; BRugs package: Ligges and Sturtz, 2005) to obtain 3 chains of length 10 000.

Convergence is assessed using the  $\hat{R}$  values from the WinBUGS summary output in R. After discarding the first 5000 iterations and retaining every fifth iteration thereafter, for each replication, we are left with a posterior sample of 3000 relative risk estimates per county.

For each replication, once the disease map is smoothed, we find the proportion of the *posterior sample* of the relative risk that exceeds the reference value of one (i.e. the *posterior proportion*, PP) for each county. We illustrate how the sensitivities and specificities are found from the PP values for one of the 50 replications. Using MCMC techniques, for each of the 120 counties, we obtain 30 000 relative risk estimates and we keep 3000 of these estimates for our *posterior sample*. For each county we find the proportion of the 3000 estimates that exceed 1. As an illustration, we consider a single threshold (cut-off) of 0.65. We record the number of hot spots that have *posterior proportions* over 0.65, the number of background-risk areas with *posterior proportions* over 0.65 and the number of hot spots and background-risk areas with *posterior proportions* under 0.65. The 2x2 table below presents the results.

Table 3.1: Illustration of cross-classification of results from the model's classification of areas using a threshold of 0.65 by true status

	True Hot Spot	True Background
Proportion > 0.65 “hot spot detected”	5	25
Proportion < 0.65 “hot spot not detected”	2	88

The estimate of the sensitivity of the model for this threshold is the proportion of hot spots that are correctly detected as a hot spot (i.e.  $5/7 = 71\%$ ). The estimate of the specificity for this threshold is the proportion of background counties that are correctly classified as such (i.e.  $88/113 = 78\%$ ).

For each  $RR - sf$  combination, we have a vector of 6000 PP's formed from pooling the 120 PP's from all 50 replications. Given a cut-off (i.e. threshold), the proportion of the  $50 \times 7$  true hot spots that are detected as such by the model gives an estimate of the sensitivity. Similarly, given a cut-off the proportion of the  $50 \times 113$  true background-risk areas that are correctly identified as background is an estimate of the specificity. By estimating the sensitivity and specificity for various cut-off values we can form the ROC curve (i.e. a plot of the sensitivity vs. 1-specificity at each cut-off).

We also find estimates of the positive and negative predictive values of the model. The positive predictive value ( $PPV$ ) is estimated by the proportion

of the declared hot spots that are true hot spots. The negative predictive value (*NPV*) is estimated by the proportion of declared background-risk counties that are true background-risk counties. In the illustrative example presented in Table 3.1, the estimate of the positive predictive value is 5/30 = 17%, while the negative predictive value is estimated to be 88/90=98%. Hence, with a cut-off of 0.65 for the PP, the positive predictive value is very low. On average, only 17% of those counties declared as hot spots will be true hot spots, whereas 98% of counties that are declared as background will be true background areas.

As described in chapter 2, there is a mathematical relationship (see 2.1 in section 2.3) between the positive predictive value, the sensitivity and specificity of the test and the prevalence of the ‘positives’ in the population. Using this relationship, we examine the prevalence that is needed to have at least a 50% chance of a declared hot spot being a true hot spot.

As mentioned in section 3.1, two of the four simulation studies that evaluated the discriminatory power of the BYM model used a cut-off of 0.8 and another study used a cut-off of 0.75. To determine the adequacy of a cut-off of 0.8, for each *RR – sf* combination, we find the threshold or cut-off that jointly maximizes sensitivity and specificity. The method we will use requires us to find the cut-off that minimizes

$$d^2 = (1 - \text{sensitivity})^2 + (1 - \text{specificity})^2. \quad (3.2)$$

This corresponds to finding the cut-off and resulting sensitivity/specificity pair that minimizes the Euclidean distance from the point (1,1) (i.e. the ideal situation where sensitivity and specificity are both 1). This cut-off

minimizes the total number of errors (i.e. false positives and false negatives are weighted equally).

In addition, we further explore the choice of cut-off by following the methods of Richardson et al (2004), that is, we find, for each  $RR - sf$  combination, the threshold that minimizes the three loss functions. Each loss function is a linear combination of the false-negative rate ( $fn$ ) and the false-positive rate ( $fp$ ):

$$\begin{aligned} loss1 &= \frac{fn + fp}{2} \\ loss2 &= \frac{2 \cdot fn + fp}{3} \\ loss3 &= \frac{fn + 2 \cdot fp}{3} \end{aligned} \quad . \quad (3.3)$$

As Richardson et al (2004) limit their assessment of accuracy to sensitivity and specificity estimates for a threshold of 0.8 for all  $RR - sf$  combinations, we compare the sensitivities that result from this choice to those obtained in their study.

To summarize the methods used for assessing accuracy, we begin by producing a ROC curve for every  $RR - sf$  combination along with the corresponding AUCs. From there, using (3.2), we find the ‘optimal’ cut-off that jointly maximizes the sensitivity and specificity. Using this cut-off we find the positive and negative predictive values,  $PPV$  and  $NPV$  respectively. We explore the (hot spot) prevalence requirements that are needed to have a  $PPV$  of at least 50% by producing graphs of  $PPV$  versus prevalence for each  $RR - sf$  combination.

We also wish to compare our methods and results to those of Richardson et al (2004) so we start by finding, for each  $RR - sf$  combination, the cut-off that minimizes the loss functions (3.3) and compare these cut-offs to those found by minimizing the Euclidean distance to (1,1). Richardson et al (2004) chose a single cut-off of 0.8 for all  $RR - sf$  combinations, so we produce sensitivities for each of the seven hot spots using their cut-off of 0.8 and again using our  $RR - sf$  specific ‘optimal’ cut-off. This will demonstrate that the sensitivity depends highly on the choice of cut-off used to separate a hot spot from a background-risk area.

### 3.3 Results

To summarize the results for all 120 counties we present, in Figures 3.2 and 3.3, disease maps of the posterior proportion exceeding 1 and the posterior mean relative risk for the simulations with hot spot RR equal to 1.5 and scale factor equal to 1. These are averaged over the 50 replications. The disease maps for all  $RR - sf$  combinations are presented in Appendix A. We see that the BYM model, through the spatially structured term, elevates some background area relative risks due to the presence of neighbouring hot spots. In other words, the spatially structured term of the BYM model spreads the hot spot high risk to some of the neighbouring background-risk areas. This is particularly obvious in the western tip of the state. For example, hot spot county #73 has the largest expected count of all hot spots and strongly influences the risks in the surrounding areas, making those background-risk areas appear to have elevated risk as well.

Figure 3.2: *Posterior proportion* (exceeding 1), averaged over the 50 replications for the case where hot spot RR=1.5 and sf=1. True hot spots are indicated by a black dot. Area 73 has the largest expected count of all hot spots.

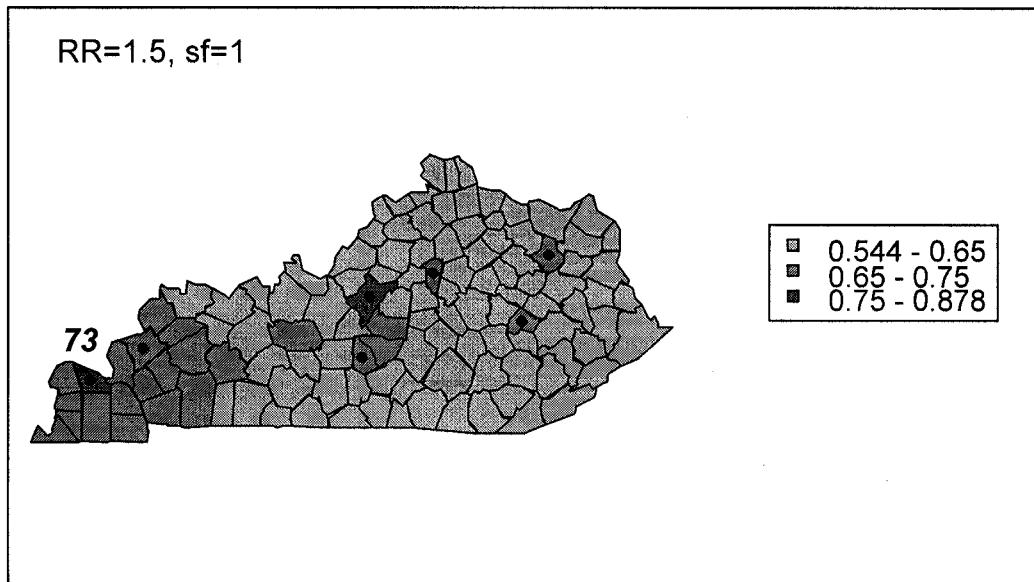


Figure 3.3: *Posterior mean relative risk* averaged over 50 replications for the case where the true hot spot RR=1.5, sf=1. True hot spots are indicated by a black dot. Area 73 is the hot spot with largest expected count.

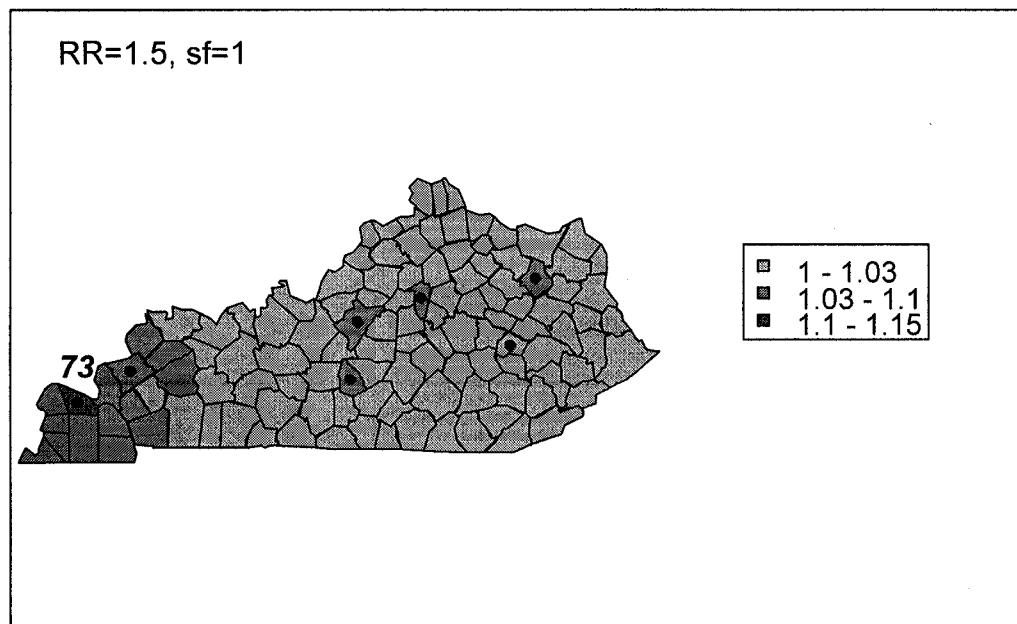


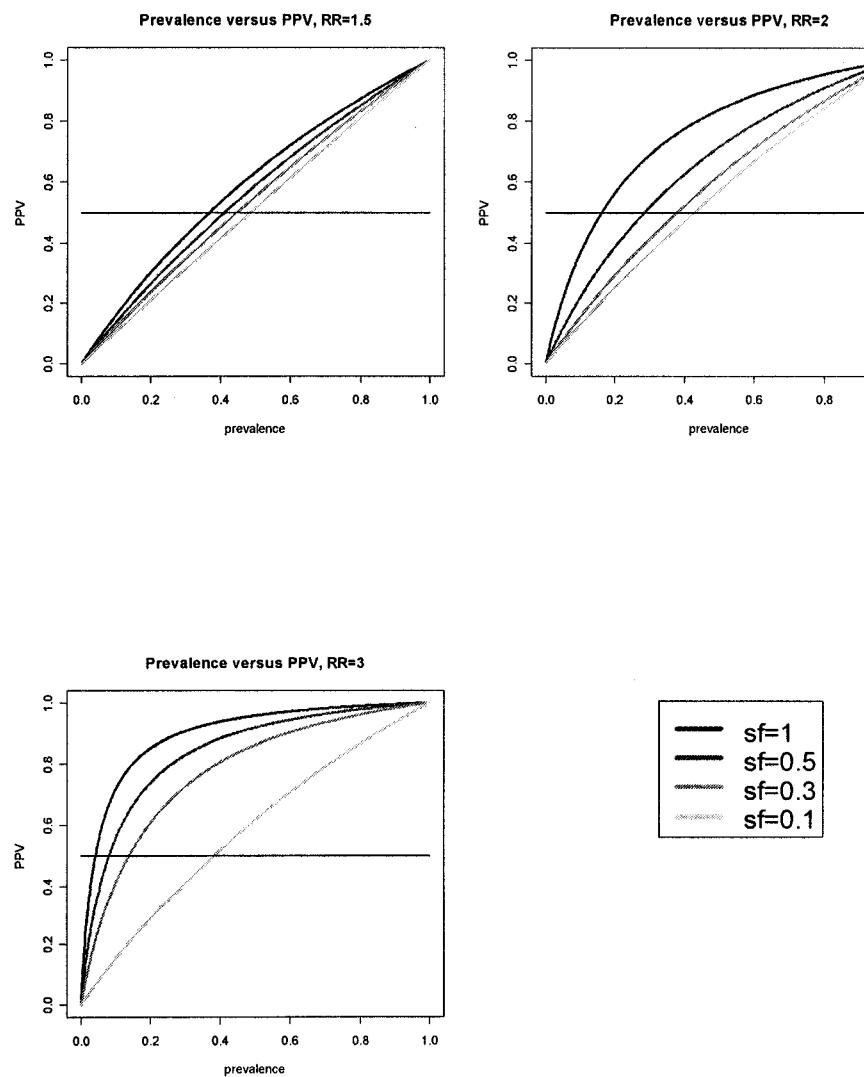
Table 3.2 shows, for each  $RR - sf$  combination, the cut-off that minimizes  $d^2$  in (3.2), along with the jointly maximized sensitivity and specificity, predictive values and AUC (see Appendix A for ROC curves). We see that the sensitivity, specificity and AUCs indicate that the BYM model quite accurately classifies areas into high- and background-risk groups when the hot spot relative risk is 3 and the scale factor is 0.3 or more or when the relative risk is 2 and the scale factor is 1. All other RR-sf combinations lead to poor accuracy (i.e. AUCs below 0.8) and when the scale factor is 0.1 the model does hardly better than a random classification, particularly when the true hot spot relative risk is 1.5 or 2.

Table 3.2: Area under the ROC curve (AUC), predictive values ( $PPV$  and  $NPV$ ), cut-off that minimizes Euclidean distance of (sensitivity, specificity) to the point (1,1) and the corresponding sensitivity and specificity for that cut-off.

Relative Risk	Scale Factor	Range of $E_i$ in Hot Spots	Cut-off	Max Sens	Max Spec	$PPV$	$NPV$	AUC
1.5	1	3.90 – 36.30	0.68	0.69	0.60	0.10	0.97	0.70
	0.5	1.95 -18.15	0.62	0.56	0.61	0.08	0.96	0.62
	0.3	1.17-10.89	0.60	0.57	0.54	0.07	0.95	0.57
	0.1	0.39-3.63	0.64	0.51	0.52	0.06	0.94	0.52
2	1	3.90 – 36.30	0.78	0.83	0.84	0.25	0.99	0.92
	0.5	1.95 -18.15	0.75	0.73	0.71	0.13	0.98	0.79
	0.3	1.17-10.89	0.73	0.56	0.66	0.09	0.96	0.65
	0.1	0.39-3.63	0.67	0.50	0.59	0.07	0.95	0.56
3	1	3.90 – 36.30	0.92	0.92	0.96	0.62	0.99	0.98
	0.5	1.95 -18.15	0.84	0.91	0.92	0.42	0.99	0.97
	0.3	1.17-10.89	0.78	0.81	0.87	0.28	0.99	0.93
	0.1	0.39-3.63	0.66	0.63	0.61	0.09	0.96	0.68

The  $NPV$  is high for every  $RR - sf$  combination. As explained in section 2.3, this is due to the fact that, given sensitivity and specificity values, a higher prevalence of positives increases  $PPV$  while a higher prevalence of negatives increases  $NPV$ . The graphs of *positive predictive value* versus prevalence, produced from (2.1), are given in Figure 3.4.

Figure 3.4: Plot of prevalence versus *positive predictive value* ( $PPV$ ). These graphs were formed using the sensitivity and specificity from the ‘optimal’ cut-off in equation (2.1). Each curve corresponds to a scale factor (i.e.  $sf=1, 0.5, 0.3, 0.1$ ).



Where the graph cuts the black horizontal line (i.e. 0.5) we have the prevalence of hot spots required for a greater than 50% chance of a detected hot spot corresponding to a true hot spot. We see that the prevalence needed to have a  $PPV$  over 50% varies according to the hot spot relative risk values and the scale factors. When the true hot spot relative risk is 1.5 and the scale factor is 1 (i.e. hot spot expected count between 3.9 and 36.3), we would need at least 34% of the areas of the map to be hot spots. This drops to 4% when the elevated relative risk is 3. As the scale factor decreases we see that the prevalence requirements increase, but it is common to use the map with finest partition; hence, we generally do not control the size of the expected counts. The only other way to increase the  $PPV$  is to raise the cut-off. For example, when true hot spot relative risk is 1.5 and the scale factor is 1, if we raise the cut-off to 0.8, the  $PPV$  increases to 15% (from 10%) and the  $NPV$  drops slightly to 96% (from 97%). Raising the cut-off to 0.95 raises the  $PPV$  to 52% and lowers the  $NPV$  to 95%. Although the cut-off is extremely high so that we may classify few, if any areas as high-risk, any detected hot spot has little more than a 50% chance of being a true hot spot. Thus, even this alternative is inadequate.

We compare our optimal cut-offs from minimizing the Euclidean distance (Table 3.2) to the cut-offs from minimizing the loss functions. The graphs of the loss functions versus the cut-offs are presented in Figures 3.5 to 3.7 and they show that the cut-off depends on the hot spot relative risk, the scale factor and the loss function we wish to minimize. If there is a greater loss associated with false-positives we should choose a large cut-off (e.g. over 0.8). If we place the same importance on true positives as true negatives we should choose a cut-off between 0.65 and 0.85 and if false-negatives result in a greater loss, a cut-off below 0.8 is appropriate. As the hot spot relative risk increases, the cut-off that minimizes the loss increases and when the scale factor increases, the required cut-off also

increases. In comparison with our ‘optimal’ cut-offs, we see that the cut-off at which the three loss functions intersect is very close to the ‘optimal’ cut-offs found by minimizing the Euclidean distance. For example the intersection points for simulations with true hot spot relative risks of 1.5 and scales factors equal to 1, 0.5, 0.3, and 0.1 are 0.68, 0.62, 0.6 and 0.64 with our method, compared to 0.7, 0.61, 0.6 and 0.64 with the intersection of the loss functions, respectively.

In Tables 3.3 and 3.4 we present the sensitivities using the Richardson et al (2004) cut-off of 0.8 and the ‘optimal’ cut-offs from Table 3.2. We see that the sensitivities are always higher using the ‘optimal’ cut-offs except for a few cases when the true hot spot relative risk is 3. This makes sense since the cut-off of 0.8 is too low for some simulations with RR=3, resulting in higher sensitivities at the cost of a lower specificity.

We present in Table 3.5, the maximum  $\hat{R}$  values, for the relative risks, from three trial runs for each  $RR - sf$  combination. The maximum  $\hat{R}$  values are well under 1.2, thus convergence is achieved at 10 000 iterations.

Figure 3.5: Graphs of loss functions versus cut-off for the case where the hot spot RR is 1.5. Loss 1 is the average of the false positive and false negative rates given a cut-off value. Loss 2 assumes that a false negative is twice as costly as a false positive and Loss 3 assumes that a false positive is twice as costly as a false negative.

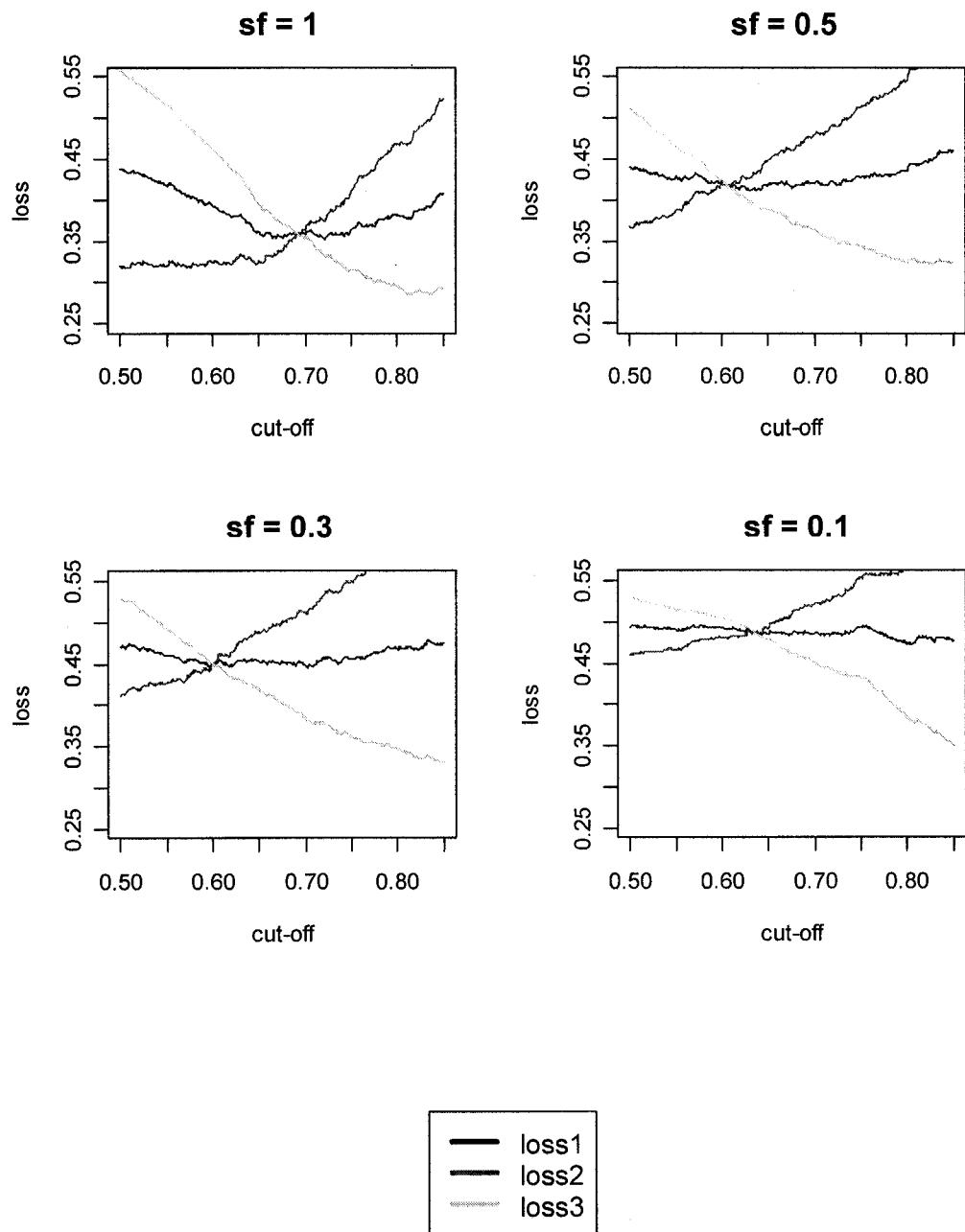


Figure 3.6: Graphs of loss functions versus cut-off for RR=2. Loss 1 is the average of the false positive and false negative rates given a cut-off value. Loss 2 assumes that a false negative is twice as costly as a false positive and Loss 3 assumes that a false positive is twice as costly as a false negative.

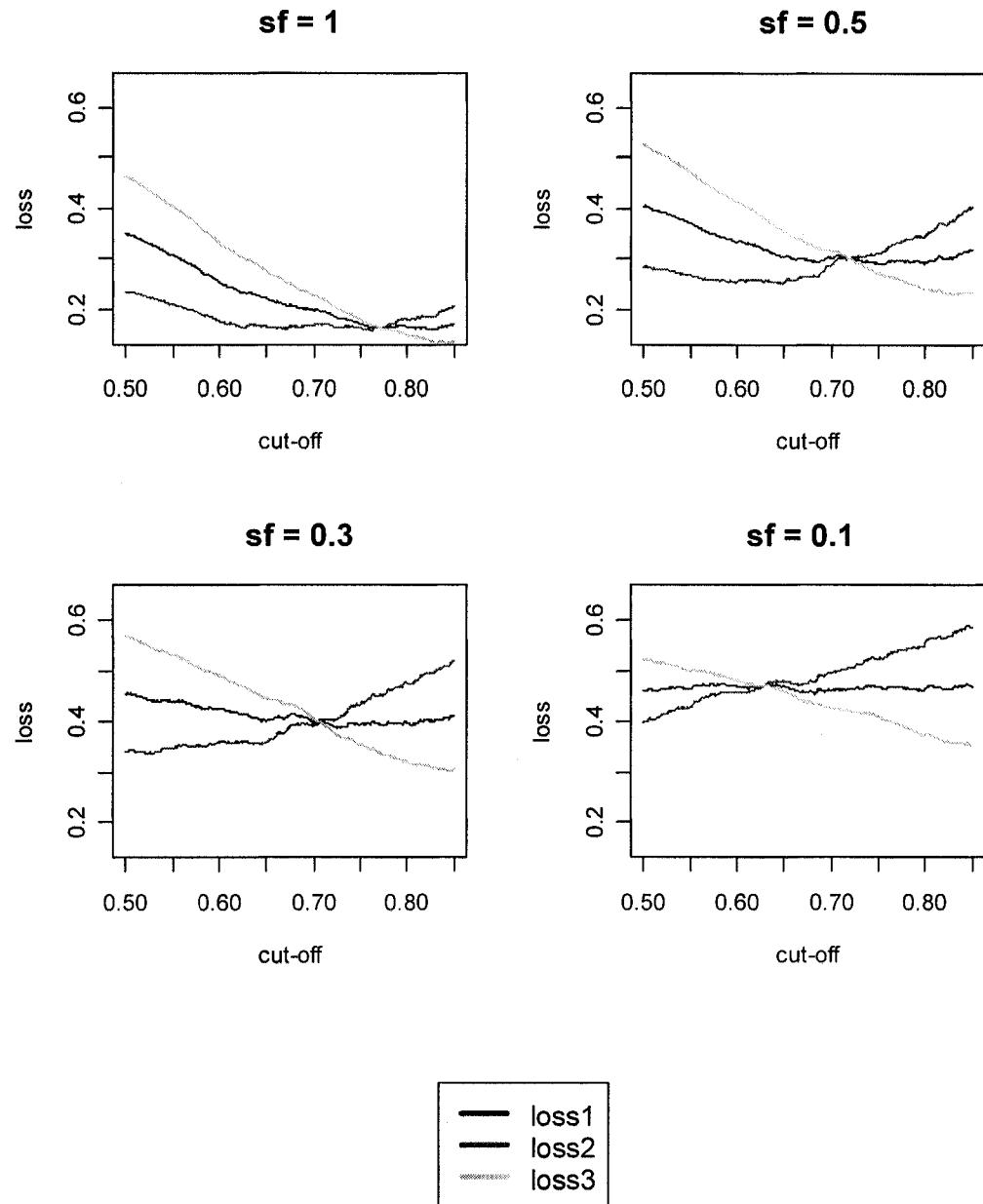


Figure 3.7: Graphs of loss functions versus cut-off for RR=3. Loss 1 is the average of the false positive and false negative rates given a cut-off value. Loss 2 assumes that a false negative is twice as costly as a false positive and Loss 3 assumes that a false positive is twice as costly as a false negative.

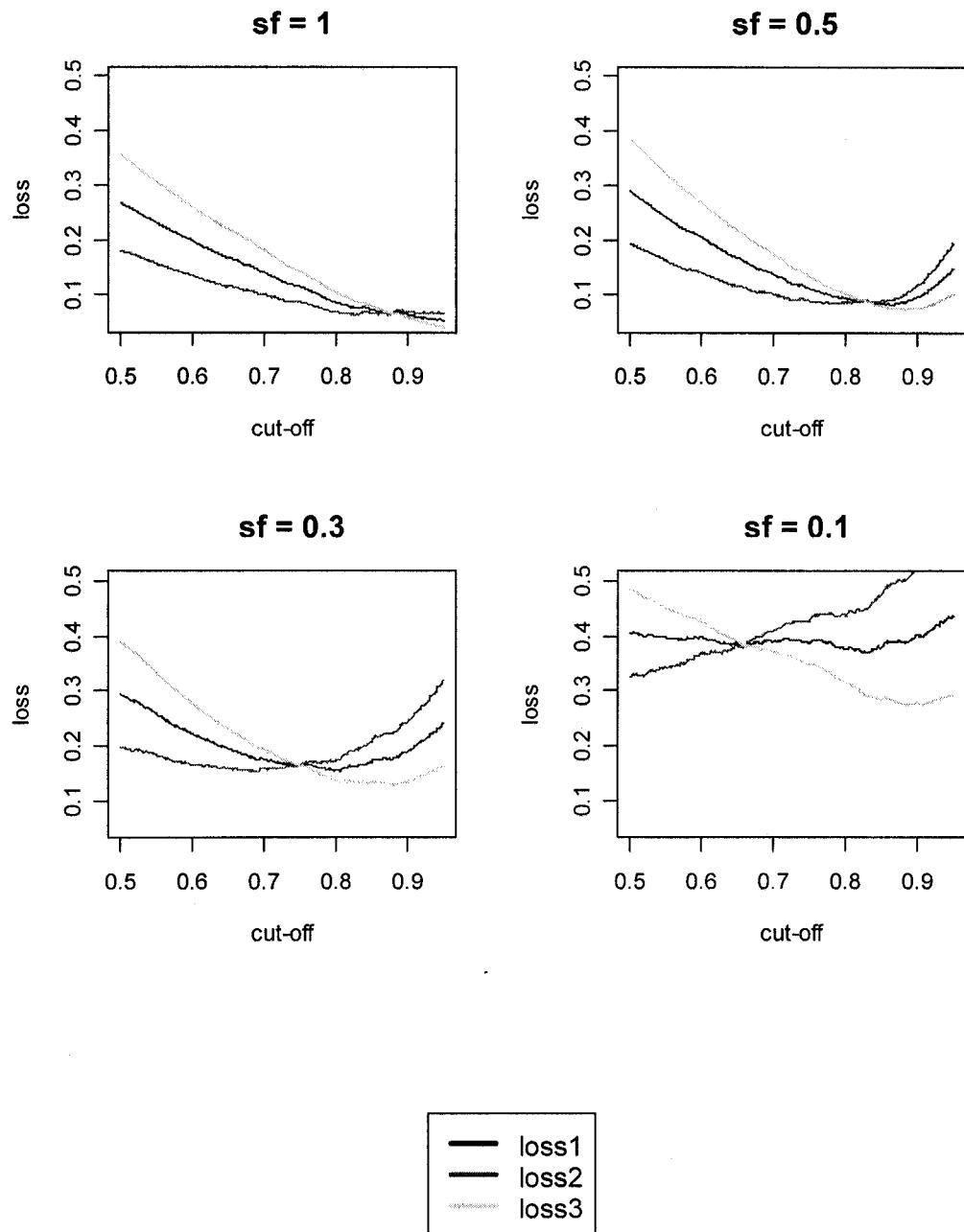


Table 3.3: Sensitivity for individual hot spots using a threshold of 0.8.  $E_i$  is the expected count for the area multiplied by the scale factor (sf). RR is the true hot spot relative risk.

AREA #		65	28	44	103	120	90	73
$E_i$ (sf=1)		3.9	4.9	5.9	11.5	12.1	18.8	36.3
sf	RR							
1	1.5	0.14	0.24	0.22	0.28	0.34	0.54	0.74
	2	0.54	0.62	0.64	0.92	0.88	0.94	1.00
	3	0.92	0.94	0.9	1.00	1.00	1.00	1.00
$E_i$ (sf=0.5)		1.95	2.45	2.95	5.75	6.05	9.40	18.15
0.5	1.5	0.18	0.12	0.12	0.22	0.24	0.28	0.48
	2	0.18	0.46	0.38	0.54	0.6	0.74	0.94
	3	0.78	0.84	0.92	0.98	0.98	1.00	1.00
$E_i$ (sf=0.3)		1.17	1.47	1.77	3.45	3.63	5.64	10.89
0.3	1.5	0.1	0.16	0.12	0.1	0.16	0.24	0.3
	2	0.22	0.3	0.24	0.32	0.42	0.44	0.64
	3	0.58	0.62	0.58	0.82	0.9	1.00	1.00
$E_i$ (sf=0.1)		0.39	0.49	0.59	1.15	1.21	1.88	3.63
0.1	1.5	0.22	0.26	0.22	0.26	0.28	0.32	0.26
	2	0.22	0.24	0.22	0.3	0.28	0.26	0.34
	3	0.30	0.32	0.32	0.42	0.42	0.58	0.70

Table 3.4: Sensitivity for individual hot spots using the RR-sf specific 'optimal' cut-off.  $E_i$  is the expected count for the area multiplied by the scale factor (sf). RR is the true hot spot relative risk.

AREA #		65	28	44	103	120	90	73
$E_i$ (sf=1)		3.9	4.9	5.9	11.5	12.1	18.8	36.3
sf	RR							
1	<b>1.5</b>	0.14	0.64	0.56	0.60	0.68	0.76	0.88
	<b>2</b>	0.62	0.64	0.70	0.92	0.9	0.96	1.00
	<b>3</b>	0.68	0.86	0.88	1.00	1.00	1.00	1.00
$E_i$ (sf=0.5)		<b>1.95</b>	<b>2.45</b>	<b>2.95</b>	<b>5.75</b>	<b>6.05</b>	<b>9.40</b>	<b>18.15</b>
0.5	<b>1.5</b>	0.48	0.52	0.48	0.5	0.56	0.64	0.76
	<b>2</b>	0.32	0.60	0.54	0.62	0.66	0.84	0.96
	<b>3</b>	0.68	0.80	0.92	0.98	0.98	1.00	1.00
$E_i$ (sf=0.3)		<b>1.17</b>	<b>1.47</b>	<b>1.77</b>	<b>3.45</b>	<b>3.63</b>	<b>5.64</b>	<b>10.89</b>
0.3	<b>1.5</b>	0.48	0.5	0.46	0.52	0.56	0.6	0.72
	<b>2</b>	0.48	0.46	0.44	0.5	0.58	0.6	0.76
	<b>3</b>	0.64	0.64	0.58	0.86	0.94	1.00	1.00
$E_i$ (sf=0.1)		<b>0.39</b>	<b>0.49</b>	<b>0.59</b>	<b>1.15</b>	<b>1.21</b>	<b>1.88</b>	<b>3.63</b>
0.1	<b>1.5</b>	0.42	0.58	0.52	0.46	0.48	0.52	0.58
	<b>2</b>	0.42	0.42	0.46	0.54	0.5	0.52	0.58
	<b>3</b>	0.5	0.5	0.5	0.6	0.62	0.74	0.88

Table 3.5: Maximum  $\hat{R}$  values from three trial runs for each relative risk-scale factor combination.

<b>Relative Risk</b>	<b>Scale Factor</b>	$\hat{R}$		
1.5	<b>1</b>	1.03	1.04	1.08
	<b>0.5</b>	1.01	1.03	1.04
	<b>0.3</b>	1.02	1.02	1.01
	<b>0.1</b>	1.02	1.02	1.01
2	<b>1</b>	1.01	1.01	1.00
	<b>0.5</b>	1.02	1.02	1.00
	<b>0.3</b>	1.01	1.02	1.02
	<b>0.1</b>	1.04	1.02	1.04
3	<b>1</b>	1.01	1.00	1.01
	<b>0.5</b>	1.00	1.00	1.00
	<b>0.3</b>	1.00	1.01	1.01
	<b>0.1</b>	1.06	1.04	1.07

### 3.4 Discussion

Fully Bayesian disease mapping models have been used extensively to describe the spatial variation of disease risk and to locate areas with elevated risk as compared to the risk of a reference region. When using these models it is important to know the extent to which we can have

confidence in the results we obtain. The purpose of our simulation study was to assess the accuracy of the BYM model in classifying areas into high-risk and background-risk groups, where accuracy is measured by sensitivity, specificity, predictive values and ROC curves. We consider the case where hot spots consist of a single area, surrounded by background-risk areas.

To our knowledge, there have only been five published simulation studies that have assessed the accuracy of this model in discriminating between high-risk and background-risk areas and these studies had limitations that make it difficult to draw conclusions regarding the accuracy of the model. For example, a previous simulation study that assessed accuracy through ROC curves used only 6 replications, a relatively small number (as noted even by the study's authors) (Bernardinelli et al, 1995b). Other simulation studies involved limited assessments of accuracy consisting of sensitivity and specificity calculations and proportions of correctly classified areas (Jarup et al, 2002; Richardson et al, 2004; Best et al, 2005; Aamodt et al, 2006). Richardson et al (2004), for example, used a single threshold of 0.8 for the *posterior proportions* to find the sensitivities and specificities. They concluded that the BYM model has high specificity with low sensitivity except when the expected counts are over 50 per area or the hot spot relative risks are over 2. However, this was actually an assessment of the model with a cut-off of 0.8 rather than an assessment of the model itself. They say further, that for some simulations, particularly those with smaller expected counts, a cut-off below 0.7 resulted in higher accuracy (i.e. a cut-off below 0.7 minimized the loss functions) yet they applied a universal cut-off of 0.8 to all simulations. Thus, it is not clear whether the model or the cut-off of 0.8 caused the low sensitivity and high specificity. Another inappropriate criterion of diagnostic accuracy is the proportion of correctly classified areas. This proportion is used in the simulation study by Aamodt et al (2006). The proportion of correctly

classified areas is inadequate in assessing overall accuracy because it is dominated by the specificity when the prevalence of hot spots is small. For example, assume that there are 100 areas, 5 of which are hot spots. If the model fails to detect any of the hot spots but correctly classifies all 95 background-risk areas, the sensitivity is estimated at 0, the estimate of specificity is 1.00 and the proportion of correctly classified areas will be impressive at 0.95 even though the model failed to detect any of the 5 true hot spots.

We find that the BYM model has low accuracy (i.e. low sensitivity, low specificity and small AUC), except when the hot spot relative risk is 2 and the scale factor is 1 (corresponding to hot spot expected counts of approximately 4, 5, 6, 11.5, 12, 19 and 36), or the RR is 3 and the scale factor is at least 0.3. What is more disconcerting, however, is the fact that the positive predictive values are extremely low due to the small percentage of the counties that are hot spots (i.e.  $7/120=6\%$ ). Given that the prevalence of hot spots in a disease mapping study is usually small, in practice, this would imply that detected hot spots rarely correspond to true high-risk areas.

The plots of the positive predictive value versus prevalence of hot spots (i.e. proportion of areas that are true high-risk areas) determined the required prevalence to achieve a positive predictive value exceeding 0.5, assuming that the sensitivity and specificity would remain constant as the prevalence increased. However, larger prevalences of hot spots would no longer correspond to a situation of isolated high-risk areas, but instead, would correspond to a spatial pattern of clustered high-risk areas. The simulation study by Richardson et al (2004) examined this spatial pattern of clustering and found that, for some simulated maps, the BYM model was better at detecting clusters of high-risk, leading to slightly higher sensitivity and specificity. However, the improvement in the clustered hot

spots scenario is far from obvious, thus the assumption of constant sensitivity and specificity should hold approximately. We found that to have a positive predictive value over 0.5, the required prevalence for maps with true hot spot relative risk equal to 1.5 ranged from 0.34 to 0.47. These high proportions are not typical, implying that for most studies the probability that a detected hot spot corresponds to a true high-risk area is under 0.5.

In comparing ‘optimal’ cut-offs we found that minimizing the Euclidean distance of the sensitivity/specificity pair to the point (1,1) resulted in virtually the same cut-offs as those corresponding to the intersection of the Richardson et al (2004) loss functions. The ‘optimal’ cut-off, whether it was found through loss functions or the Euclidean distance, was dependent on the hot spot relative risk as well as the scale factor. If the choice of cut-off would have depended only on the scale factor (i.e. the size of the expected counts), as was suggested by Richardson et al(2004), in practice we would only need to calculate expected counts and use the range of expected counts to find the required cut-off. As this choice also depends on the relative risks of the high-risk areas, it is impossible to determine the cut-off in advance. Rather than dichotomizing as ‘hot spot’ or ‘background’, perhaps it is more sensible to associate a degree of confidence according to ranges of *posterior proportions*. A similar approach has been used in the disease mapping studies of Pugliatti et al (2002) and Montomoli et al (2002) where prevalence proportions were mapped and areas with relative proportions exceeding 0.9 were considered to have higher prevalence than the norm, areas with relative prevalence between 0.75 and 0.9 showed some indication of having higher prevalence, etc. Since the cut-offs ranged from 0.6 to 0.92 for our realistic set of expected counts and relative risks, we could have categories of 0.65 to 0.75 corresponding to ‘possible hot spot’, 0.75 to 0.85 corresponding to ‘probable hot spot’, and 0.85 to 1 indicating a ‘hot

spot'. This would increase our chances of detecting hot spots with true relative risks of 1.5.

Our conclusions are limited to the spatial clustering pattern that we considered. We intended to study the accuracy of the BYM model given the spatial clustering pattern of several isolated hot spots consisting of a single area. The spatially structured component of the model tends to spread the high-risk to the surrounding areas, resulting in an increased rate of false negatives. Although this is the case for isolated hot spots consisting of a single area, it is possible that the BYM model achieves somewhat higher accuracy for other spatial clustering patterns. There is some suggestion of this in the results of Richardson et al (2004), where hot spots consisting of a cluster of contiguous areas were sometimes more accurately detected by the model.

We have seen from this simulation study that the BYM model has low accuracy as assessed through sensitivity, specificity and AUCs except when the hot spot relative risk is 2 and the expected counts of the hot spots are fairly large (in the range of 4 to 36) or when the relative risk is equal to 3 or more and the expected counts are over 1.2 in high-risk areas. However, due to the low prevalence of hot spots on our map, even when we find an unusually large *posterior proportion* (i.e. the model detects a hot spot) there is only a small chance that this will reflect a true high-risk area. This has a profound implication for disease mapping studies, namely that the most popular fully Bayesian disease mapping model may locate hot spots but these detected hot spots are unlikely to reflect areas of true elevated risk.

## CHAPTER 4 COMPARISON OF SPATIAL AND SPATIO- TEMPORAL SMOOTHING MODELS

It has been demonstrated that the BYM model has low accuracy in discriminating between high- and background-risk areas (Richardson et al, 2004) except in the case where the relative risk of the high-risk areas is large (i.e. over 2) and/or the expected counts are large. We have confirmed this (see chapter 3) and demonstrated that even when high-risk areas are detected by the model, the chance that these areas correspond to true high-risk areas is small. In this chapter, we consider alternative spatial and spatio-temporal models and assess their accuracy in discriminating between high- and background-risk areas.

### 4.1 Introduction

In chapter 3, we saw that the BYM model had low power in discriminating between high-risk and background-risk areas. Therefore, in this chapter we compare the accuracies of the BYM model, two spatial models and four spatio-temporal models in order to find a more accurate alternative to the BYM model in the context of locating isolated hot spots consisting of a single area.

When analyzing space-time data with a temporal component, for each area in the region, relative risks are estimated at each time point in the study period. Spatio-temporal modeling requires assumptions about the risk behaviour through time and these assumptions need to be incorporated via one or more temporal terms in our model. For example, if we believe in smooth transitions from one time period to the next, we can

add a temporal term that follows a random walk model. Once we have applied our spatio-temporal model, we obtain, for each area, a *posterior sample* of the ‘overall’ relative risk estimates in an appropriate way. The estimated *posterior proportion* (PP) is used to decide whether the area is a hot spot. There are two potential advantages to performing a spatio-temporal analysis instead of a strictly spatial analysis for the purpose of hot spot detection. First, a spatio-temporal analysis provides information regarding the pattern of risk through time. Second, the spatio-temporal model may stabilize the relative risk without relying exclusively on *local* and *global* spatial smoothing. The BYM model expresses the log of the relative risk as the sum of a spatially structured term and spatially unstructured term. The spatially structured term results in *local* spatial smoothing. *Local* spatial smoothing assumes spatial dependence and consequently shrinks extreme relative risk estimates to the average relative risk of surrounding areas (under an Intrinsic CAR prior). As a result, the spatially structured term may ‘oversmooth’ the disease map by masking small discontinuities in the risk surface. In other words, the spatially structured term may reduce the variance of the relative risk estimates but at the cost of increasing the bias of the estimates to the point where we may lose true hot spots. Temporal smoothing shrinks unusual relative risk estimates to the relative risks at other time points; thus, unlike *local* spatial smoothing there is not a loss of spatial differentiation in relative risks. If an area’s relative risks consistently follow a pattern through time except at one or two time points where the relative risk is unusually high or low, then the spatio-temporal modeling should produce a more accurate relative risk estimate because it will ‘shrink’ the unusual relative risk to the temporal pattern. In some real-world contexts one or two extreme RR’s through time may be realistic. In other contexts, there may be several small deviations from the true temporal pattern of relative risks that causes the RR for the study period to be over or underestimated. In this situation temporal smoothing may not be as

beneficial, but it is difficult to know if it will decrease the accuracy of classification. It is possible that accuracy may diminish due to the estimation of the additional (unhelpful) temporal parameters.

In this chapter we have two main objectives. First, we want to examine the changes in accuracy that result from including a spatially structured effect with an Intrinsic CAR prior. More specifically, we want to test the hypothesis that the spatially structured term hinders the detection of small, isolated hot spots. Second, we want to examine the effects on accuracy of including a temporal component to the spatial models.

To meet the first objective, we examine the accuracy of the BYM model without the spatially structured term. These models are referred to, by some, as *log-normal models*. They assume that the relative risks vary without spatial pattern. We also compare spatio-temporal models with the same temporal component that differ in whether or not they have spatially structured effects. For the second objective, we add temporal terms to the *log-normal* and BYM models and examine changes in accuracy. We also, compare two space-time interaction models that have either spatially structured effects or spatially unstructured effects.

We apply the *log-normal models* to the generated 'observed' counts from the assessment of the BYM model (chapter 3). For the spatio-temporal models we distribute the overall generated count amongst the 8 years of the study period 1995-2002. The generated count is distributed in different ways reflecting different types of temporal patterns. The six models are applied to the generated data and for each model we obtain a *posterior sample* of relative risk estimates for each of the 120 counties. Using the *posterior proportion*, the ROC curve is constructed and the corresponding area under the curve (AUC) estimated. Finally, we make

pair wise comparisons of model accuracy by finding the confidence intervals for the difference in AUCs between models.

## 4.2 Data and Methods

### 4.2.1 A review of the method of data generation in the assessment of the BYM model

Data generation in chapter 3 was based on the cervical cancer rate in Kentucky and the at-risk person-years in the state's 120 counties. For county  $i$  we calculated the expected count  $E_i$  and assigned a relative risk,

$RR_i$ . Background-risk areas were assigned a relative risk of 1 and hot spots were assigned a relative risk of 1.5, 2 or 3. The set of expected counts were multiplied by scale factors ( $sf$ ) equal to 0.5, 0.3 and 0.1 to examine model accuracy under varying degrees of stability. Recall that the amount of smoothing that takes place depends, in part, on the size of the expected counts. Counts of cases,  $Y_i$ ,  $i = 1, 2, \dots, 120$  were generated for each county by drawing from independent Poisson distributions with mean  $RR_i \cdot sf \cdot E_i$ . To summarize:

$$E_i = \text{at-risk person years}_i \times \text{cervical cancer rate}$$

$$Y_i \sim \text{Poisson}(RR_i \cdot sf \cdot E_i) \quad i = 1, 2, \dots, 120$$

$$sf = 0.1, 0.3, 0.5, 1; \quad RR_i = \begin{cases} 1.5, 2, \text{ or } 3 & \text{for hot spots} \\ 1 & \text{for background} \end{cases} .$$

This process was replicated 50 times to reduce the effects of random variability in the data generation. Thus 12 sets of 50 replications were generated corresponding to the 12  $RR - sf$  combinations.

In chapter 3, we applied the BYM model to the generated data and we found the AUCs for all twelve  $RR - sf$  combinations. In this chapter we consider two additional spatial smoothing models without spatially structured random effects. The log-normal model expresses the log of the relative risk as the sum of an intercept and a spatially unstructured effect for *global spatial smoothing*. This component is called a “heterogeneity” term and the amount of *global smoothing* that takes place is controlled, to some extent, by the choice of hyperprior for its precision. If we let  $Y_i$ ,  $E_i$  and  $RR_i$  represent the observed count, the expected count and the area-level relative risk, respectively, then the model specification is:

#### Log-normal model (LN)

$$\begin{aligned}
 Y_i &\sim Poisson(E_i \cdot RR_i) \\
 \log(RR_i) &= \alpha + V_i \\
 V_i &\sim N(0, \tau_V) \\
 \tau_V &\sim Gamma(\cdot, \cdot) \tag{4.1} \\
 \alpha &\sim Uniform(-\infty, +\infty) \\
 i &= 1, \dots, 120
 \end{aligned}$$

where  $V_i$  is the spatially unstructured effect for area  $i$ ,  $\alpha$  represents the overall level of  $\log(RR_i)$  and  $\tau_V$  is the precision of the spatially unstructured term. The first log-normal model we consider has a  $Gamma(0.001, 0.001)$  hyperprior for the precision and we will refer to this

model as LN1. The second, LN2, has precision that is assigned a  $\text{Gamma}(0.5, 0.0005)$  hyperprior.

#### **4.2.2 Generating time series of ‘observed’ counts for each area**

In chapter 3, for each county, we generated an ‘observed’ count,  $Y_i$ , for the study period. For the spatio-temporal analyses we need to distribute these counts amongst the eight years of the study period because spatio-temporal models are applied to time-series of observed (and expected) counts. Therefore, to assess the accuracy of the models with a temporal component, we generate sets of time series of observed counts,  $(Y_{i1}, Y_{i2}, \dots, Y_{i8})$ , for every  $RR - sf$  combination described in chapter 3.

Essentially, each ‘observed’ count of cervical cancer that was generated in the assessment of the BYM model is distributed amongst the 8 years of the study period, giving for each area an ‘observed’ count for each year from 1995 to 2002.

##### **4.2.2.1 Generating time series to reflect different temporal patterns of risk**

Recall that the ‘observed’ counts for the study period were generated from a Poisson distribution with mean equal to  $RR_i \cdot sf \cdot E_i$  for area  $i$ . Due to Poisson variability, the generated count differs from  $RR_i \cdot sf \cdot E_i$  (i.e. the number of cases that would result in a relative risk estimate equal to the true assigned relative risk). To distribute the generated count for the study period we consider that the deviation from  $RR_i \cdot sf \cdot E_i$  may have been the accumulation of ‘deviations from the truth’ occurring at several time points of the study period or it may have been due to ‘wild’ counts

occurring at one or two time points. We generate scenarios to reflect these two extremes. We consider four scenarios of the risk evolution over time. Each scenario determines the way in which the observed counts are allocated to each year in the study period. The first three scenarios assume that the reason for the difference in the ‘observed’ number of cases and  $RR_i \cdot sf \cdot E_i$  is the accumulation of several deviations from the true risk occurring at several time points in the study period. The fourth scenario assumes that the deviation from  $RR_i \cdot sf \cdot E_i$  occurred at a single time point in the study period.

To generate counts for area  $i$  through time,  $(Y_{i1}, Y_{i2}, \dots, Y_{i8})$ , with the overall relative risk,  $RR_i$ , that was assigned in chapter 3 (i.e. 1.5, 2, or 3 for hot spots and 1 for background-risk areas) we distribute the counts for 1995–2002 according to a multinomial distribution with number of objects equal to  $Y_i$ , the generated (observed) count for the area (i.e. county) from chapter 3.

#### 4.2.2.1.1 Scenario 1

For the first scenario we assume that the relative risks for the hot spot areas are elevated at a constant level through time. For both hot spots and background areas, the 8x1 probability vector for the multinomial consists of the proportion of person-years contributed by each year. Therefore, if the at-risk population is largest in the 5<sup>th</sup> year, for example, there is a higher probability of allocating a case to that year.

$$(Y_{i1}, Y_{i2}, \dots, Y_{i8}) \sim \text{Multinomial}(Y_i, p)$$

$$p = (at.risk_{i1} / at.risk_i, \dots, at.risk_{i8} / at.risk_i)$$

where  $at.risk_{it}$  is the at-risk person years for area  $i$  at time  $t$ , and  $at.risk_i$  is the total at-risk person-years for area  $i$  for the study period.

#### 4.2.2.1.2 Scenario 2

The second scenario assumes that the relative risks are increasing through time. For Scenario 2, the probability vector for hot spot areas is (8/148, 11/148, 14/148, ..., 29/148). For the background areas the counts are distributed according to the probability vector in Scenario 1.

$$(Y_{i1}, Y_{i2}, \dots, Y_{i8}) \sim Multinomial(Y_i, p)$$

$$p = (at.risk_{i1} / at.risk_i, \dots, at.risk_{i8} / at.risk_i) \quad \text{for background areas}$$

$$p = (8/148, 11/148, \dots, 29/148) \quad \text{for hot spots}$$

#### 4.2.2.1.3 Scenario 3

For the third scenario we distribute the cases to reflect a pattern of relative risks that increases to a maximum by the fifth year and stays at that level through to the eighth year. The probability vector for hot spot areas is (1/36, 2/36, 3/36, 4/36, 13/72, 13/72, 13/72, 13/72) and the counts for the background areas are distributed according to the probability vector in Scenario 1.

$$(Y_{i1}, Y_{i2}, \dots, Y_{i8}) \sim multinomial(Y_i, p)$$

$$p = (at.risk_{i1} / at.risk_i, \dots, at.risk_{i8} / at.risk_i) \quad \text{for background areas}$$

$$p = (1/36, 2/36, 3/36, 4/36, 13/72, 13/72, 13/72, 13/72) \quad \text{for hot spots}$$

#### 4.2.2.1.4 Scenario 4

For a single area, we generated a count that represents the number of cases occurring in the study period. Due to random variability, this generated count is either higher or lower than the number of cases that would result in the assigned (i.e. ‘true’) relative risk. We consider, in this scenario, that this deviation from the truth arose from an unusual number of cases occurring at a single time point, the seventh year of the study period, and the result is that the crude SMR for the study period is either larger or smaller than the true RR.

For Scenario 4, we distribute  $RR_i \cdot sf \cdot E_i$ , the number of cases that would result in the true relative risk,  $RR_i$ , rather than the generated count for the study period. We use a multinomial with probability vector as in Scenario 2 but add to the generated count in the seventh year, represented by  $X$  below, the difference between  $RR_i \cdot sf \cdot E_i$  and the generated count for the study period,  $Y_i$ . Thus the sum of the counts through time will equal the generated count for the study period but any deviation from  $RR_i \cdot sf \cdot E_i$  (i.e. the number of cases that results in the assigned  $RR_i$ ) is contributed by the seventh year (i.e. 2001).

$$(Y_{i1}, Y_{i2}, \dots, X, Y_{i8}) \sim \text{multinomial}(RR_i \cdot sf \cdot E_i, p)$$

$$p = (\text{at.risk}_{i1} / \text{at.risk}_i, \dots, \text{at.risk}_{i8} / \text{at.risk}_i) \quad \text{for background areas}$$

$$p = (8/148, 11/148, \dots, 29/148) \quad \text{for hot spots}$$

$$Y_{i7} = X + (Y_i - RR_i \cdot sf \cdot E_i)$$

### 4.2.3 Spatio-temporal models

Four different spatio-temporal models are applied to the four sets of possible disease count time series. The first model assumes that the log of the relative risk is the sum of a spatially unstructured random effect with a temporal term representing the effect of time. This is the *log-normal* model with  $\text{Gamma}(0.5, 0.0005)$  hyperprior with an added temporal term that follows a random walk model. The random walk model is flexible because it does not require the time series to be stationary, as opposed to the AR(1) model that has been suggested by some authors (Knorr-Held and Richardson, 2003; Waller et al, 1997). The stationarity assumption is restrictive because it requires that the statistical properties of the time series do not change over time. In particular, there can be no systematic change in mean or variance, or periodic variations. Although the random walk model is flexible, it assumes smooth changes from one time period to the next, which is an assumption that might not be appropriate for all disease mapping studies. We will refer to this model as a *temporal random walk model* (TRW). It is important to note that, here, the effect of time on the log relative risks is not modelled for each area independently. Rather, it estimates the effect of the time period on the log relative-risk for the entire study region.

Throughout this section, for all model specifications, we will use  $Y_{it}$  to represent the observed count for area  $i$  at time  $t$ ,  $RR_{it}$  will represent the relative risk in area  $i$  at time  $t$ , and  $E_{it}$  will denote the expected count in area  $i$  at time  $t$ . The specification of the Temporal Random Walk model is as follows:

## Temporal Random Walk (TRW) model

$$\begin{aligned}
Y_{it} &\sim \text{Poisson}(RR_{it} \cdot E_{it}) \\
\log(RR_{it}) &= \alpha + V_i + \delta_t \\
V_i &\sim N(0, \tau_V) \\
\delta_t \mid \delta_{-t} &\sim N(\delta_{t+1}, \cdot) \quad , t = 1 \\
&\sim N((\delta_{t-1} + \delta_{t+1}) / 2, \cdot) \quad , t = 2, \dots, T - 1 \\
&\sim N(\delta_{t-1}, \cdot) \quad , t = T \\
\sum_t \delta_t &= 0 \\
\alpha &\sim Uniform(-\infty, +\infty) \\
\tau_V &\sim Gamma(0.5, 0.0005) \\
\tau_\delta &\sim Gamma(0.01, 0.01) \tag{4.2} \\
i &= 1, \dots, n; t = 1, \dots, T.
\end{aligned}$$

where  $\delta_{-t}$  is the vector of all  $\delta$ 's except  $\delta_t$  and CAR is the Intrinsic CAR prior based on the normal distribution. The  $\delta_t$  can be interpreted as the effect (on  $\log(RR)$ ) of year  $t$ . In WinBUGS we use the *car.normal* prior to model the random walk for the  $\delta$ 's. Normally an intercept is not included in the model but, as the WinBUGS *car.normal* prior automatically constrains the sum of the  $\delta$ 's to zero, the inclusion of the intercept term  $\alpha$  is appropriate.

The second spatio-temporal model, introduced by Bernardinelli et al (1995a), is a linear trend model with spatially unstructured random effect terms for the intercept and slope. This model allows each area to have a different baseline log relative-risk with different linear rate of change. While there is an assumption that the intercepts and slopes are not too different from area to area, we do not assume that adjacent areas have

more similar intercepts and slopes. We call this a Temporal Linear Trend model (TLT) and the specification of the model is as follows:

### Temporal Linear Trend (TLT) model

$$\begin{aligned}
 Y_{it} &\sim Poisson(RR_{it} \cdot E_{it}) \\
 \log(RR_{it}) &= \alpha_i + \delta_i \cdot (t - \bar{t}) \\
 \alpha_i &\sim N(\mu_\alpha, \tau_\alpha) \\
 \delta_i &\sim N(\mu_\delta, \tau_\delta) \\
 \mu_\alpha &\sim N(0, 1000 000) \\
 \mu_\delta &\sim N(0, 1000 000) \\
 \tau_\alpha &\sim Gamma(0.001, 0.001) \\
 \tau_\delta &\sim Gamma(0.001, 0.001) \\
 i &= 1, \dots, n; t = 1, \dots, T.
 \end{aligned} \tag{4.3}$$

where  $\alpha_i$  is the area-specific baseline log relative-risk and  $\delta_i$  is the area-specific rate of change in the log relative risk. We center time ( $t$ ) to improve convergence of the MCMC implementation and the quality of the resulting *posterior sample*.

Two spatio-temporal models with terms for spatially structured effects are evaluated. The first is the Knorr-Held and Besag (1998) space-time model (KHB), that is, the BYM model with a term for temporally structured effects. The temporal effect assumes smooth changes through time and is implemented with a random walk model. Basically, this is the TRW model with an additional spatially structured effect for *local spatial smoothing*.

The model specification is as follows:

### Knorr-Held and Besag (KHB) model

$$\begin{aligned}
 Y_{it} &\sim \text{Poisson}(RR_{it} \cdot E_{it}) \\
 \log(RR_{it}) &= \alpha + U_i + V_i + \delta_t \\
 U_i \mid U_{j \neq i} &\sim \text{CAR}(\tau_U) \\
 V_i &\sim N(0, \tau_V) \\
 \delta_t \mid \delta_{-t} &\sim N(\delta_{t+1}, \cdot) \quad , t = 1 \\
 &\sim N((\delta_{t-1} + \delta_{t+1}) / 2, \cdot) \quad , t = 2, \dots, T-1 \\
 &\sim N(\delta_{t-1}, \cdot) \quad , t = T \\
 \sum_t \delta_t &= 0 \\
 \alpha &\sim \text{Uniform}(-\infty, +\infty) \\
 \tau_U &\sim \text{Gamma}(0.5, 0.0005) \\
 \tau_V &\sim \text{Gamma}(0.5, 0.0005) \\
 \tau_\delta &\sim \text{Gamma}(0.01, 0.01) \tag{4.4} \\
 i &= 1, \dots, n; t = 1, \dots, T.
 \end{aligned}$$

where, as before, the  $U_i$ ,  $V_i$  and the  $\delta_t$  represent the spatially structured term, the spatially unstructured term and the temporally structured term, respectively.

The last spatio-temporal model is the space-time interaction model of Bernardinelli et al (1995a) that assumes spatial structure (BC). It models the log relative-risk as a linear trend model with a spatially structured area-specific intercept and coefficient of the linear trend. We will abbreviate this model as BC.

The formulation of the BC model is:

Bernardinelli et al (1995a) (BC) model

$$\begin{aligned}
 Y_{it} &\sim Poisson(RR_{it} \cdot E_{it}) \\
 \log(RR_{it}) &= \alpha + U_i + (\beta + \delta_i) \cdot (t - \bar{t}) \\
 U_i \mid U_{j \neq i} &\sim CAR(\tau_U) \\
 \delta_i \mid \delta_{j \neq i} &\sim CAR(\tau_\delta) \\
 \alpha &\sim Uniform(-\infty, +\infty) \\
 \beta &\sim Uniform(-\infty, +\infty) \\
 \tau_U &\sim Gamma(0.5, 0.0005) \quad (4.5) \\
 \tau_\delta &\sim Gamma(0.5, 0.0005) \\
 i &= 1, \dots, n; t = 1, \dots, T.
 \end{aligned}$$

This space-time interaction model allows spatially structured variation in risk between areas at baseline and in the linear temporal trend. It modifies the Temporal Linear Trend model by imposing the assumption that adjacent areas have intercepts and slopes that are more alike than non-adjacent areas.

#### **4.2.4 Obtaining relative risk estimates for the study period from the results of the spatio-temporal models**

A spatio-temporal analysis results in area-level risk estimates through time. We consider the situation where we are interested in forming a single disease map for the study period. To this end, for each area, we produce a smoothed estimate of the SMR by summing the fitted (i.e. smoothed) counts through time and dividing by the expected count for the

study period (as in SMR=observed/expected). The fitted count for area  $i$  at time  $t$  is given by the product of the relative risk estimate,  $\hat{RR}_{it}$ , and the expected count,  $E_{it}$ . Thus, for each area, we combine the relative risks through time as follows:

$$\hat{RR}_i = \frac{\sum_{t=1}^8 (\hat{RR}_{it} \cdot E_{it})}{ref.rate \cdot at.risk_i}$$

$$E_{it} = ref.rate \cdot at.risk_i \quad (4.6)$$

where  $E_{it}$  is the expected count for area  $i$  at time  $t$ ,  $\hat{RR}_i$  represents the relative risk (or SMR) estimate for area  $i$ ,  $\hat{RR}_{it}$  is the relative risk estimate for area  $i$  at time  $t$ , and  $ref.rate$  is the reference rate for the study period (e.g. the incidence rate of cervical cancer for Kentucky for the 8 year period from 1995 to 2002).  $at.risk_i$  represents the at-risk person years for area  $i$  from 1995 to 2002, while  $at.risk_{it}$  is the at-risk person years for area  $i$  at time  $t$ . Thus, for each estimate in the *posterior sample* of  $(\hat{RR}_{i1}, \hat{RR}_{i2}, \dots, \hat{RR}_{i8})$ , we form one estimate for the *posterior sample* of  $\hat{RR}_i$ .

#### 4.2.5 Implementation

Data generation was accomplished with R software (R Development Core Team, 2005). Before applying the smoothing models to the generated data, we conducted preliminary runs on an additional set of data so that we could properly calibrate the number of iterations required for

convergence. Visual inspection of the trace plots and  $\hat{R}$  values(Gelman et al, 2003) from several runs suggested a minimum of 10 000 iterations were required per chain. To obtain smoothed risk estimates we used WinBUGS 1.4 (Spiegelhalter et al, 2002) and R (R Development Core Team, 2005; Thomas A et al, 2006) and convergence of the MCMC chains was assessed using the  $\hat{R}$  statistic in R (Gelman et al, 2003). Three chains of length 10 000 were obtained; every fifth iteration was retained after discarding the first 5000 as burn-in estimates giving a sample of 3000 estimates for each area.

#### **4.2.6 Comparing the accuracy of the models' classification**

The accuracy of the models in discriminating between high-risk and background-risk areas is assessed through ROC curves and their corresponding AUCs. The methods of Hanley and McNeil (1983) are used to test pair wise differences in models' AUCs. Their methods are described in section 2.3.

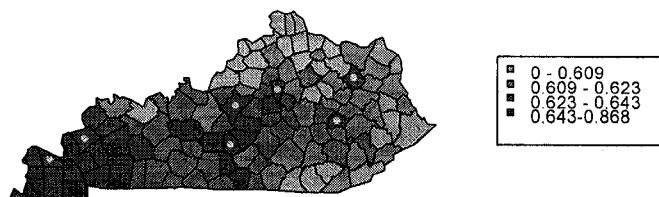
### **4.3 Results**

Figures 4.1 and 4.2 summarize the RR estimates through disease maps of the PP and *posterior mean* for Scenario 1 when the true hot spot RR is 1.5 and the scale factor is 1. The remaining disease maps for Scenario 1 are presented in Appendix A. The central cut-points in the legend correspond to the quartiles of the distribution of PP's and *posterior means* averaged over the 50 replications. There is some suggestion of a west-to-east gradient in the disease maps from the BYM, KHB, and BC models, even though a gradient was not generated in the true map. This is likely due to the effects of *local spatial smoothing* in which the hot spot with the largest

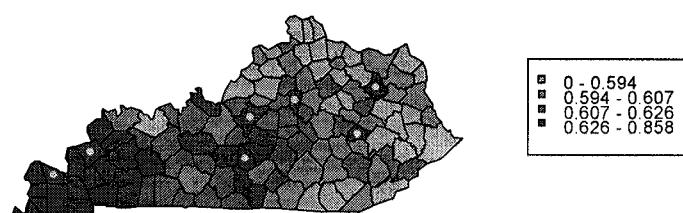
expected count of 36.3 in the west of the state has a heavy influence on the neighbouring areas. This effect is probably made worse by the other hot spot in the western tip (i.e. area 28). BC's gradient is most pronounced and this may be explained by the lack of a spatially unstructured term in this model.

Figure 4.1: Disease Maps of the average *posterior proportion* over 50 replications for Scenario 1, RR=1.5, sf=1. True high-risk areas are indicated by a white dot. The central cut-points correspond to the quartiles of the PP distribution.

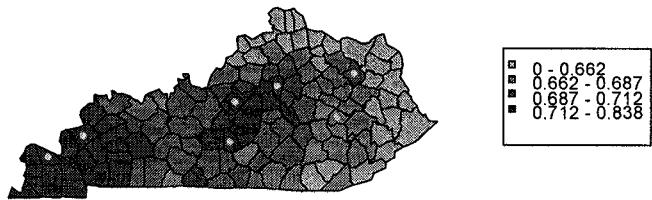
BYM



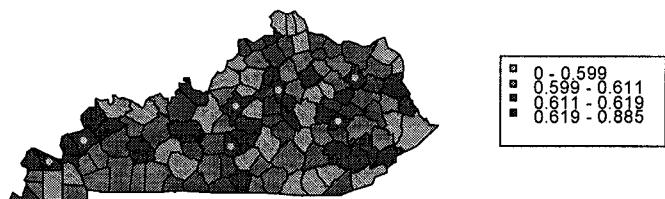
KHB



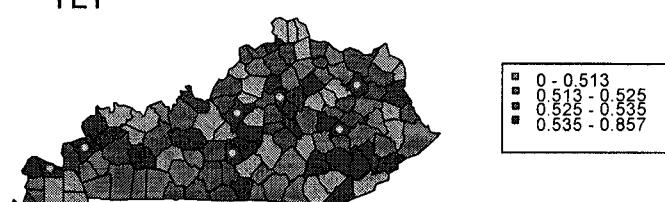
BC



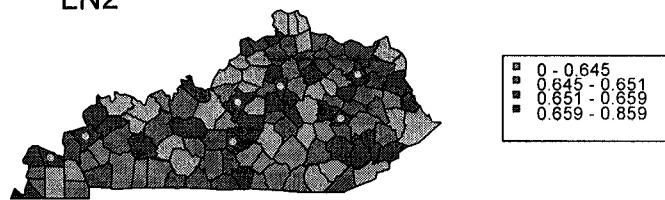
LN1



TLT



LN2



TRW

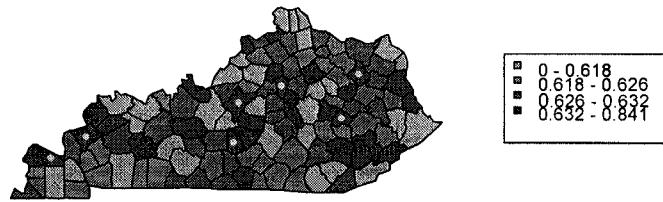
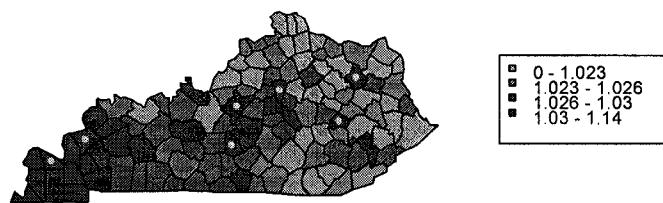
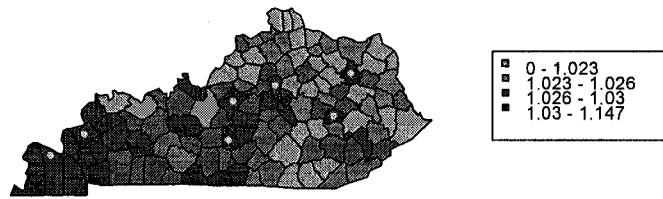


Figure 4.2: Disease maps of the average *posterior mean* RR over 50 replications, for Scenario 1, RR=1.5, sf=1. Hot spots are indicated by a white dot.

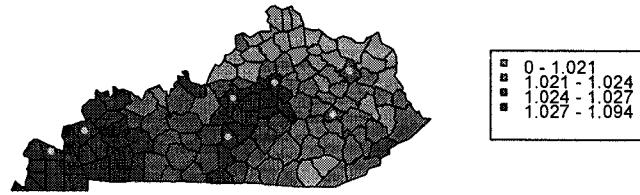
BYM



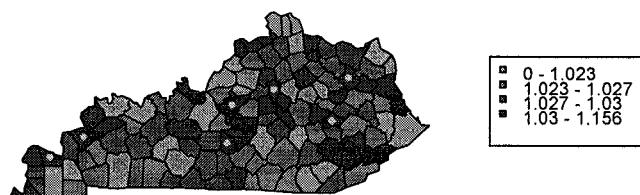
KHB



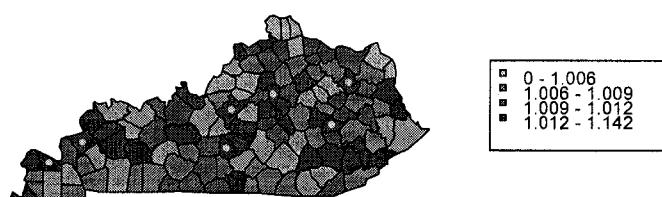
BC



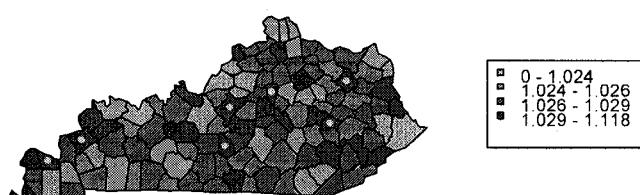
LN1

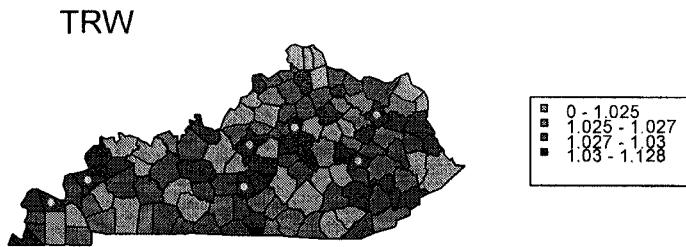


TLT



LN2





Tables 4.1 to 4.4 present the AUCs when the true hot spot RR is 1.5 and the scale factor is 1 for Scenario's 1 to 4. There is little difference in the results for Scenarios 1, 2 and 3 but there is a dramatic difference between the results for Scenarios 1 to 3 and those of Scenario 4. Table 4.5 and 4.6 gives the confidence limits for the differences in AUCs based on the Hanley and MacNeil (1983) test statistic. The ROC curves for RR=1.5 for Scenarios 1 and 4 are presented in Figures 4.3 and 4.4. See Appendix A for tables of confidence limits for all RR-sf combinations of Scenario1 and 4 and for RR=1.5 and sf=1 for Scenario 2 and 3. Also, in Appendix A, are the ROC curves for all four scenarios. The results indicate that there is little difference in accuracy from including a spatially structured term in the model as long as a spatially unstructured term is also included. We see this by looking at the differences in AUC between the BYM and LN2 models, the KHB and TRW models and the BC and TLT models. The difference between the BC and TLT models is likely due to the fact that the BC model has intercept and coefficient of linear trend modelled with a spatially structured effect, while the TLT model has spatially unstructured effects. In addition, the TLT model has a  $\text{Gamma}(0.001, 0.001)$  hyperprior for the spatially unstructured effects. The influence of the choice of hyperprior for the precisions is further confirmed by comparing the LN1 and LN2 models, where the LN1 model has significantly greater accuracy than the LN2 and BYM models (see Table 4.5). The PP disease maps

using the optimal cut-off for the LN1 and LN2 model show that LN2 results in more smoothing so that the three hot spots with the lowest expected counts are missed, whereas the LN1 model misses only two hot spots (Figure 4.4).

Table 4.1: Area under the ROC curve (AUC) for Scenario 1.

<b>Hot Spot RR</b>	<b>sf</b>	<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>1.5</b>	<b>1</b>	0.70	0.70	0.59	0.74	0.82	0.68	0.70
	<b>0.5</b>	0.62	0.63	0.54	0.67	0.76	0.60	0.61
	<b>0.3</b>	0.57	0.58	0.52	0.61	0.69	0.55	0.57
	<b>0.1</b>	0.52	0.52	0.51	0.54	0.60	0.51	0.52
<b>2</b>	<b>1</b>	0.92	0.92	0.81	0.93	0.94	0.92	0.93
	<b>0.5</b>	0.79	0.79	0.66	0.84	0.88	0.79	0.80
	<b>0.3</b>	0.65	0.64	0.55	0.72	0.81	0.63	0.64
	<b>0.1</b>	0.56	0.55	0.52	0.60	0.72	0.54	0.55
<b>3</b>	<b>1</b>	0.98	0.98	0.97	0.98	0.98	0.98	0.98
	<b>0.5</b>	0.97	0.97	0.94	0.97	0.97	0.97	0.97
	<b>0.3</b>	0.93	0.93	0.84	0.93	0.94	0.93	0.93
	<b>0.1</b>	0.68	0.67	0.57	0.75	0.83	0.67	0.68

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Table 4.2: Area under the ROC curve (AUC) for Scenario 2.

<b>Hot Spot RR</b>	<b>SF</b>	<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>1.5</b>	<b>1</b>	0.70	0.69	0.60	0.74	0.83	0.68	0.69
	<b>0.5</b>	0.62	0.61	0.54	0.67	0.77	0.60	0.60
	<b>0.3</b>	0.57	0.57	0.53	0.61	0.70	0.55	0.56
	<b>0.1</b>	0.52	0.52	0.51	0.54	0.62	0.51	0.51
<b>2</b>	<b>1</b>	0.92	0.92	0.81	0.93	0.94	0.92	0.93
	<b>0.5</b>	0.79	0.81	0.66	0.84	0.89	0.79	0.81
	<b>0.3</b>	0.65	0.65	0.56	0.72	0.81	0.63	0.65
	<b>0.1</b>	0.56	0.56	0.52	0.60	0.72	0.54	0.55
<b>3</b>	<b>1</b>	0.98	0.98	0.97	0.98	0.98	0.98	0.98
	<b>0.5</b>	0.97	0.97	0.93	0.97	0.98	0.97	0.97
	<b>0.3</b>	0.93	0.93	0.83	0.93	0.94	0.93	0.93
	<b>0.1</b>	0.68	0.68	0.57	0.75	0.83	0.67	0.69

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Table 4.3: Area under the ROC curve (AUC) for Scenario 3.

<b>Hot Spot RR</b>	<b>SF</b>	<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
1.5	<b>1</b>	0.70	0.70	0.61	0.74	0.83	0.68	0.69
	<b>0.5</b>	0.62	0.61	0.55	0.67	0.78	0.60	0.60
	<b>0.3</b>	0.57	0.57	0.53	0.61	0.69	0.55	0.56
	<b>0.1</b>	0.52	0.52	0.51	0.54	0.61	0.51	0.52
2	<b>1</b>	0.92	0.93	0.81	0.93	0.94	0.92	0.93
	<b>0.5</b>	0.79	0.81	0.66	0.84	0.89	0.79	0.81
	<b>0.3</b>	0.65	0.65	0.56	0.72	0.81	0.63	0.65
	<b>0.1</b>	0.56	0.56	0.52	0.60	0.72	0.54	0.55
3	<b>1</b>	0.98	0.98	0.97	0.98	0.98	0.98	0.98
	<b>0.5</b>	0.97	0.97	0.94	0.97	0.98	0.97	0.97
	<b>0.3</b>	0.93	0.93	0.83	0.93	0.94	0.93	0.93
	<b>0.1</b>	0.68	0.68	0.57	0.75	0.83	0.67	0.68

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Table 4.4: Area under the ROC curve (AUC) for Scenario 4.

<b>Hot Spot RR</b>	<b>SF</b>	<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>1.5</b>	<b>1</b>	0.70	0.91	0.83	0.74	0.94	0.68	0.91
	<b>0.5</b>	0.62	0.78	0.72	0.67	0.88	0.6	0.78
	<b>0.3</b>	0.57	0.64	0.62	0.61	0.79	0.55	0.62
	<b>0.1</b>	0.52	0.54	0.53	0.54	0.64	0.51	0.53
<b>2</b>	<b>1</b>	0.92	0.99	0.97	0.93	0.99	0.92	0.99
	<b>0.5</b>	0.79	0.97	0.87	0.84	0.97	0.79	0.97
	<b>0.3</b>	0.65	0.87	0.76	0.72	0.93	0.63	0.87
	<b>0.1</b>	0.56	0.60	0.60	0.60	0.81	0.54	0.58
<b>3</b>	<b>1</b>	0.98	1.00	1.00	0.98	1.00	0.98	1.00
	<b>0.5</b>	0.97	1.00	1.00	0.97	1.00	0.97	1.00
	<b>0.3</b>	0.93	1.00	0.99	0.93	1.00	0.93	1.00
	<b>0.1</b>	0.68	0.95	0.76	0.75	0.96	0.67	0.95

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Table 4.5: Confidence limits for difference in AUCs for Scenario 1, with hot spot relative risk equal to 1.5 and scale factor equal to 1. Entries correspond to the row heading AUC minus the column heading AUC. Confidence limits in bold are significant differences in model AUCs. Significance does not take into account multiple comparisons.

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0406	<b>0.0696</b>	<b>-0.0794</b>	<b>-0.1536</b>	-0.0191	-0.0379
	<b>97.5%</b>	0	0.0257	<b>0.1369</b>	<b>-0.0140</b>	<b>-0.0844</b>	0.0476	0.0284
<b>KHB</b>	<b>2.5%</b>	-0.0257	0	<b>0.0768</b>	<b>-0.0719</b>	<b>-0.1457</b>	-0.0117	-0.0305
	<b>97.5%</b>	0.0406	0	<b>0.1445</b>	<b>-0.0067</b>	<b>-0.0774</b>	0.0549	0.0358
<b>BC</b>	<b>2.5%</b>	<b>-0.1369</b>	<b>-0.1445</b>	0	<b>-0.1853</b>	<b>-0.2609</b>	<b>-0.1236</b>	<b>-0.1429</b>
	<b>97.5%</b>	<b>-0.0696</b>	<b>-0.0768</b>	0	<b>-0.1146</b>	<b>-0.1837</b>	<b>-0.0546</b>	<b>-0.0731</b>
<b>LN1</b>	<b>2.5%</b>	<b>0.0140</b>	<b>0.0067</b>	<b>0.1146</b>	0	<b>-0.1046</b>	<b>0.0281</b>	<b>0.0093</b>
	<b>97.5%</b>	<b>0.0794</b>	<b>0.0719</b>	<b>0.1853</b>	0	<b>-0.0400</b>	<b>0.0937</b>	<b>0.0745</b>
<b>TLT</b>	<b>2.5%</b>	<b>0.0844</b>	<b>0.0774</b>	<b>0.1837</b>	<b>0.0400</b>	0	<b>0.0979</b>	<b>0.0801</b>
	<b>97.5%</b>	<b>0.1536</b>	<b>0.1457</b>	<b>0.2609</b>	<b>0.1046</b>	0	<b>0.1685</b>	<b>0.1484</b>
<b>LN2</b>	<b>2.5%</b>	-0.0476	-0.0549	<b>0.0546</b>	<b>-0.0937</b>	<b>-0.1685</b>	0	-0.0523
	<b>97.5%</b>	0.0191	0.0117	<b>0.1236</b>	<b>-0.0281</b>	<b>-0.0979</b>	0	0.0143
<b>TRW</b>	<b>2.5%</b>	-0.0284	-0.0358	<b>0.0731</b>	<b>-0.0745</b>	<b>-0.1484</b>	-0.0143	0
	<b>97.5%</b>	0.0379	0.0305	<b>0.1429</b>	<b>-0.0093</b>	<b>-0.0801</b>	0.0523	0

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Table 4.6: Confidence limits for difference in AUCs for Scenario 4, with hot spot relative risk equal to 1.5 and scale factor equal to 1. Entries correspond to the row heading AUC minus the column heading AUC. Confidence limits in bold are significant differences in model AUCs. Significance does not take into account multiple comparisons.

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	<b>-0.2404</b>	<b>-0.1665</b>	<b>-0.0794</b>	<b>-0.2732</b>	-0.0191	<b>-0.2468</b>
	<b>97.5%</b>	0	<b>-0.1780</b>	<b>-0.0992</b>	<b>-0.0140</b>	<b>-0.2084</b>	0.0476	<b>-0.1835</b>
<b>KHB</b>	<b>2.5%</b>	<b>0.1780</b>	0	<b>0.0477</b>	<b>0.1326</b>	<b>-0.0538</b>	<b>0.1912</b>	-0.0283
	<b>97.5%</b>	<b>0.2404</b>	0	<b>0.1050</b>	<b>0.1924</b>	<b>-0.0094</b>	<b>0.2555</b>	0.0164
<b>BC</b>	<b>2.5%</b>	<b>0.0992</b>	<b>-0.1050</b>	0	<b>0.0520</b>	<b>-0.1359</b>	<b>0.1125</b>	<b>-0.1110</b>
	<b>97.5%</b>	<b>0.1665</b>	<b>-0.0477</b>	0	<b>0.1203</b>	<b>-0.0799</b>	<b>0.1816</b>	<b>-0.0535</b>
<b>LN1</b>	<b>2.5%</b>	<b>0.0140</b>	<b>-0.1924</b>	<b>-0.1203</b>	0	<b>-0.2247</b>	<b>0.0281</b>	<b>-0.1979</b>
	<b>97.5%</b>	<b>0.0794</b>	<b>-0.1326</b>	<b>-0.0520</b>	0	<b>-0.1635</b>	<b>0.0937</b>	<b>-0.1390</b>
<b>TLT</b>	<b>2.5%</b>	<b>0.2084</b>	<b>0.0094</b>	<b>0.0799</b>	<b>0.1635</b>	0	<b>0.2222</b>	<b>0.0042</b>
	<b>97.5%</b>	<b>0.2732</b>	<b>0.0538</b>	<b>0.1359</b>	<b>0.2247</b>	0	<b>0.2878</b>	<b>0.0471</b>
<b>LN2</b>	<b>2.5%</b>	-0.0476	<b>-0.2555</b>	<b>-0.1816</b>	<b>-0.0937</b>	<b>-0.2878</b>	0	<b>-0.2610</b>
	<b>97.5%</b>	0.0191	<b>-0.1912</b>	<b>-0.1125</b>	<b>-0.0281</b>	<b>-0.2222</b>	0	<b>-0.1977</b>
<b>TRW</b>	<b>2.5%</b>	<b>0.1835</b>	-0.0164	<b>0.0535</b>	<b>0.1390</b>	<b>-0.0471</b>	<b>0.1977</b>	0
	<b>97.5%</b>	<b>0.2468</b>	0.0283	<b>0.1110</b>	<b>0.1979</b>	<b>-0.0042</b>	<b>0.2610</b>	0

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Figure 4.3: ROC for Scenario 1, for maps with hot spot relative risk equal to 1.5. Each curve corresponds to the ROC curve of a particular spatial or spatio-temporal model over 50 replications.

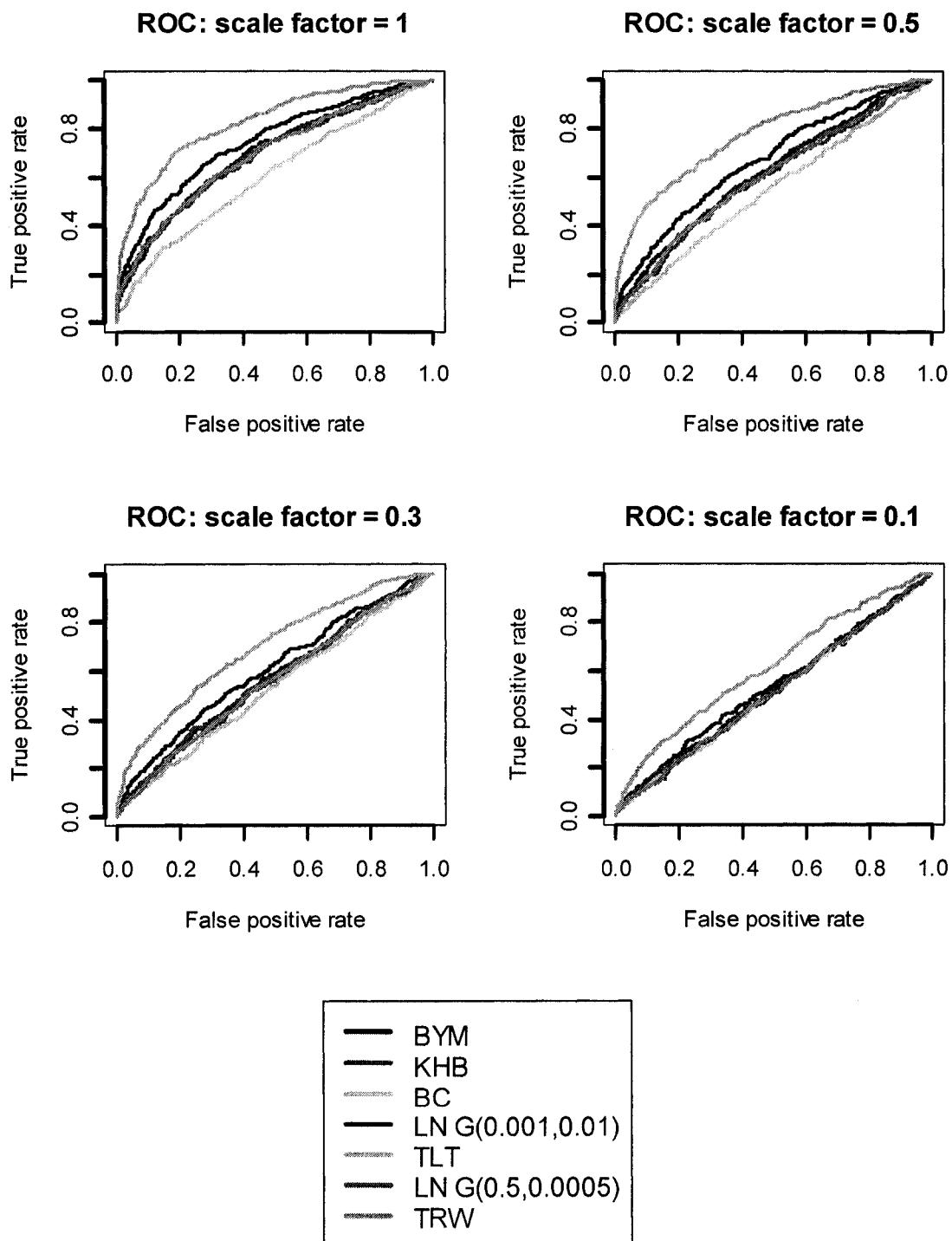


Figure 4.4: ROC for Scenario 4, for maps with hot spot relative risk equal to 1.5. Each curve corresponds to the ROC curve of a particular spatial or spatio-temporal model over 50 replications.

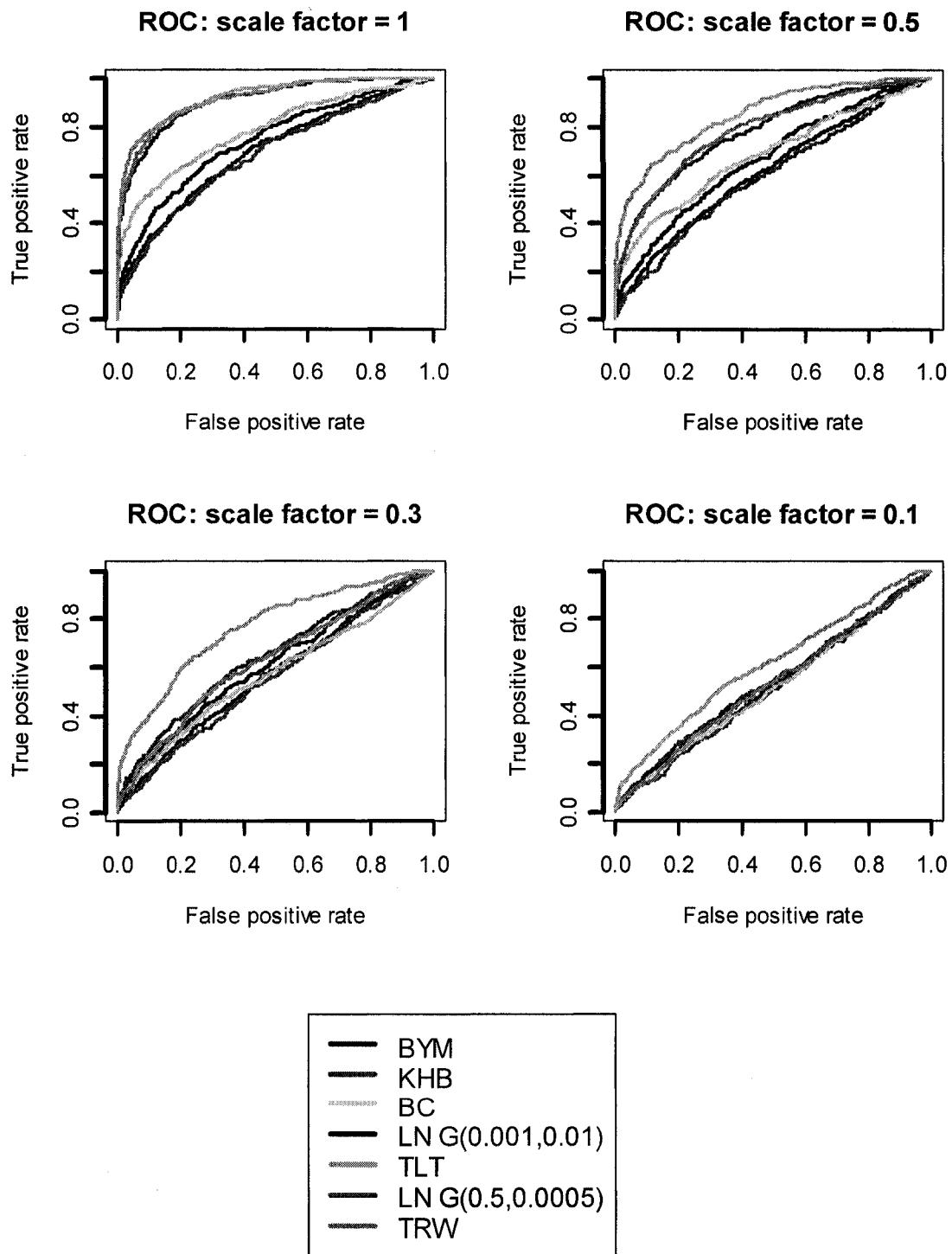
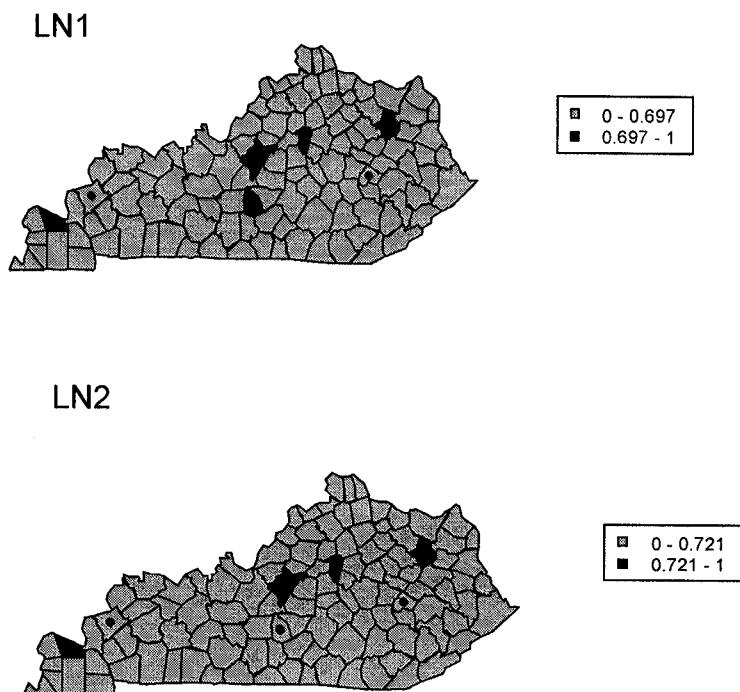


Figure 4.5: Disease maps of PP for log-normal models with ‘optimal’ threshold. The ‘optimal’ threshold for LN1 and LN2 are 0.697 and 0.721, respectively. Areas with black dots are true hot spots.



The maximum  $\hat{R}$  value amongst the  $120 \times 8$  relative risk estimates (120 estimates for spatial models) for a single replication was used to assess convergence for each model. The maximum  $\hat{R}$  values for all models and  $RR - sf$  combinations for Scenario 1 are presented in Table 4.7 (see Appendix A for maximum  $\hat{R}$  values for Scenario 2 to 4). We see that most maximum  $\hat{R}$  values are less than 1.1, thus convergence has been achieved in those cases. For three  $RR - sf$  combinations, the BC model has maximum  $\hat{R}$  values over 1.1 (i.e. 1.11, 1.15 and 1.16) and for the LN2 model, in two cases we have values over 1.1 (i.e. 1.12 and 1.14). These are the largest values from all 960  $RR$  estimates for the BC model and all

120  $RR$  estimates for the LN2 model; thus, it is unlikely that convergence of the algorithm had an important effect on the results.

Table 4.7: Maximum  $\hat{R}$  value for a single simulation of Scenario 1.

Hot Spot RR	SF	BYM	KHB	BC	LN1	TLT	LN2	TRW
1.5	<b>1</b>	1.02	1.02	1.03	1.01	1.01	1.03	1.00
	<b>0.5</b>	1.05	1.01	1.02	1.02	1.01	1.02	1.10
	<b>0.3</b>	1.03	1.01	1.03	1.02	1.02	1.04	1.01
	<b>0.1</b>	1.01	1.03	1.15	1.04	1.01	1.02	1.00
2	<b>1</b>	1.02	1.01	1.10	1.00	1.00	1.00	1.01
	<b>0.5</b>	1.03	1.01	1.02	1.00	1.01	1.00	1.01
	<b>0.3</b>	1.08	1.03	1.11	1.01	1.01	1.12	1.00
	<b>0.1</b>	1.10	1.04	1.16	1.03	1.02	1.03	1.02
3	<b>1</b>	1.00	1.01	1.01	1.00	1.01	1.00	1.00
	<b>0.5</b>	1.00	1.01	1.01	1.00	1.00	1.00	1.01
	<b>0.3</b>	1.02	1.01	1.02	1.00	1.01	1.00	1.01
	<b>0.1</b>	1.04	1.05	1.01	1.02	1.03	1.14	1.03

Thus, the results indicate that the models with spatially structured terms add an artefactual gradient to the map but do not result in lower accuracy unless a spatially unstructured term is omitted from the model. However, the choice of hyperprior for the precision of the spatially unstructured term is crucial. The  $Gamma(0.5, 0.0005)$  results in more smoothing than the  $Gamma(0.001, 0.001)$  and hot spots with small expected counts are not detected.

For Scenarios 1 to 3, adding a temporal component to the spatial model does not significantly change the accuracy in classification. The exception

is the TLT model. Adding a term for linear trend to the LN1 model significantly increases accuracy. This is probably a consequence of the space-time interaction which gives another opportunity to express the heterogeneity in relative risk between areas. The TRW model, however, estimates the effect of time for the entire study region thus it should not be as accurate as the TLT model. In general, the TLT model appears to yield better estimates of the relative risks of background areas. The PP maps show that, in general, the 75<sup>th</sup> percentile of TLT's distribution of PP's is lower than that of the other models. For example, for Scenario 1, RR=1.5 and sf=1, the 75<sup>th</sup> percentile of the distribution of PP's is 0.54 for the TLT model, but 0.62 or higher for the other models.

In generating counts for Scenario 4, we assumed that the seventh year had an unusual count (either smaller or larger than what it should be, given the temporal pattern and the true RR of the area). For this scenario, there was a substantial improvement in accuracy for all spatio-temporal models (as compared to the first three scenarios). With few exceptions, the spatio-temporal models were more accurate than the spatial models. Even the BC model performed significantly better than the spatial models for many *RR – sf* combinations (Table 4.6 and Appendix).

#### 4.4 Discussion

The BYM model is the most popular spatial model in the non-statistical literature, yet, of the seven models that were assessed in this simulation study, it was not the most accurate for detecting small, isolated hot spots. In our simulations, the BYM model with  $\text{Gamma}(0.5, 0.0005)$  hyperpriors for the precisions is never significantly better than the log-normal model with  $\text{Gamma}(0.001, 0.001)$  hyperprior and, in fact, is usually inferior. However, for the log-normal models we found a significant influence of the priors for the purposes of identifying high-risk areas. The log-normal model with  $\text{Gamma}(0.5, 0.0005)$  hyperprior for the precision was not significantly more (or less) accurate than the BYM model. This is an interesting finding considering that Bernardinelli et al (1995b) conclude that in their simulation study investigating the impact of priors on the posterior distribution of the RR estimates from applying the BYM model, the choice of prior had little impact on the posterior distribution. Furthermore, the  $\text{Gamma}(0.5, 0.0005)$  is an often used hyperprior for the precisions of the BYM model, the log-normal model and various spatio-temporal models. The motivation for its use comes from the belief that other more non-informative priors place too much mass on small precisions (i.e. large variances). A large variance can result in a wide range of RR values in the prior, which is considered unrealistic (Bernardinelli, 1995; Wakefield, 2000a). Thus, the  $\text{Gamma}(0.5, 0.0005)$  serves to confine the relative risk values. The WinBUGS/GeoBUGS manual uses this hyperprior in most CAR examples and other disease mapping references either recommend their use or use them extensively in their examples (Lawson et al, 2003; Wakefield et al, 2000a). We found that perhaps the  $\text{Gamma}(0.001, 0.001)$  prior allowed more ‘breathing room’ for the relative risks so that areas with true elevated risk could have larger relative risk estimates. Thus, at least for the purposes of identifying isolated high-risk areas, the

$\text{Gamma}(0.5, 0.0005)$  prior for the precision of the log-normal model is not always the optimal choice.

We also note that removing the spatially structured effect in the KHB and BYM models (giving the TRW and LN2 models, respectively) did not improve accuracy. However, replacing the spatially structured effect for a spatially unstructured effect in the BC model greatly improved accuracy, though some of this improvement may have been due to the  $\text{Gamma}(0.001, 0.001)$  prior that was used for the spatially unstructured random effects of the TLT model.

The comparison of the spatial model with and without a temporal component shows us that, in the worst case, the temporal term will not result in a significant decrease in accuracy, and can at times result in significantly higher accuracy. Thus, there is no loss in accuracy from applying the spatio-temporal model to all applications that aim to locate isolated hot spots consisting of a single area. In fact, by applying a spatio-temporal model, there is the added advantage that we are able to examine the relative risk behavior through time. For example, the space-time interaction models with linear trend can indicate which areas have the sharpest increase or decrease in trend. This can be particularly meaningful for areas with larger expected counts where the trend is less likely to have occurred by chance. For the spatio-temporal models with a random walk prior, we obtain estimates for the effect of the time period on the log of the relative risks for the study region. It is important, however, to include a spatially unstructured term in the model, as the results of the BC model clearly indicate that the model with only spatially structured random effects has very low accuracy in identifying the high-risk areas, whereas the same model with only spatially unstructured effects, i.e. the TLT model, has high accuracy.

Our simulation study focused on the detection of isolated hot spots consisting of a single area and we considered four patterns through time. The conclusions are applicable for studies in which there is an *a priori* belief that any hot spots will consist of a single area. Future research should investigate whether these conclusions also hold for hot spots consisting of clusters of areas with various cluster shapes (i.e. circular or linear clusters). Different temporal patterns should also be investigated. In this work we used the same hot spots throughout the study and it is possible that the specific positions of these seven hot spots could affect the results to some extent. An assessment of the same map with a different set of seven hot spots might be informative.

In summary, when hot spots consist of a single area, the fully Bayesian disease mapping models with spatially structured random effects are not less accurate than the same models without this term. In choosing between the Bernardinelli et al (1995a) space-time interaction model with spatially structured effects versus the model with spatially unstructured effects, it is better for the purposes of identifying isolated high-risk areas to choose the model with spatially unstructured effects. Adding a temporal term to a spatial model results in at least the same degree of accuracy as the spatial model and has the added benefit of allowing the visual examination of risk through time. When deviations from the true risk appeared to have been caused by unusual counts at one or two time periods in the study period there may be substantial increases in accuracy.

For the case of detecting isolated hot spots consisting of a single area, we found that, for our generated data, the BYM model was not the most accurate model, in fact, the simpler log-normal model with  $\text{Gamma}(0.001, 0.001)$  prior for the precision was more accurate. The TLT model had even higher accuracy than the log-normal model. An

advantage of using the TLT model is the possibility of mapping the area-level coefficient of the linear trend so that we can examine the temporal trend of relative risk estimates for each area. Therefore, before applying the BYM model, one needs to carefully consider the nature of the disease and the characteristics of the populations of the region. For example, there should be some thought as to whether spatial dependence is present given the nature of the disease and the inhabitants of the various areas. Also, if the objective is to look for high-risk areas, is it more plausible that there would be clusters of high-risk areas or single isolated high-risk areas? Time series plots of the crude SMR for the areas with larger expected counts might be revealing as well as comparisons of the crude SMRs of contiguous areas with larger expected counts. In general, because we aim to minimize ecological biases by using the map with the finest partition, the population sizes are fixed so that we do not control the sample size to increase precision. Since there is typically too little information in the data to give precise estimates, assumptions about patterns in space and time are needed to increase precision. However, the assumptions that are made should be based on the substantive problem, the amount of information in the data, the region under study and how the region is partitioned.

## CHAPTER 5 APPLICATION TO MULTIPLE SCLEROSIS

### 5.1 Introduction

Multiple Sclerosis (MS) is a chronic disease of the central nervous system that is usually diagnosed between the ages of 25 and 40 years. It affects two to three times more women than men, and Caucasians, particularly those of northern European descent, are at highest risk. In the MS patient the fatty protective layer (myelin) of the axons of the central nervous system are damaged or destroyed in patches, resulting in symptoms as varied as sensory disturbances, difficulties with muscle coordination and bladder/bowel problems. The course of the disease varies and is difficult to predict at the time of diagnosis, but many patients experience periods of remission where they are relatively symptom-free. Many patients will live a normal life span but the burden of the disease is heavy as the patient requires assistance, in various forms, for much of her life.

At this time, MS is thought to be caused by interplay between genetics and environment: the individual who is genetically susceptible (although no genes have yet been identified) and who is exposed to an environmental agent at a critical time in their development, is more likely to develop the disease when exposed to an environmental trigger (Kahana 2000; Pugliatti et al 2002; Granieri 1997). Small-area disease mapping studies have been performed to isolate potential environmental risk factors. Some investigations were more descriptive in their purpose, in that they aimed to locate high and low risk areas, while others were ecological studies that estimated the association between risk in space-time and one or more environmental factors. Examples of hypothesized risk factors based on these studies include viral agents (Kurtzke and Heltberg, 2001), smoke

preservation of meats (Lauer 1995) and climate (i.e. humidity and cold) (Alperovitch and Bouvier, 1982).

In general, MS prevalence is thought to be correlated with geographical latitude, in that a geographical position further from the equator usually corresponds to higher MS risk. Multiple sclerosis on the Island of Sardinia (off the west coast of Italy) has been found to have an unexpectedly high prevalence given its geographical latitude. On December 31, 1997, the prevalence was estimated at 150 per 100 000 population (Pugliatti et al, 2005). Sardinia has been the location for a number of epidemiological studies of MS. Bayesian disease mapping studies have been conducted to link environmental factors to risk of MS in Sardinia. The disease mapping studies by Montomoli et al (2002) and Pugliatti et al (2002) examined the distribution of MS prevalence in central and northern Sardinia, respectively. Sardinia's population is ideal for a disease mapping study that aims to identify environmental risk factors because the population is thought to be more genetically homogeneous than most other Caucasian populations (Pugliatti et al, 2002, Montomoli et al, 2002). In addition, because there is little migratory movement in Sardinia, when a case of MS arises, contact with the putative cause will have most likely occurred in the commune of residence. Montomoli et al (2002) reported higher risks amongst areas that were particularly isolated, implying increased genetic susceptibility due to inbreeding. Pugliatti et al (2002) presented disease maps of prevalence by area of residence at diagnosis and during the ages of 5 to 15 years. They report clustering in the southwestern part of the Sassari province of Sardinia and the presence of a west-to-east gradient. Pugliatti et al (2002) concluded that further spatio-temporal cluster studies could help uncover etiologic clues for MS.

In this chapter, we present the spatial analysis of MS incidence for Sassari, the northern province of Sardinia. The data consist of case

information from the 90 communes of the province of Sassari, from 1971 to 2005. Time and commune of residence at onset were ascertained for 774 cases. Spatial and spatio-temporal analyses are performed to locate areas of elevated risk, demonstrating the methods outlined in chapter 4.

## 5.2 Data and Methods

Sardinia is an Italian island off the western coast of Italy, in the Mediterranean Sea (Figures 5.1 and 5.2). It has a population of approximately 1.65 million and the geographic region is approximately 24 090 km<sup>2</sup>. Sardinia is divided into 366 administrative communes, 90 of which make up the province of Sassari, which is also known as Northern Sardinia. Northern Sardinia has six distinct linguistic geographical areas. There has been very little migratory movement into Sardinia and within the island itself. Official statistics for the region indicate that in 1997, only 1.8% of the population reported having moved to the province of Sassari from another province or country while only 1.6% moved away from Northern Sardinia (Pugliatti et al, 2002). This feature facilitates the interpretation of a spatial analysis of MS risk, as it is more feasible to link environmental exposures to a single place of residence.

Cases were ascertained through the MS Registry at the Institute of Clinical Neurology of the University of Sassari. This is the major referral centre for MS patients living in Northern Sardinia (Pugliatti et al, 2005). Dates of birth, death and onset of the disease, as well as information on current residence, residence at birth and residence during the ages 5 to 15 years were recorded. In addition, for many cases, the dates and locations of migratory movements were noted. Other information, such as the symptoms at onset and the disease classification were also recorded. Date of onset was defined as the date of appearance of the first symptom(s) of MS.

Figure 5.1: Map of Italy and the Island of Sardinia



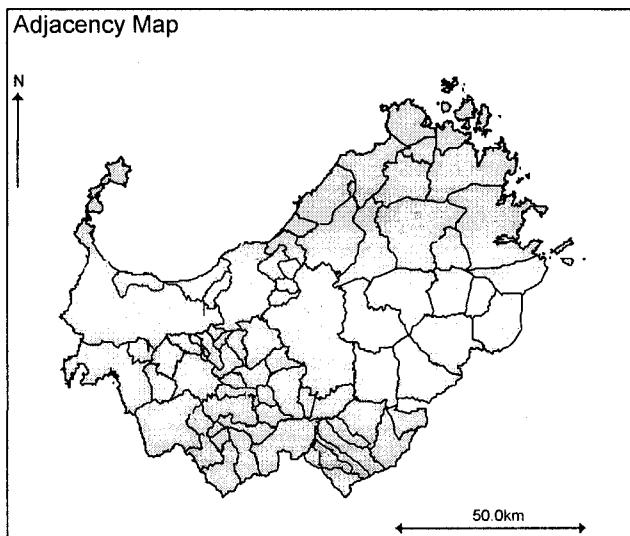
Figure 5.2: Map of the Island of Sardinia with the Province of Sassari in the North



At the time of this analysis, there were 1221 MS cases with 41 cases residing outside of the province of Sassari. Of the 1180 cases living in Sassari, 1022 had a date at onset recorded but the commune at onset was determined for only 774 cases. The year of onset for the 774 cases ranged from 1934 to 2006. A total of 690 cases had date of onset occurring in the study period, 1971 to 2005, and these were the cases that were used in the analysis. Some cases, particularly the more recently diagnosed cases, were missing detailed information regarding migratory movements. For these cases, when the same commune was recorded for residence at birth, residence during the ages 5 to 15 years and residence at registration, the 'residence at onset' was assumed to be the same. This seems reasonable, considering the low rates of internal migration that are reported for this region. Census data, for the years 1971, 1981, 1991, and 2001, were used to estimate the age-sex specific at-risk person-years for each area. The study period was divided into seven five-year intervals (i.e. 1971-1975, 1976-1980,..., 2001-2005).

The map of Sardinia had changed somewhat throughout the study period, 1971 to 2005. Nine new communes were formed from previously existing communes giving a total of 90 communes. In the analysis, these communes were re-integrated. However, when the new commune was created from two or more communes, these communes had to be combined to give consistency through time. This resulted in a map with 74 areas. Expected counts were calculated for these 74 and six were found to have very small expected counts (smaller than 1), so these were combined with adjacent communes giving a map with 69 areas (Figure 5.3). Care was taken to only combine communes from the same linguistic group since previous studies have suggested that, in Northern Sardinia, linguistic group may be associated with MS risk (Pugliatti et al, 2002). The 69 areas were numbered from 1 to 69 and we refer to the areas by these labels.

Figure 5.3: Map of Sassari partitioned into 69 areas



Combining areas to compensate for too small expected counts is motivated by the concluding remarks from the Richardson et al (2004) study evaluating the accuracy of Bayesian disease mapping models. They suggest that there should be some balancing of area size with magnitude of expected counts and *a priori* belief in the magnitude of the hot spot relative risks. It seems sensible that through the choice of partition of the study region, there should be some balancing of ecological bias (resulting from aggregating groups with different risks) with instability of the estimates (due to small population size). Although disease maps are typically constructed using the finest subdivision available to minimize ecological biases, we have to consider the size of the expected counts as well, since imprecise risk estimates may also lead to erroneous conclusions.

MS risk is strongly associated with age and sex, so *internal standardization* of the expected counts is necessary to account for the

differences in area-level age-sex distributions. *Internal standardization* uses the age-sex rates of the study region as reference rates to calculate the expected counts (MacNab and Dean, 2002). To find the expected count for an area we sum over the expected counts for each age-sex strata. Letting  $E_{it}$  represent the expected count for area  $i$  at time  $t$ ,  $ref.rate_{jk}$  represent the strata-specific reference rate of MS for Northern Sardinia, for age group  $j$  and sex  $k$ , and  $at.risk_{ijkt}$  represent the at-risk person-years for age group  $j$  and sex  $k$  in area  $i$  at time  $t$ , the expected count is calculated as follows:

$$E_{it} = \sum_{j,k} ref.rate_{jk} \cdot at.risk_{ijkt}$$

To find the overall expected count for an area (i.e. for the study period),  $E_i$ , we sum the area's expected counts through time.

$$E_i = \sum_t E_{it}$$

The age groups used in this analysis are < 5, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, and >74 years of age.

An assumption of the disease mapping model based on the Poisson (or Binomial) distribution, is the *proportionality assumption* (Wakefield et al, 2000a). The proportionality assumption requires that the expectation of the rate of disease in the  $j^{th}$  stratum of area  $i$ ,  $p_{ij}$ , is proportional to the area-level relative risk,  $RR_i$ , times the rate of disease in the  $j^{th}$  stratum of the study region,  $p_j$ . To check the validity of the assumption, it is customary to plot on the same axes, the observed  $\hat{p}_{ij}$  versus the

observed  $\hat{p}_j$  for areas with larger expected counts (Pascutto et al, 2000). Proportionality can be assumed to hold when the points for each area form a line with slope equal to one and the lines are parallel.

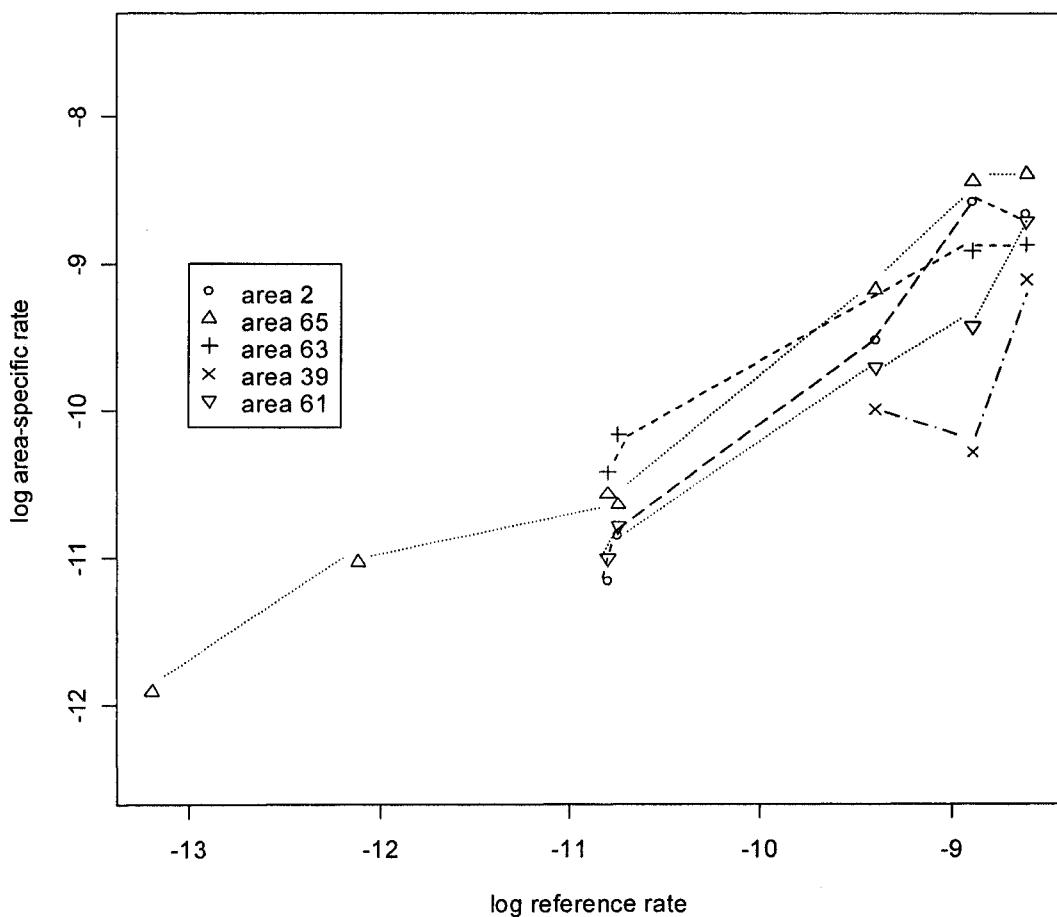
To produce a smoothed map and to identify potential hot spots we employ four of the seven models that were evaluated in Chapter 4: the BYM model, the log-normal model, the TLT and TRW models. We will not use the Bernardinelli et al (1995a) model with spatially structured effects for the intercept and coefficient of the linear trend because it did not perform as well as the others in any of the scenarios of the simulation study. In addition, the KHB model will not be used because  $\hat{R}$  values indicated that, even after 30 000 iterations, the KHB model was not converging very well. As the KHB model and the TRW model gave similar results for every simulation in chapter 4, only the TRW was used to analyse the MS data. All methods are described in section 4.2.

The Deviance Information Criterion (DIC) was used to compare model fit. The DIC is the sum of the posterior mean deviance (i.e. -2 times the log likelihood) plus a measure of the effective number of parameters,  $p_D$  (Spiegelhalter et al, 2002). The posterior mean deviance is a measure of the overall fit of the model, while  $p_D$  penalizes overly complex models that would otherwise tend to result in better fit. If there is much 'borrowing of strength' in that the prior provides much of the information in the posterior distribution, the effective number of parameters will be small relative to the number of datapoints (Best et al, 2005). The model with the smaller DIC is said to be the better model, though models with differences in DIC of 1 or 2 are thought to be equivalent, a DIC within 3 to 7 of the best model is not as adequate and larger differences indicate an inferior model (Spiegelhalter et al, 2002).

### 5.3 Results

We plot the age-specific rates of MS for Northern Sardinia versus the area-level age specific rates for five areas with larger expected counts and found that; overall, the proportionality assumption appears to hold (Figure 5.4).

Figure 5.4: Graphs of the log of the age-specific reference rates versus the log of the area-level age-specific rates for five areas with large expected counts. For the proportionality assumption to hold for each area the points should roughly form a line with slope equal to one and area-level graphs should be parallel.



Comparing the distribution of expected counts of cervical cancer in Kentucky and the distribution of expected counts of Multiple Sclerosis in Northern Sardinia (Table 5.1), we see that the expected counts of MS are smaller than the expected counts of cervical cancer by a factor of 0.3 to 0.5. In chapter 4, when the scale factor was 0.3 or 0.5, the models were able to detect high-risk areas with relative risks of 3. When the relative risks were smaller, only the TLT model had relatively decent accuracy in classifying areas into high- and background-risk groups. This should be kept in mind when interpreting the results.

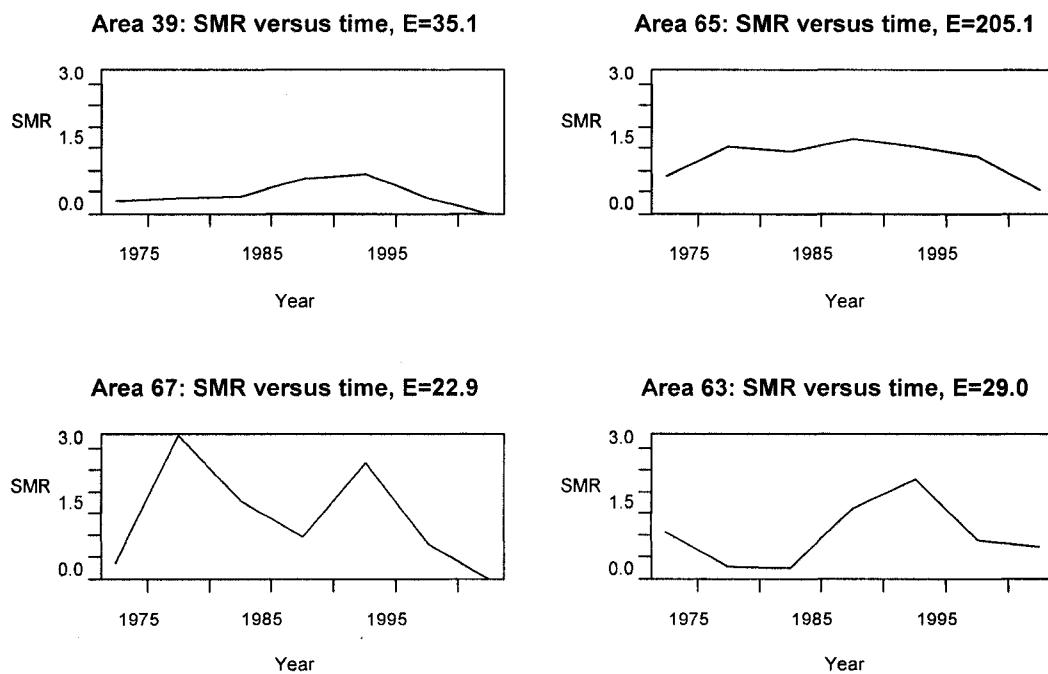
Table 5.1: Comparison of the distributions of expected counts from the Kentucky cervical cancer simulations and the Northern Sardinia Multiple Sclerosis data

	Minimum	1st Quartile	Median	3rd Quartile	Maximum
Kentucky Cervical Cancer	1.167	6.182	8.996	16.820	366.800
Northern Sardinia MS	1.006	1.819	3.219	7.154	205.100

For an exploratory assessment of the degree of spatial dependence and the patterns of relative risk through time, we plot the Standardized Morbidity Ratio (SMR) versus time for four areas with larger expected counts, two of which are neighbours (Figure 5.5). The graphs show that the temporal patterns vary from area to area. Differential missing ‘residence at onset’ information for the last time period may be responsible for the decrease in SMR in the later years. Ignoring the SMRs in the last time period (i.e. 2001 to 2005), area 67 appears to have overall

decreasing trend, while the trend seems to increase for areas 39 and 63. Modelling the relative risks with the TLT model may not be unreasonable for these data since the slopes can differ from area to area. However, decreasing trends may be exaggerated due to missing data in the last decade. The shapes of the graphs are different between areas (e.g. area 39 and area 67) so the TRW model may have more difficulties in modeling the relative risks through time since it models the effect of the time period for the entire study region, rather than for each area independently. Note also that areas 39 and 65 are neighbours with relatively large expected counts, yet their overall SMR levels are very different. Therefore, assuming spatial dependence, as in the BYM model, may not be appropriate for all areas of the map.

Figure 5.5: SMR through time for four areas of Northern Sardinia. E represents the expected count for the area.



We present, in Figures 5.6 and 5.7, disease maps displaying the area-level expected counts and the crude SMRs, respectively. For all disease maps, the numbers in parentheses that appear in the legend refer to the number of areas falling in each category.

Figure 5.6: Disease map of expected counts with areas that were detected as 'hot spots', by at least one of the four models, identified

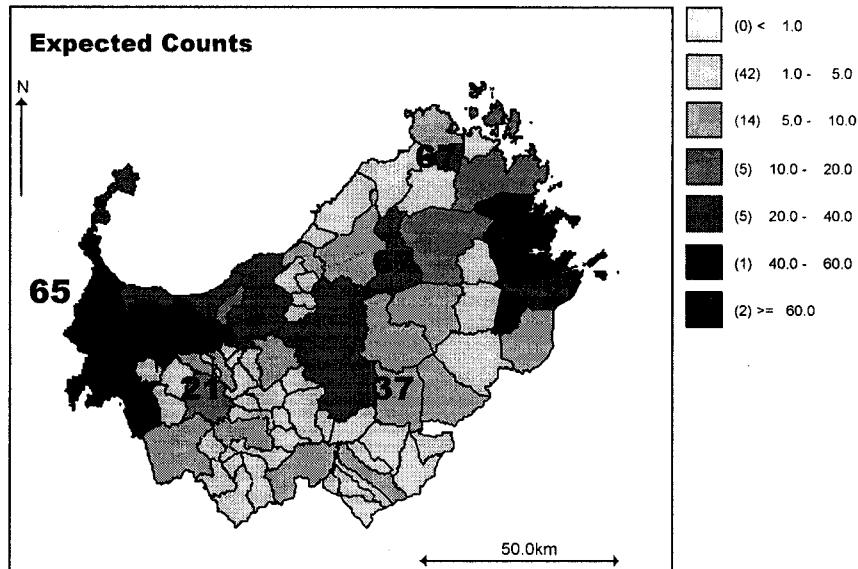


Figure 5.7: Disease map of crude SMR. The SMR is the observed number of cases divided by the expected count.

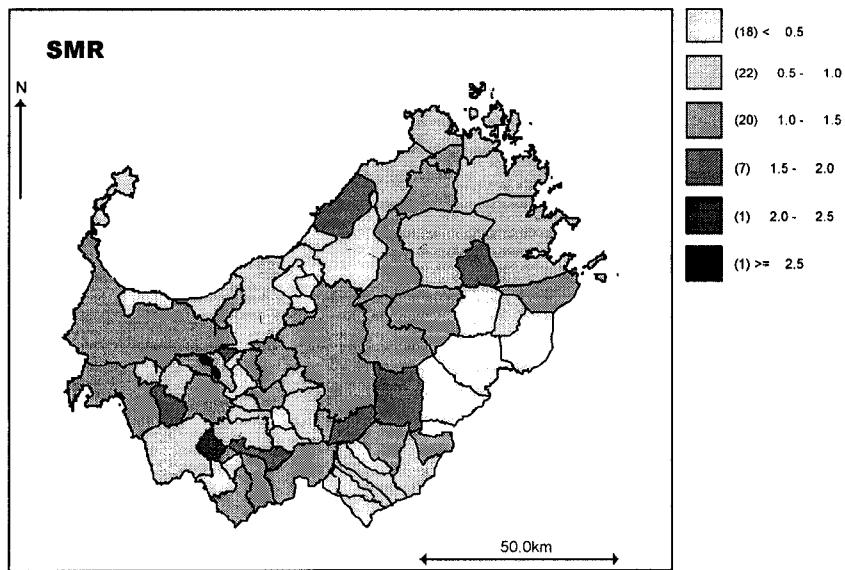


Table 5.2 gives the list of hot spots identified using each model. Areas with PP's between 0.65 and 0.75 are considered 'possible' hot spots, areas with PP's between 0.75 and 0.85 are 'probable' hot spots and areas with PP's over 0.85 are 'hot spots'. The estimated *posterior mean relative risks* and the *posterior proportions* for the proposed hot spots are presented in Tables 5.3 to 5.5. Figure 5.8 shows the disease map with all proposed hot spots indicated. Some of the areas that appeared to have high SMRs (i.e. over 1.5), for example, areas 50, 54 and 64, have only slightly higher risk, according to the posterior means and proportions and only the TLT model identified these areas as high-risk areas. Many of the same areas appear in lists of the four models, namely areas 21, 37, 48, 65 and 67. These are scattered throughout the map and the disease maps of the *posterior proportions* and *means* for all fours models (Figures 5.9 to 5.16) do not show a gradient in risk. The map appears to be dominated by unstructured heterogeneity. This is even true for the BYM model that includes a term for spatially structured variability.

Table 5.2: Areas identified as hot spots by each model.

	<b>BYM</b>	<b>LN</b>	<b>TLT</b>	<b>TRW</b>
<b>Possible Hot Spot</b> $(0.65 \leq PP < 0.7)$	48	37, 48	30, 33, 45, 50, 54, 64	48
<b>Probable Hot Spot</b> $(0.75 \leq PP < 0.8)$	21, 37, 67	21, 67	48	21, 37, 67
<b>Hot Spot</b> $(PP \geq 0.85)$	65	65	21, 37, 65, 67	65

Figure 5.8: Disease map showing all areas that are identified as hot spots ( $PP > 0.65$ ) by at least one model.

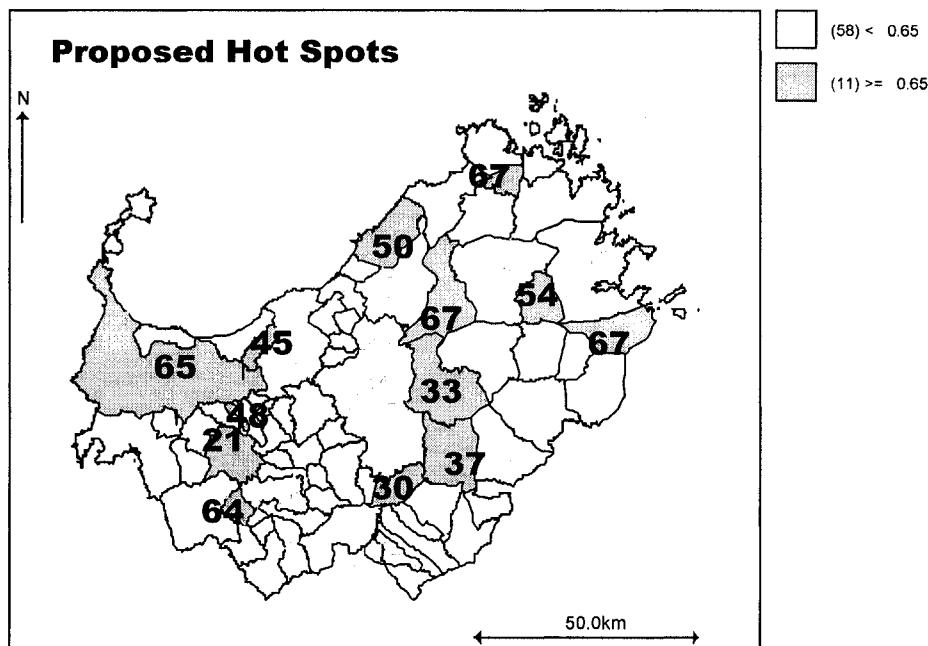


Table 5.3: Posterior mean relative risk (RR) and posterior probability (PP) of areas that were classified as 'possible hot spots' by at least one model. Areas that are considered 'possible' hot spots have PP's in the range 0.65 to 0.75.  $E_i$  is the expected count for area  $i$ .

AREA	$E_i$	BYM		Log-normal		Temporal Linear Trend		Temporal Random Walk	
		RR	PP	RR	PP	RR	PP	RR	PP
30	2.00	1.03	0.48	1.05	0.51	1.14	0.67	1.09	0.53
33	6.00	1.07	0.56	1.08	0.58	1.16	0.73	1.11	0.61
37	5.51	1.22	0.75	1.22	0.74	1.28	0.86	1.28	0.78
45	10.73	1.03	0.52	1.02	0.51	1.10	0.67	1.04	0.54
48	2.14	1.19	0.67	1.18	0.67	1.26	0.80	1.21	0.66
50	2.99	1.03	0.47	1.04	0.51	1.14	0.68	1.06	0.53
54	2.60	1.06	0.51	1.07	0.54	1.17	0.71	1.11	0.58
64	1.21	1.04	0.48	1.04	0.50	1.15	0.66	1.08	0.53

Table 5.4: Posterior mean relative risk (RR) and posterior probability (PP) of areas that were classified as ‘probable hot spots’ by at least one model. Areas that are considered ‘probable’ hot spots have PP’s in the range 0.75 to 0.85.  $E_i$  refers to the area expected count.

AREA	$E_i$	BYM		Log-normal		Temporal Linear Trend		Temporal Random Walk	
		RR	PP	RR	PP	RR	PP	RR	PP
21	13.66	1.17	0.76	1.17	0.77	1.22	0.86	1.2	0.79
37	5.51	1.22	0.75	1.22	0.74	1.28	0.86	1.28	0.78
48	2.14	1.19	0.67	1.18	0.67	1.26	0.80	1.21	0.66
67	22.9	1.16	0.79	1.16	0.82	1.21	0.89	1.19	0.82

Table 5.5: Posterior mean relative risk (RR) and posterior probability (PP) of areas that were classified as ‘hot spots’ by at least one model. Areas considered as ‘hot spots’ have PP’s between 0.85 and 1.0.  $E_i$  is the area-level expected count.

AREA	$E_i$	BYM		Log-normal		Temporal Linear Trend		Temporal Random Walk	
		RR	PP	RR	PP	RR	PP	RR	PP
21	13.66	1.17	0.76	1.17	0.77	1.22	0.86	1.2	0.79
37	5.51	1.22	0.75	1.22	0.74	1.28	0.86	1.28	0.78
65	205.1	1.25	1.00	1.25	1.00	1.26	1.00	1.25	1.00
67	22.9	1.16	0.79	1.16	0.82	1.21	0.89	1.19	0.82

Figure 5.9: Disease map of PP for BYM model.

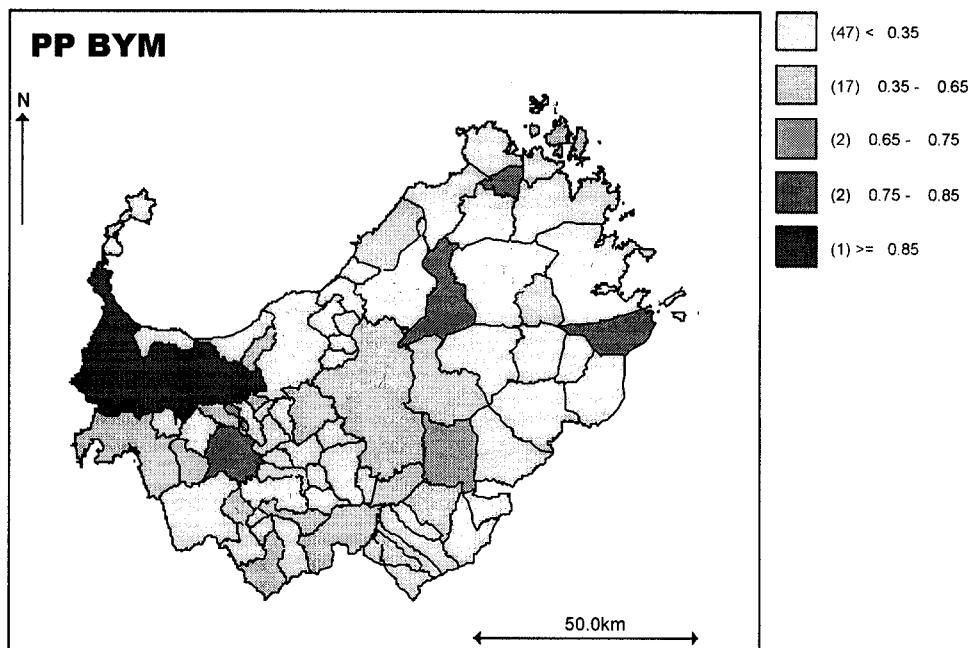


Figure 5.10: Disease map of PP for log-normal model with  $\text{Gamma}(0.001, 0.001)$  hyperprior for the precision.

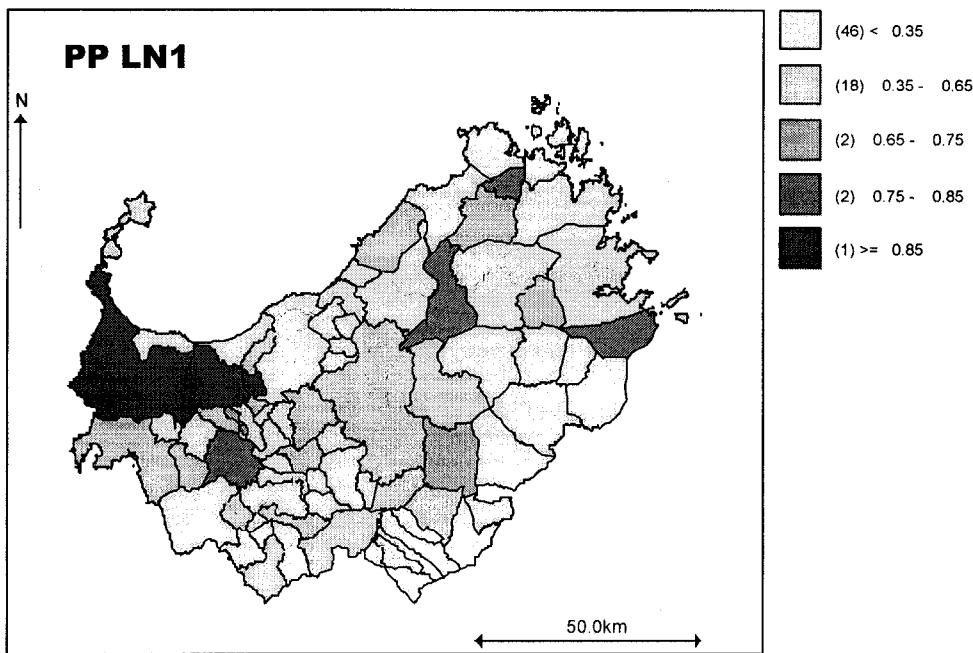


Figure 5.11: Disease map of PP for Temporal Linear Trend model.

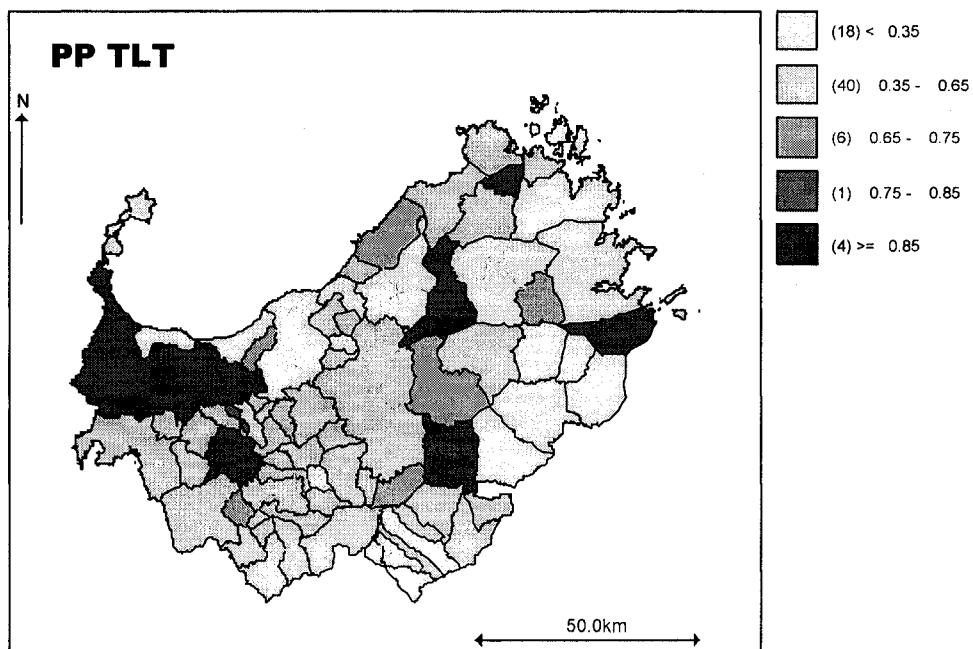


Figure 5.12: Disease map of PP for Temporal Random Walk model.

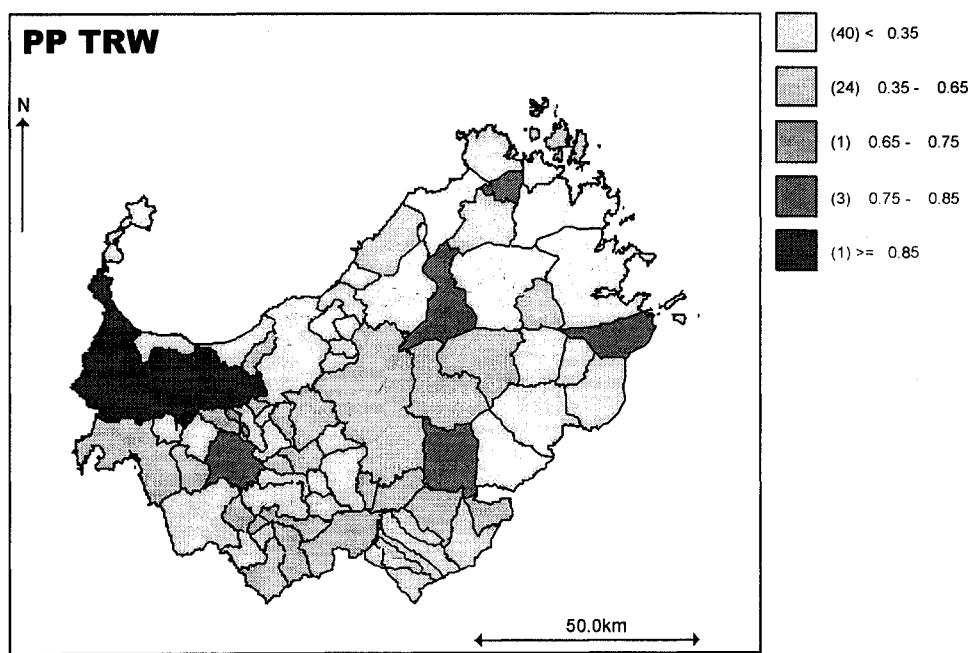


Figure 5.13: Disease map of *posterior mean relative risk* for BYM model.

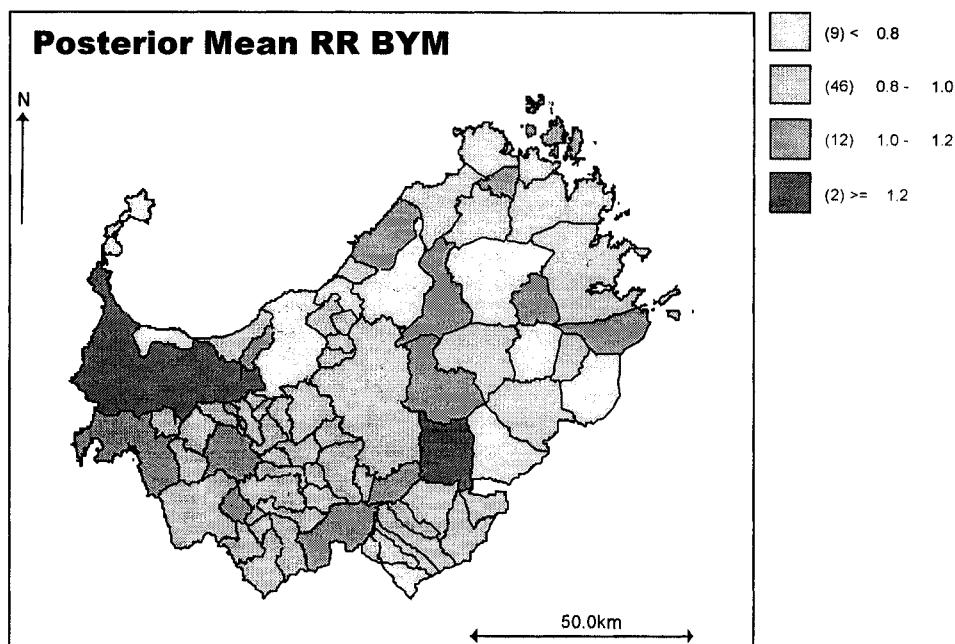


Figure 5.14: Disease map of *posterior mean relative risk* for LN1 model.

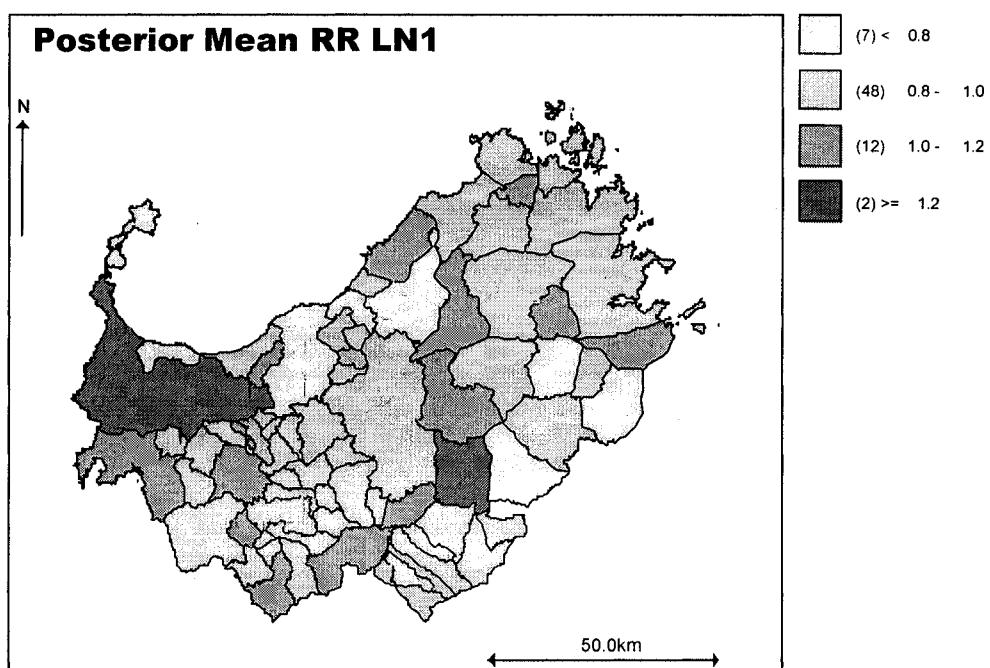


Figure 5.15: Disease map of *posterior mean relative risk* for TLT model.

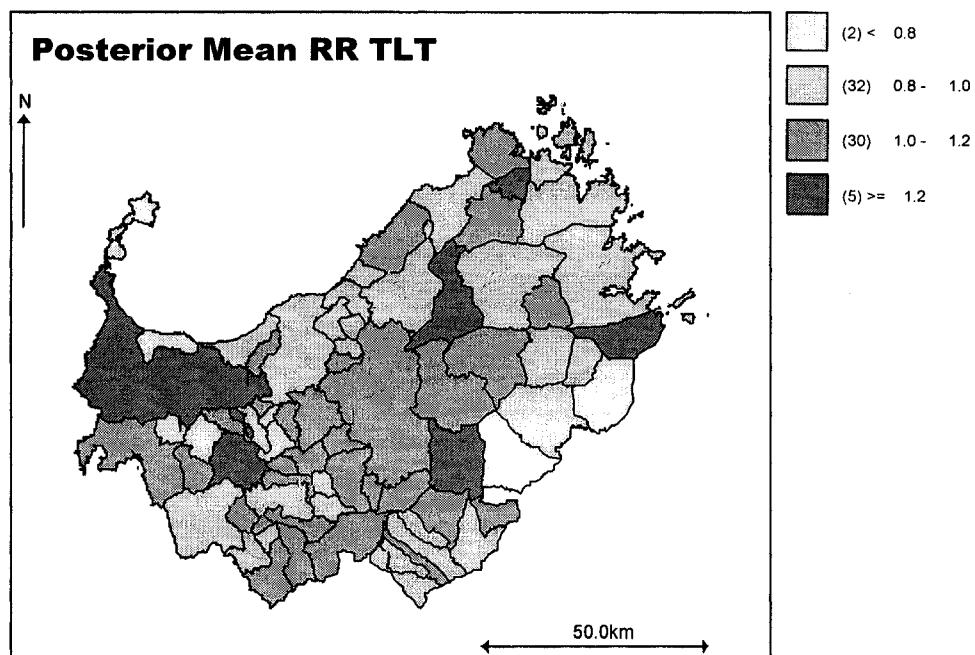
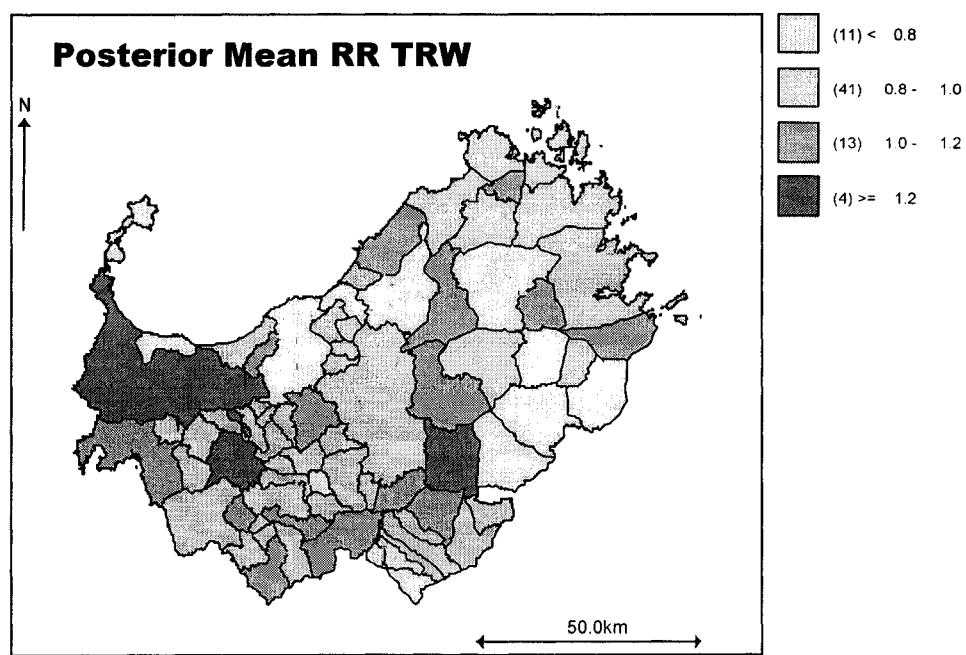
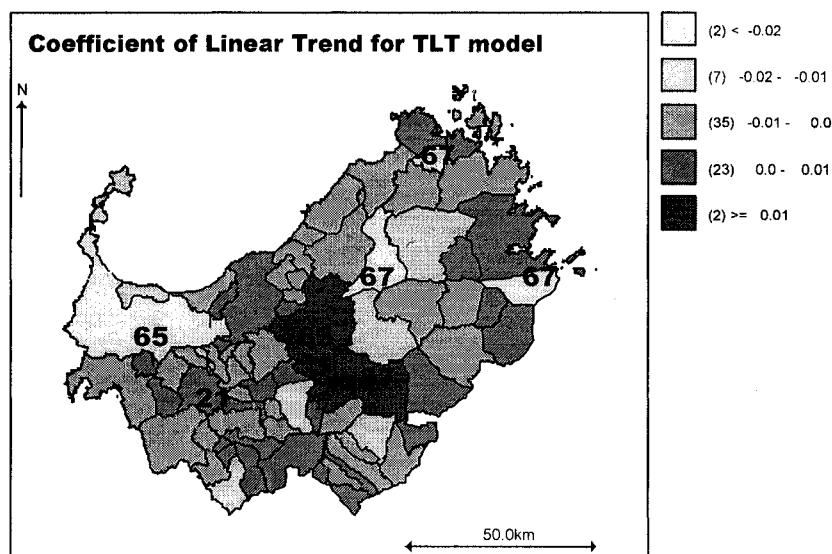


Figure 5.16: Disease map of *posterior mean relative risk* for TRW model.



We produced a disease map of the coefficient of linear trend from the TLT model (Figure 5.17) and found that two hot spots, areas 65 and 67, have the sharpest decrease in trend, while the sharpest increases are observed for areas 37 (a hot spot) and 63 (a background-risk area). However, it is possible that differential missingness may be partially responsible for the estimated rates of increase. Area 67 has coefficient equal to -0.026. This implies that the risk of MS in area 67 relative to the overall risk for Northern Sardinia decreases by a factor of  $e^{-0.026} = 0.97$  every five years. The coefficient of trend for area 63 is 0.012, which translates to an increase in relative risk of 1.2 % (i.e.  $e^{0.012} = 1.012$ ) every five years. Areas 65, 67 and 63 have expected counts over 20, thus the trend is unlikely to have arisen by chance. This is not the case for area 37 which has an expected count equal to 5.51.

Figure 5.17: Disease map of coefficient of linear trend from TLT model.



We assessed sensitivity to the choice of prior by comparing the disease maps of the posterior proportion with different hyperpriors for the precisions. For the BYM model the hyperprior for the precision of the spatially unstructured effect was changed to a  $\text{Gamma}(0.001, 0.001)$  distribution but comparing the PP maps for the two priors, we observed little change in the PP's (Figures 5.9 and 5.18) and no change in the list of

proposed hot spots. The maps of the log-normal model with  $\text{Gamma}(0.001, 0.001)$  hyperprior for the precision were compared to the maps for the log-normal model with  $\text{Gamma}(0.5, 0.0005)$  hyperprior (Figures 5.10 and 5.19). There are only small differences in the PP's and, again, no change in the list of proposed hot spots. The results of the TLT model with  $\text{Gamma}(0.001, 0.001)$  hyperpriors for the precisions were compared to the model with  $\text{Gamma}(0.01, 0.001)$  hyperpriors (Figures 5.11 and 5.20). There was little change in the PP's, but area 45, previously classified as a 'possible' hot spot, became a background-risk area. The PP for area 45 was 0.67 for the first set of priors and for the second set the PP was 0.64. The TRW precision for the random walk was changed to a  $\text{Gamma}(0.01, 0.001)$  and for the spatially unstructured effect we tried a  $\text{Gamma}(0.001, 0.001)$  prior (Figures 5.12 and 5.21). Again, there was no difference in the list of proposed hot spots. Comparing the PP maps (Figures 5.9 to 5.12 and 5.18 to 5.21) we find that the posterior distributions for this dataset are only slightly affected by changes in the prior for a given model.

Figure 5.18: Disease map of PP for BYM model with  $\text{Gamma}(0.001, 0.001)$  hyperprior for the precision of the unstructured effect.

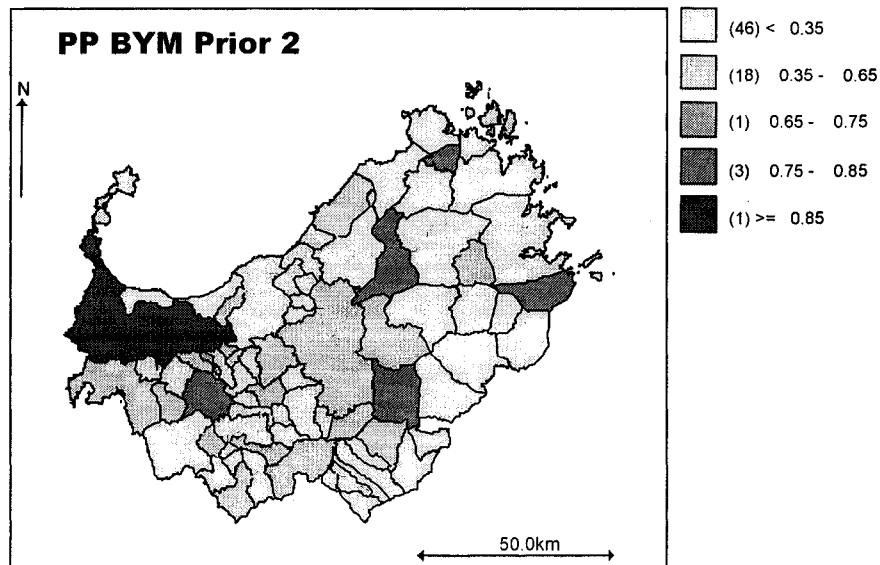


Figure 5.19: Disease map of PP for log-normal model with  $\text{Gamma}(0.5, 0.0005)$  hyperprior for the precision of the unstructured effect.

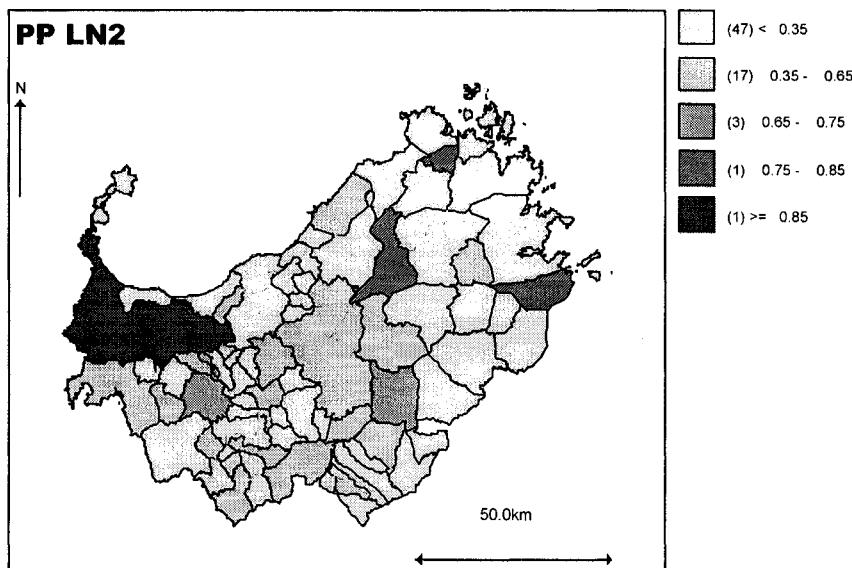


Figure 5.20: Disease map of PP for TLT model with  $\text{Gamma}(0.1, 0.0001)$  hyperprior for the precision of the unstructured effect.

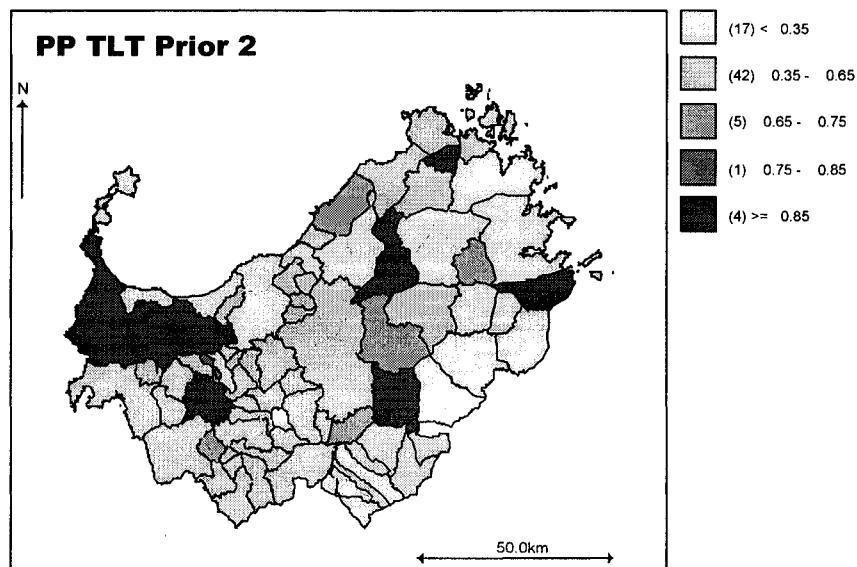
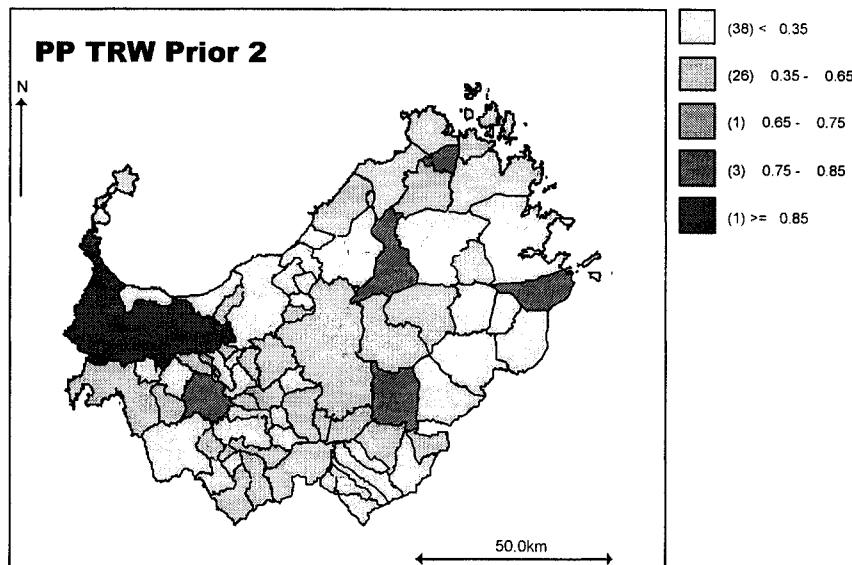


Figure 5.21: Disease map of PP for TRW model with  $\text{Gamma}(0.001, 0.001)$  hyperprior for the precision of the random walk and  $\text{Gamma}(0.001, 0.001)$  hyperprior for the precision of the unstructured effect.



For the spatial models, the DIC's are presented in Table 5.6 and they suggest that the log-normal model with  $\text{Gamma}(0.001, 0.001)$  hyperprior is superior to the other spatial models in terms of adequacy of model fit given model complexity. For the spatio-temporal models, the TRW model with the second set of priors ('Prior 2') has lower DIC, indicating better model fit when taking into account the complexity of the models (Table 5.7).

Table 5.6: Deviance Information Criterion (DIC) and effective number of parameters,  $p_D$ , for spatial smoothing models

Model	Prior for unstructured effect $V_i$	DIC	$p_D$
BYM Prior 1	$\text{Gamma}(0.5, 0.0005)$	321.7	45.4
BYM Prior 2	$\text{Gamma}(0.001, 0.001)$	319.6	44.9
Log-normal Prior 1	$\text{Gamma}(0.001, 0.001)$	314.5	39.0
Log-normal Prior 2	$\text{Gamma}(0.5, 0.0005)$	322.7	45.2

Table 5.7: Deviance Information Criterion (DIC) and effective number of parameters,  $p_D$ , for spatio-temporal models

Model	Parameter	Prior	DIC	$p_D$
TLT Prior 1	Intercept	$Gamma(0.001, 0.001)$	1066.1	47.2
	Precision			
	Linear trend	$Gamma(0.001, 0.001)$	1068.3	48.9
	Precision			
TLT Prior 2	Intercept	$Gamma(0.1, 0.0001)$	1068.3	48.9
	Precision			
	Linear trend	$Gamma(0.001, 0.001)$	1001.8	72.0
	Precision			
TRW Prior 1	Unstructured	$Gamma(0.5, 0.0005)$	1001.8	72.0
	Precision			
	Random Walk	$Gamma(0.01, 0.01)$	998.2	69.4
	Precision			
TRW Prior 2	Unstructured	$Gamma(0.01, 0.001)$	998.2	69.4
	Precision			
	Random Walk	$Gamma(0.001, 0.001)$	998.2	69.4
	Precision			

To summarize the results, all four models identified five areas as hot spots. Only two of the areas were contiguous and there did not appear to be a gradient in any of the smoothed disease maps. The TLT model identified six additional areas as hot spots but all six had posterior proportions under 0.73. Results were relatively insensitive to the choice of prior. However, the DIC indicated that the models are not equivalent in terms of model fit (given the complexity of the models). Of the spatial models, the log-normal model with  $Gamma(0.001, 0.001)$  prior fit the data better according to the DIC criterion. The TRW model with

$\text{Gamma}(0.01, 0.001)$  prior for the precision of the spatially unstructured term and  $\text{Gamma}(0.001, 0.001)$  for the random walk was superior to the other three models. It is not clear, however, how this impacts on the accuracy in the classification of areas into high- and background-risk groups.

#### 5.4 Discussion

In recent years, disease mapping studies of commune-level MS prevalence have been conducted for both Northern Sardinia and Central Sardinia (Pugliatti et al, 2002; Montomoli et al, 2002). The prevalence map of Central Sardinia exhibited somewhat higher risk for two areas located in the centre of Central Sardinia. The map of Northern Sardinia showed higher prevalence in the western commune of Sassari and many of its neighbouring areas. Prevalence maps were also constructed according to residence between the ages of 5 to 15 years (the hypothesized critical age period of susceptibility to the environmental exposure). These revealed further clustering in the south-western area of the province. Although Sardinia is said to have little internal migration, prevalence maps are not ideal in generating hypotheses of disease etiology because, for some diseases, cases tend to move to areas with better health care facilities.

In this research, we construct maps using *incidence* data, where the residence of the incident case is defined as the commune of residence at the time of onset. Mapping the RR (or SMR) of MS for each commune (area) of Northern Sardinia can reveal higher and lower risk areas, relative to the overall risk of MS in Northern Sardinia. This is important for MS research because, although MS is thought to be caused by an environmental exposure in the genetically susceptible individual, it is further hypothesized that there may be a triggering ‘event’ (e.g.

environmental exposure) that initiates the disease process (Granieri, 1997). Mapping the relative risk using incident cases defined in this way may allow us to generate hypotheses of environmental triggers.

Several areas were identified as hot spots. The communes of Ittiri, Tissi and Sassari combined with Stintino in the west, Pattada in the central region and the combined communes of Tempio Pausania and Loiri Porto San Paolo in the east of the province (comprising three non-contiguous areas) all displayed higher relative risks. According to the TLT model, the communes of Sassari/Stintino and Tempio Pausania/Loiri Porto San Paolo showed the greatest decline in relative risk from 1971 to 2005, while another proposed hot spot, Pattada, had the second highest rate of increase through time (second to a combined group of communes in the central region).

The small-area prevalence cluster detection study by Pugliatti et al (2002) found a west to east gradient in risk. We did not find this using incidence data. As mentioned earlier, prevalence studies are not ideal for this kind of investigation because cases have a greater tendency to relocate to areas to have better access to health care. We found that 28 of the 47 cases that relocated after diagnosis moved to the commune of Sassari (in the north-west) or one of its neighbouring areas. This may partly explain the west-east gradient found in the Pugliatti et al (2002) study. It is also possible that the prior for the spatially structured random effects resulted in greater spatial dependence between areas, so that the commune of Sassari, a high-risk area with a large expected counts had a great influence on the neighbouring areas in the same way that an artefactual gradient appeared in the Kentucky generated maps of cervical cancer that were smoothed using the BYM, KHB and BC models.

At the present time the dataset is being updated, particularly with respect to the migratory changes prior to onset for more recent cases, so that more complete information should be available for future analyses. The effect of missing commune at onset for more recent cases could be seen in the disease map of the coefficient of linear trend where a majority of the areas showed a decline in relative risk. Pugliatti et al (2005) showed that, from 1965 to 1999, the incident rate of MS has been steadily increasing in all six linguistic areas of Northern Sardinia (though one area had a slight decrease at the end of the study period). Thus, we should be seeing more areas with increasing relative risks in the later time periods. It is also likely that the missing data compromised the relative risk estimation for the TLT model that assumes linear trend. For several areas, the SMR appeared to be smallest at the beginning and end of the study period (Figure 5.5), thus assuming a linear trend for these areas is not appropriate.

It is common to consider the DIC in selecting Bayesian models. When there is more ‘borrowing of strength’ due to the choice of prior, the estimate of the effective number of parameters,  $p_D$ , is small relative to the number of data points and so the DIC tends to be smaller as long as there is adequate fit (as measured by the mean deviance). For spatial models, the DIC may not always correlate with model accuracy. Best et al (2005) in their simulation study that compared several Bayesian spatial disease mapping models found that the model that had lowest DIC in 3 of 5 simulations was less accurate in classifying areas into high-risk and background-risk groups than some of the other models that had higher DIC. Spatial models that oversmooth the risk surface may mask hot spots, yet the effective number of parameters,  $p_D$ , will decrease, making the DIC appear to improve. The posterior mean deviance should be larger in this case but it may not compensate for the smaller value of the effective number of parameters. For the spatio-temporal models we deliberately smooth out irregularities so that there may be substantial

differences in the fitted and observed values at some time points, but this facilitates the detection of high-risk areas in some contexts. A spatio-temporal model that smoothes the relative risks through time to a greater extent may have smaller effective number of parameters but the deviance may be larger. Thus, a lower DIC may not imply better discrimination between high- and background-risk areas, so in this case the DIC may not be the best criteria for model selection.

## CHAPTER 6 DISCUSSION

Disease mapping studies, based on areal data, have become increasingly popular in recent years as a statistical tool for stabilizing disease maps. Technological advances have made areal data more readily available and have facilitated the implementation of computationally intensive Bayesian disease mapping models. The most widely used spatial Bayesian disease mapping model is the Besag, York and Mollie (1991) (BYM) model. Several studies have compared the BYM model to other disease mapping models with respect to performance in meeting various disease mapping objectives, such as goodness-of-fit to the data and accuracy of the classification of areas into high-risk and background-risk groups (e.g. Bernardinelli, 1995b; Lawson et al, 2000; Richardson et al, 2004; Best et al, 2005; Aamodt, 2006). In general, the findings suggested that the BYM model performed at least as well as the other spatial models with respect to many criteria. Nevertheless, the overall accuracy of classification and goodness-of-fit of the spatial models was not very good unless there was a lot of information in the data (e.g. large expected counts). Furthermore, the studies that investigated the accuracy of classification suffered from limitations that may have affected their conclusions regarding the accuracy of the BYM model. This thesis concerns the assessment of accuracy of classification of several Bayesian disease mapping models that differ with respect to the combination of spatially structured effects, spatially unstructured effects, temporal terms, and space-time interaction terms. As the BYM model is the most popular model and has performed relatively well in other studies, the accuracy measurements of the BYM model serve as a benchmark for the other models that we investigate.

## 6.1 Assessing the accuracy of the BYM model

To our knowledge, only two studies, Richardson et al (2004) and Aamodt et al (2006) were specifically designed with the goal of assessment of accuracy. Three others (Bernardinelli et al, 1995b; Jarup et al, 2002; Best et al, 2005) included such an assessment as part of their analyses. All five studies had limited assessments of accuracy and the three studies for which the assessment of accuracy was not a main objective, had too few replications (e.g. 5 or 6) for us to have confidence that the results were not due to the set of generated data. The largest studies, with the greatest number of replications and types of spatial clustering, had accuracy assessments that were based on sensitivity and specificity for a single cut-off (i.e. the threshold for the *posterior proportion* that separates a 'high-risk' from 'background-risk' classification). As the sensitivity and specificity depend on the choice of cut-off, by restricting the assessment to a single cut-off, they had measured the accuracy of the model *at that value of the cut-off* rather than the accuracy of the model alone. In practice, there is no universally accepted cut-off and we found that the 'optimal' cut-off that maximizes sensitivity and specificity depends on the (unknown) true hot spot RR's as well as the distribution of expected counts. In this thesis we assess accuracy independently of the choice of cut-off through ROC curves and their corresponding AUCs. In addition, we find the 'optimal' cut-off that jointly maximizes sensitivity and specificity and we use these values of the sensitivity and specificity to find positive and negative predictive values. The positive and negative predictive values measure the probability that a 'detected' hot spot corresponds to a true hot spot and the probability that an area classified as 'background' corresponds to a true background-risk area, respectively. The predictive values depend on the sensitivity and specificity of the classification procedure as well as the prevalence of 'positives' in the population. When a classification procedure has low positive predictive value it means that, in practice,

when the model identifies a number of areas as high-risk, only a small proportion of these will be true high-risk areas. Mathematically, the positive predictive value can be expressed in terms of the sensitivity, specificity and the prevalence of ‘positives’. Although, sensitivity and specificity have some effect on the positive predictive value, small prevalence values can decrease the positive predictive value beyond what you would expect for a given sensitivity and specificity. Another objective of this thesis is to determine the prevalence requirement that is needed to have a positive predictive value over 0.50, so that a detected hot spot has a better than 50% chance of being a true hot spot. Thus, unlike the previous five studies, we conduct a simulation study with many replications to measure the accuracy of classification, independently of the choice of cut-off, and we determine the extent to which we can have confidence that a detected hot spot corresponds to a true hot spot.

Our results suggest that the BYM model has fairly low accuracy in detecting areas with elevated risk, as measured by AUCs, except when the hot spot relative risk is large (at least 2) and the hot spot expected counts are over 3 or 4. We found that the ‘optimal’ cut-off varied according to the relative risk in high-risk areas and the size of the expected counts. When the hot spot relative risk was small (i.e. 1.5), the ‘optimal’ cut-off was between 0.6 and 0.7, but simulations with larger hot spot relative risks resulted in ‘optimal’ cut-offs as high as 0.92. Four of the five simulation studies that assessed accuracy of classification used a single cut-off for all their simulations. One study used a cut-off of 0.8 and concluded that the BYM model’s classification is highly specific with low sensitivity (Richardson et al, 2004). To appreciate the effect of the choice of cut-off on the sensitivity and specificity, we found sensitivities and specificities using a cut-off of 0.8 and compared them to the sensitivities and specificities found from applying the ‘optimal’ cut-offs. The ‘optimal’ cut-off resulted in much higher sensitivities for simulations with smaller hot

spot relative risks and expected counts. In practice, we need to consider the disease, the populations under study and the purpose of the disease mapping study when deciding which cut-off to use. For example, it may be important to identify all hot spots even at the cost of a higher false-positive rate. In this case a smaller cut-off may be more appropriate. On the other hand, when false-alarm are costly, a larger cut-off is more suitable. Larger expected counts and an *a priori* belief in larger hot spot relative risks suggest a higher cut-off, if the cost of false-positives and false-negatives is the same. If we do not know *a priori* what the hot spot relative risks should be, then we can consider categorizing areas into several groups according to their *posterior proportion (PP)*. For example, we may consider areas with PP's between 0.35 to 0.65 as background-risk areas, PP's between 0.65 to 0.75 as 'possible' hot spots, PP's between 0.75 to 0.85 as 'probable' hot spots and PP's over 0.85 as 'hot spots'. This way, we may still catch areas with truly high relative risks even though the true relative risk is not as high as 3 or the expected counts are not large.

The maximum sensitivities and specificities (i.e. from the 'optimal' cut-off) were used to find the predictive values. For all simulations, the BYM model had very low positive predictive value ( $PPV$ ) and high negative predictive value ( $NPV$ ) (i.e. the  $PPV$  never exceeded 62%) because of the low prevalence of hot spots on the map. We considered the prevalence that would give a  $PPV$  over 50% for each relative-risk and scale factor combination and found that when the true hot spot relative risk is 1.5, we would need for 35 to 50% of the areas on the map to be high-risk areas. When the true hot spot relative risk is 2, we need a prevalence proportion between 18 and 42%. This decreases to a range of 2 to 38% when the true hot spot relative risk is 3. The preceding prevalences would give a greater than 50% chance that a detected hot spot corresponds to a true hot spot. However, these prevalence proportions are unrealistic.

Thus, in practice, when the model flags an area as 'high-risk', there is only a small chance that this area is a true high-risk area.

## **6.2 Assessing the accuracy of alternative spatial and spatio-temporal models**

Given that the prevalence of true hot spots is fixed, we can only hope to increase the positive predictive value by increasing the sensitivity and specificity. Thus it is worthwhile to seek out alternatives to the BYM model for identifying high-risk areas. One alternative is to use the popular spatial scan statistic, SaTScan (Kulldorff M, 1997). However, Aamodt et al (2006) did not find an important difference in the sensitivity and specificity of the spatial scan statistic as compared to the BYM model when the relative risk in high-risk areas exceeded 1.5 (Aamodt G et al, 2006). Sensitivity and specificity assessments for the BYM model were misleading when the relative risk was 1.5 or less. For these simulations the specificities were often found to be 1.0 and the sensitivities were close to zero. Thus the threshold was probably too high, making it appear that the BYM model is not adequate when the hot spot RR is 1.5 or less. Thus, it is reasonable to consider alternatives to these models. In chapter 4, we considered 6 alternative models, 2 spatial models and 4 spatio-temporal models. The two spatial models are less complex than the BYM model in that they only have the spatially unstructured effects and these spatial models were not considered in the other five simulation studies that assessed accuracy of classification into high-risk and background-risk groups. For the spatio-temporal models we used a novel approach to combine each area's relative risks through time to form a measure of the area's relative risk for the study period. Summarizing through time was necessary to facilitate a comparison with the usual (spatial) approaches which examine only the cumulative disease counts for the study period under consideration. ROC

curves, AUCs and confidence intervals of the difference in model AUCs were used to assess accuracy.

We found that as long as we accounted for unstructured variability, the model's accuracy was not worse than the BYM model. In fact, the simpler log-normal model with  $\text{Gamma}(0.001, 0.001)$  was superior to the BYM model in many simulations. For the log-normal model the choice of hyperprior made a difference in accuracy. Although the  $\text{Gamma}(0.5, 0.0005)$  prior for the precisions is recommended in references and software manuals (Thomas et al, 2004; Wakefield et al, 2000a) we found that for this particular set of generated data and clustering pattern, the  $\text{Gamma}(0.001, 0.001)$  resulted in greater discriminatory power.

The results of the assessment of the spatio-temporal models indicated that for all simulated temporal patterns the linear trend model with spatially unstructured effects for the intercept and coefficient of the linear trend (i.e. the Bernardinelli et al (1995) space-time interaction model with spatially unstructured effects) had the highest accuracy while the same model with only spatially structured effects was the least accurate model. When we applied the spatio-temporal models to the generated time series of counts with one unusual count that did not fit the temporal pattern, overall, the spatio-temporal models were much more accurate than the spatial models. Thus, we found that, as long as the model accounted for unstructured heterogeneity, the spatio-temporal models were as accurate as or more accurate than the corresponding (spatial) model without the temporal term.

### **6.3 Mapping the incidence of MS in Northern Sardinia**

In chapter 5, we applied two spatial and two spatio-temporal models to the problem of mapping the relative risk of multiple sclerosis in the province of Sassari in Sardinia, Italy. A previous study of the spatial distribution of point prevalence of MS cases and prevalence according to residence from ages 5-15 years indicated higher prevalence in some areas in the west of the province. In this research we used incidence data, where the time at onset was defined as the time at which the first symptoms appeared and the residence at onset was defined as the administrative commune at time of onset. We mapped the relative risk of multiple sclerosis and all four models identified the same five areas as having high-risk. Three areas are located in the west of the province (i.e. the communes of Sassari/Stintino, Ittiri, and Tissi), one area is in the centre (i.e. Pattada) and one area (comprised of three non-contiguous areas, Tempio Pausania and Loiri Porto San Paolo) is located in the east of the province. The high-risk areas in the west were also found to have higher prevalence in the prevalence study (Pugliatti et al, 2002). However, the higher prevalence in the west may have been partly due to the fact that, after diagnosis, cases migrated to the large cities in the west to have easier access to the health care facilities there. The remaining two high-risk areas in the centre and east of the province may provide additional clues to the etiology of MS. The linear trend model with (spatially unstructured intercept and coefficient of linear trend) provided a means to map the coefficient of linear trend to detect areas with increasing or decreasing trend. We found that two of the hot spots, the communes of Sassari/Stintino and Tempio Pausania/Loiri Porto San Paolo, displayed the greatest decline in relative risk through time, whereas, the areas of Pattada (a hot spot) and the area including Ozieri, Erula, Tula, Chiaramonti, Perfugas in the centre of the province had the sharpest increase in relative risk during the 35 year study period (i.e. 1971 to 2005).

## **6.4 Strengths and limitations**

### **6.4.1 Strengths and limitations of the simulation study**

The simulation study has many positive features. First, a large number of replications were used to minimize the effect of random variation on the results of accuracy. Secondly, a realistic range of expected counts, hot spot relative risks and temporal patterns were used to generate the data. Another strength of the simulation study is the more complete assessment of accuracy of the BYM model that was achieved through ROC curves, AUCs, sensitivity, specificity and positive and negative predictive values. Finally, we considered alternatives to the BYM model that are not much more difficult to implement in WinBUGS 1.4 (Spiegelhalter et al, 2002). This strength is important because the common use of the BYM model may be due to the ease of its implementation in the more recent versions of WinBUGS (Spiegelhalter et al, 2002).

The simulation study was fairly large in terms of the number of models that were assessed and the number of replications. However, only one clustering pattern was considered, that of isolated hot spots consisting of a single area. It is possible that the spatio-temporal models may not perform as well, relative to the BYM model, when hot spots are clusters of high-risk areas as when they consist of a single area. Furthermore, we selected areas as hot spots according to the size of their expected count. This variety in size of expected count enabled us to examine the accuracy of classification under varying degrees of instability (i.e. the smaller the expected count, the more unstable the estimate of relative risk). However, we only used this particular set of hot spots for all simulations and it is possible that the results of accuracy are sensitive to the positions of the hot spots on our map. In the same vein, we could have considered a multitude of other temporal patterns for the spatio-temporal analyses.

The conclusions from the simulation study are restricted to the identification of isolated hot spots consisting of a single area and for the four temporal patterns we considered. However, in practice, it is difficult to determine whether these assumptions regarding spatial clustering and temporal patterns will hold because of the random variation that dominates the crude SMRs. Even the assumption of spatial dependence is difficult to verify. There may be parts of the map that are sparsely populated (e.g. rural areas) and the areas in that part of the map may display large relative risks but it is difficult to determine whether the extreme estimates are due to a common risk factor that increases the risk in these areas or if they are due to random variation causing extreme estimates to occur because they are based on small population sizes (Pascutto et al, 2000). In chapter 5, there were contiguous areas of the map with relatively large expected counts that had very different crude SMRs. This would appear to be evidence against the presence of spatial dependence and the clustering of high-risk areas, but this is unlikely to be sufficient evidence. It is difficult to judge *a priori* if high-risk areas will cluster or whether they will be isolated. Bernardinelli et al (1995a) suggest using spatially unstructured effects for their space-time interaction model when there is an *a priori* belief that hot spots are no bigger than the area size, otherwise, when we expect high-risk areas to cluster, we should use spatially structured effects. Bernardinelli et al (1995a) chose spatially unstructured effects when the map was more coarsely partitioned and spatially structured effects for the high-resolution map. But, again, there are no established criteria that can be used to differentiate maps with a coarse partition from maps with a finer partition. In general, it is difficult to know at the start of a disease mapping study, whether the assumptions of our simulation study will apply.

## **6.4.2 Strengths and limitations of the spatial analysis of MS incidence**

The Multiple Sclerosis data set from Northern Sardinia (the province of Sassari) had many qualities that make it ideal for a spatial analysis of MS risk. There is only a small amount of migration from outside of the province and little internal migration as well (Pugliatti et al, 1997). This feature makes it easier to define the area in which the exposure to the environmental agent took place. In addition, data have been collected since 1965, thus a temporal analysis is feasible when we model the number of cases occurring in each 5-yr period.

However, there were some limitations of the data. First, census data were used to estimate the person-time at risk during each five year period. Though there are formulas to compute intercensal estimates and postcensal projections of population size, they require additional information such as births and deaths during the period of interest and military movement in and out of the areas of the study region (Arnold et al, 2000). For this reason, projection of population sizes was not attempted. Another limitation of the data was the missingness in the case records, particularly for more recently diagnosed cases. Some case records had missing 'time at onset'. These cases were immediately deleted. The commune at onset had to be deciphered from the migration information. When the migration information was missing and when the commune at registration coincided with the commune at birth and commune of residence during the ages 5 to 15 years, the commune at onset was taken to be the same. When the residence information was missing, the case was deleted. Missing data in the last time periods would certainly have affected the estimated coefficient of linear trend for the TLT model and may have caused bias if some communes/areas had more complete

migration information recorded. Therefore, when the records have been updated, if the missing data is negligible, we hope to re-analyse the data.

## 6.5 Future Research

Future work should consider the accuracy of these models given various forms of spatial clustering and temporal patterns to determine if the accuracy is very sensitive to the assumptions. For example, accuracy in detecting linear and circular clusters of areas should be investigated for all models. Linear clusters could occur along rivers, for instance. Aamodt et al (2006) considered various spatial clustering patterns in their simulation study and found that the BYM model had difficulty in identifying linear clusters. In addition, we should consider the accuracy of these models in the presence of true low-risk areas. We assumed, in our simulation study, that all areas that were not hot spots had relative risks of one, but it is not clear that the results would be the same if some of the areas had relative risks much smaller than one.

In the simulation study, we did not examine the choice of prior for the precisions because this had been considered extensively in other studies (Bernardinelli et al, 1995b; Kelsall and Wakefield, 1999). We used the recommended hyperprior for the precisions of spatially structured and unstructured effects except for one log-normal model and the linear trend model with spatially unstructured effects, where we used a  $\text{Gamma}(0.001, 0.001)$  hyperprior for the precisions. The clear superiority of the  $\text{Gamma}(0.001, 0.001)$  hyperprior was surprising. Future work should investigate the accuracy for different hyperpriors for the precisions under different clustering patterns to determine whether this is specific to the maps with small, isolated hot spots or whether this extends to other clustering patterns.

In chapter 5, we used the DIC to aid in determining the ‘best’ model. However, it is not clear when (or even whether) a better (i.e. lower) DIC indicates better classification of models into high-risk and background-risk groups. The simulation study by Best et al (2005) showed that better fit according to the DIC did not necessarily correlate with better discrimination between high-risk and background-risk areas. The relationship between the DIC and accuracy of classification should be further investigated as it could be useful to establish a set of criteria for choosing the model that is best in classifying areas as hot spots. The choice of the ‘best’ model should always first be based on assumptions about the spatial and temporal pattern of risk which should, in turn, be based on the disease, the study region and its inhabitants. However, we should continue to investigate whether statistical criteria, such as the DIC, could be useful as well.

Another topic for future research is to investigate any changes in accuracy from employing other models for the area-level counts of cases, such as the Negative Binomial and *zero-inflated Poisson model* (Angers and Biswas, 2003), rather than a Poisson model at the first stage of the hierarchical model. Although it is used extensively, the Poisson model is generally not appropriate when there are too many zero counts (i.e. no cases occurring in many areas) (Congdon, 2001). An alternative to the Poisson distribution is the *zero-inflated Poisson model* that assumes the population can be divided into two subpopulations, one with zero mean and the other with a mean that depends on the area-specific covariates. Unmeasured covariates are represented by the spatially unstructured effects, spatially structured effects and temporal terms. These models might be more appropriate than the Poisson model for rare diseases.

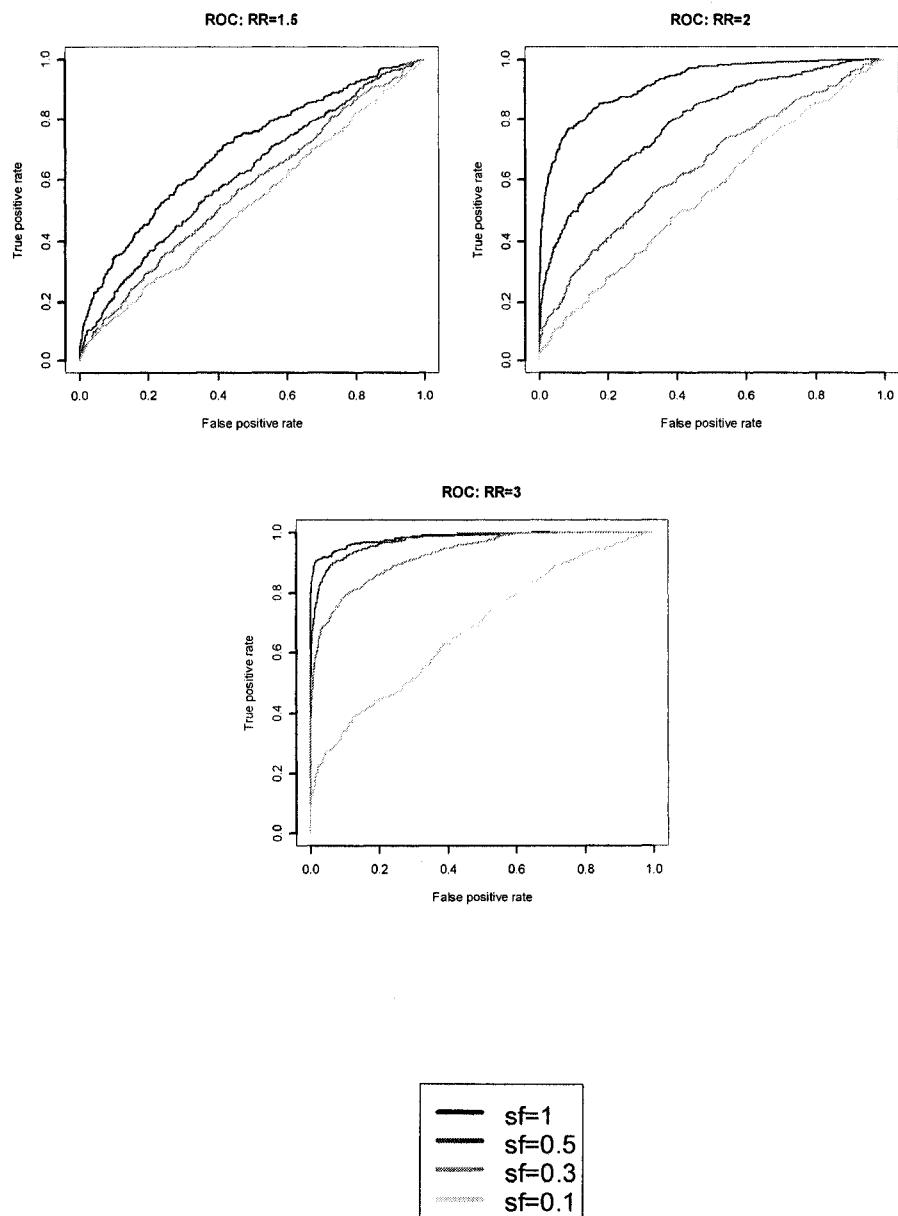
As with many simulation studies the intention was to examine the performance of a manageable number of models given a manageable set of generated data. These models are computationally costly in terms of time, in RAM memory, and in storage space. Although we could have investigated the performance of many additional models' performance, under many different spatial and temporal scenarios, we feel that these findings are encouraging, in that they support the exploration of alternative models to the Besag, York and Mollie(1991), that at least in some scenarios have higher accuracy in classifying areas into high-risk or background-risk groups.



## APPENDIX A

- A1. Receiver Operating Characteristic curves for the BYM model**
- A2. Disease maps of the posterior proportion and the posterior mean for BYM model**
- A3. Disease maps for spatial and spatio-temporal models, Scenario 1**
- A4. Confidence limits for differences in AUCs**
- A5. Receiver Operating Characteristic curves for spatial and spatio-temporal models**
- A6. Maximum  $\hat{R}$  values**

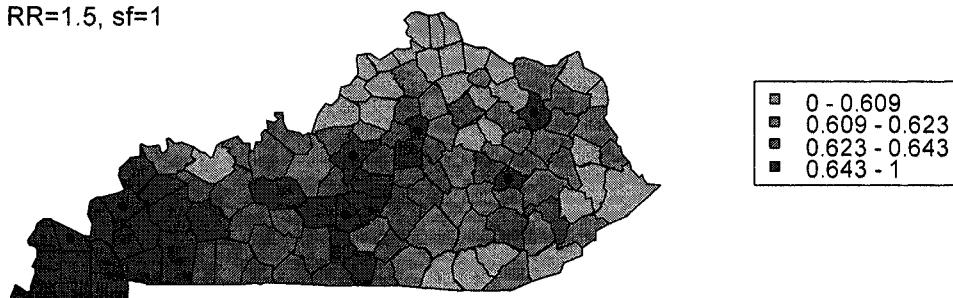
## A1. Receiver Operating Characteristic curves for the BYM model



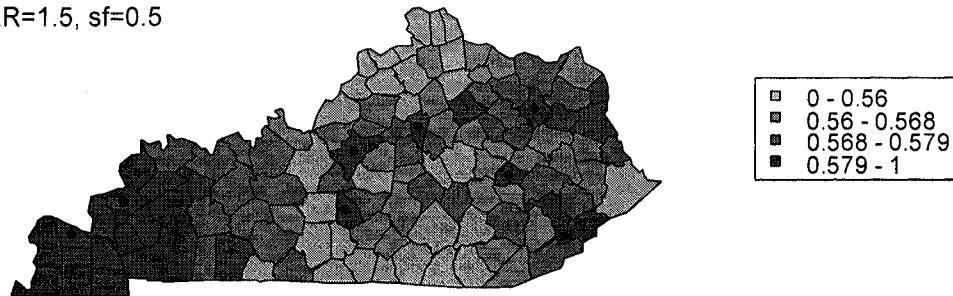
## A2. Disease maps of the posterior proportion and the posterior mean for the BYM model

Disease map of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 1.5

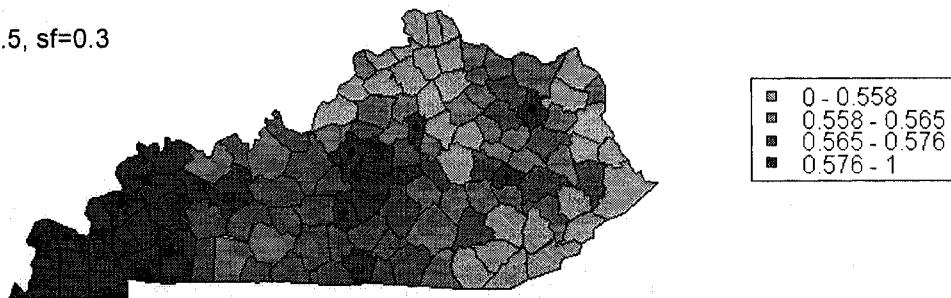
RR=1.5, sf=1



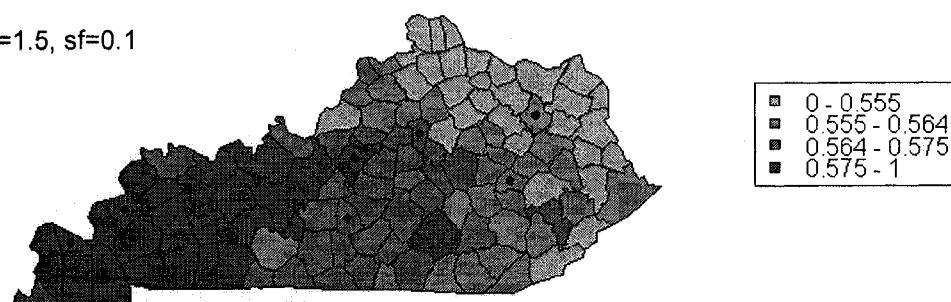
RR=1.5, sf=0.5



RR=1.5, sf=0.3

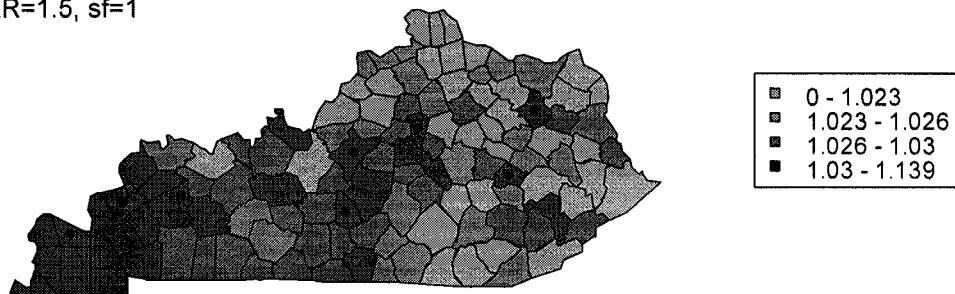


RR=1.5, sf=0.1

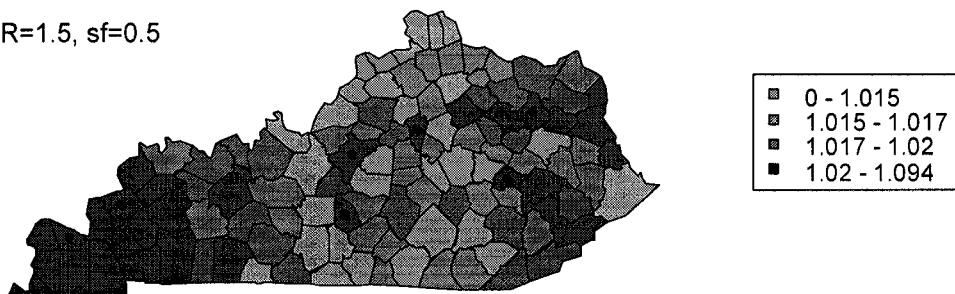


Disease maps of the posterior mean of the relative risk for simulations with true hot spot relative risk equal to 1.5.

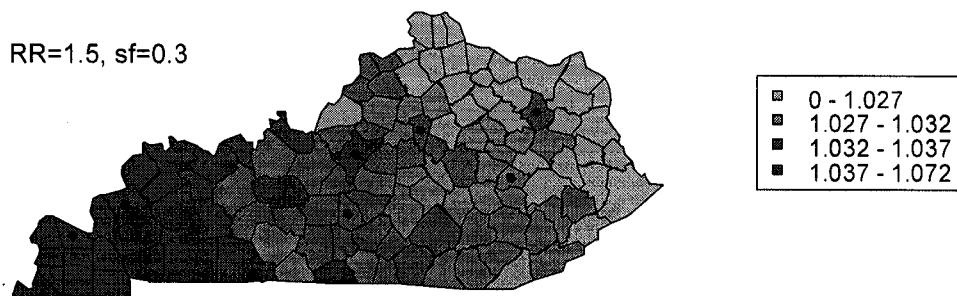
RR=1.5, sf=1



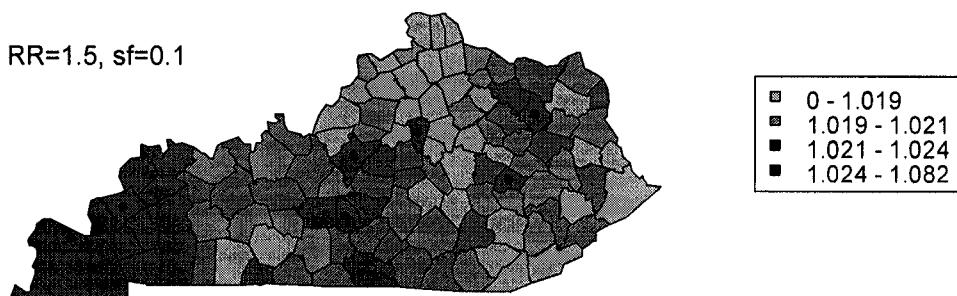
RR=1.5, sf=0.5



RR=1.5, sf=0.3

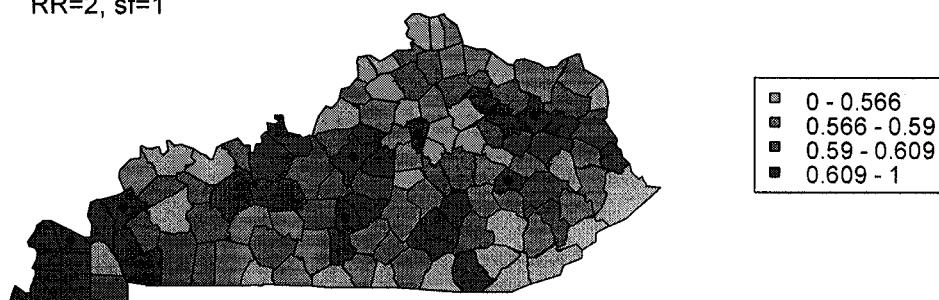


RR=1.5, sf=0.1

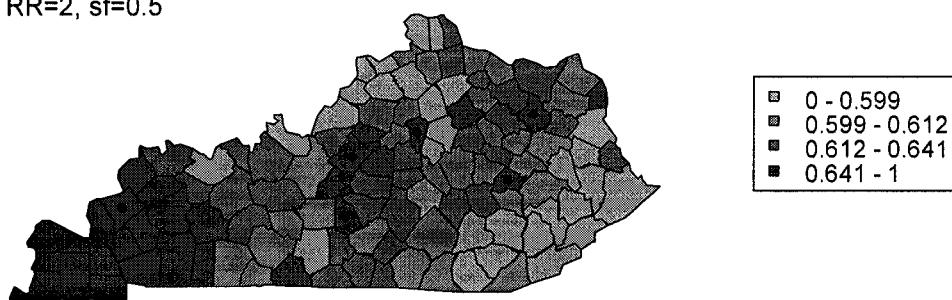


Disease maps of posterior proportion for the BYM model, for simulations with true hot spot relative risk equal to 2 and 3

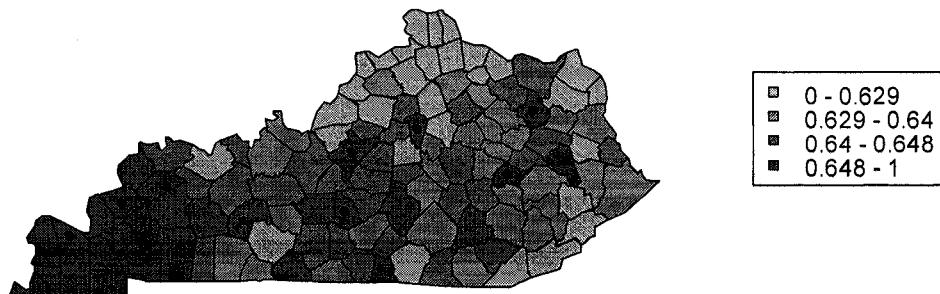
RR=2, sf=1



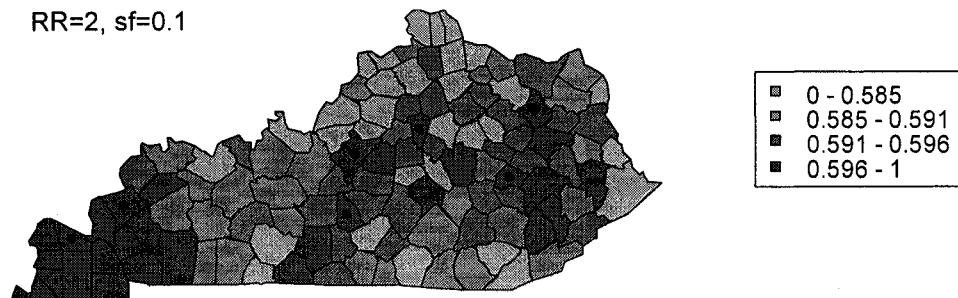
RR=2, sf=0.5



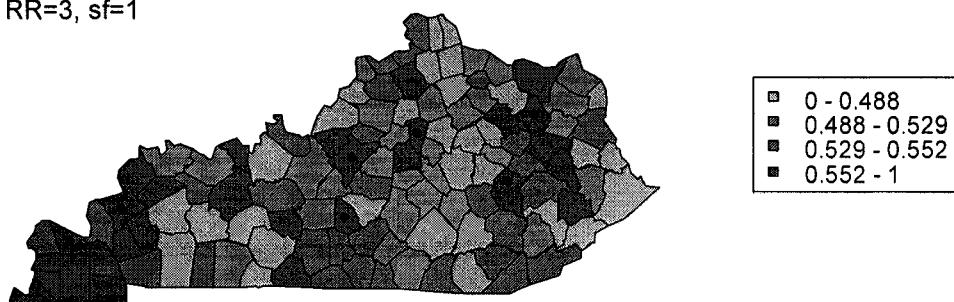
RR=2, sf=0.3



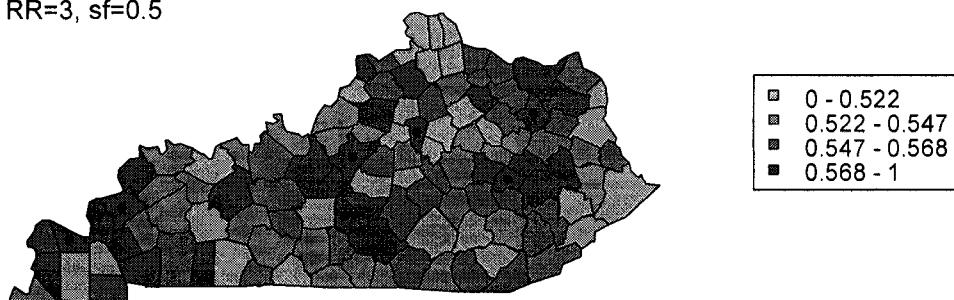
RR=2, sf=0.1



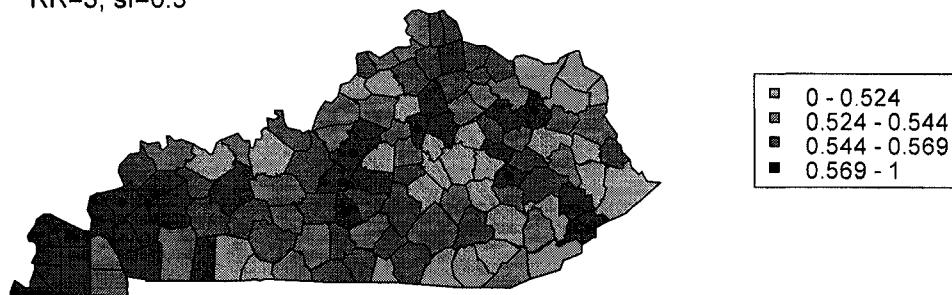
RR=3, sf=1



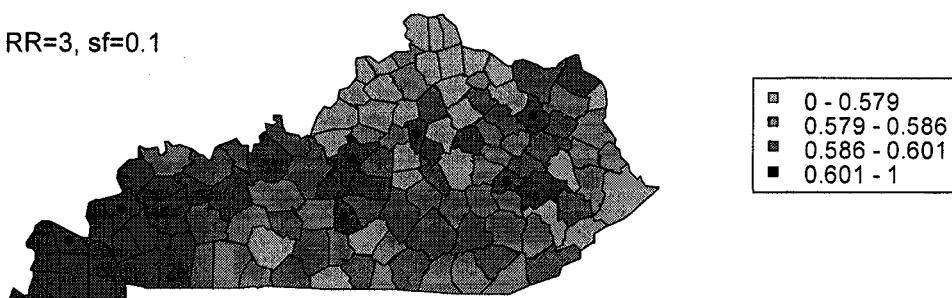
RR=3, sf=0.5



RR=3, sf=0.3

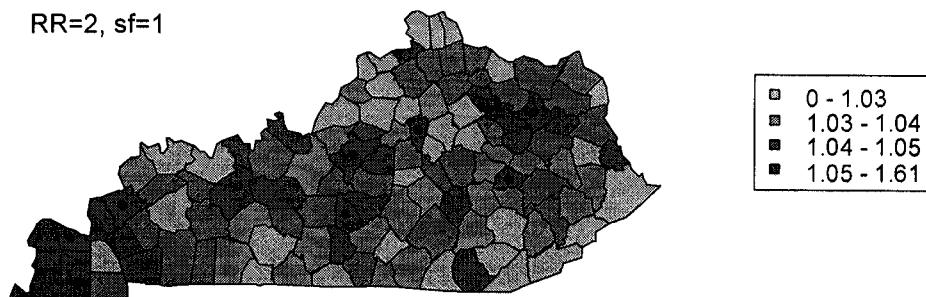


RR=3, sf=0.1

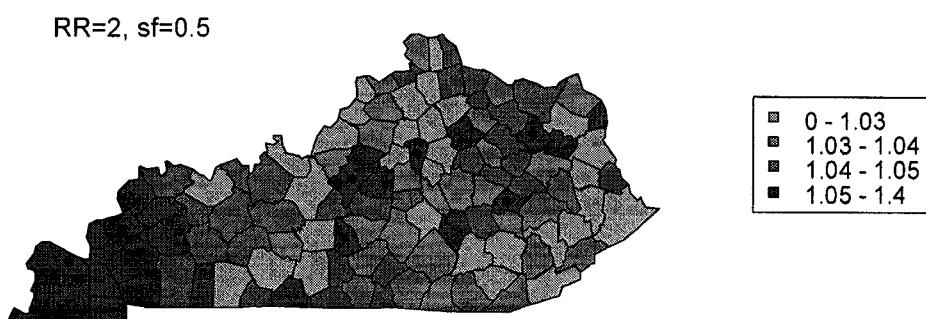


Disease maps of posterior mean relative risk for the BYM model, for simulations with true hot spot relative risk equal to 2 or 3

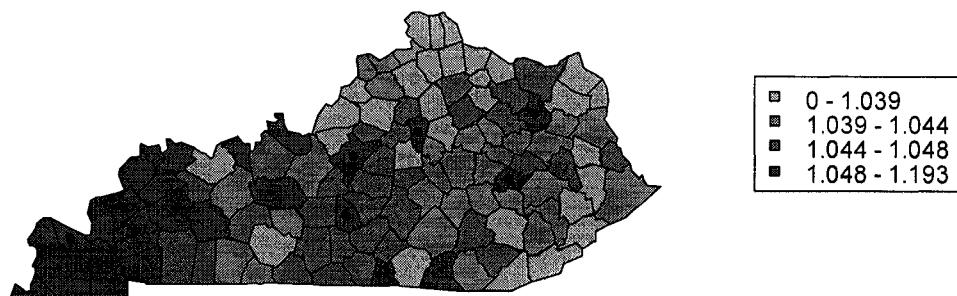
RR=2, sf=1



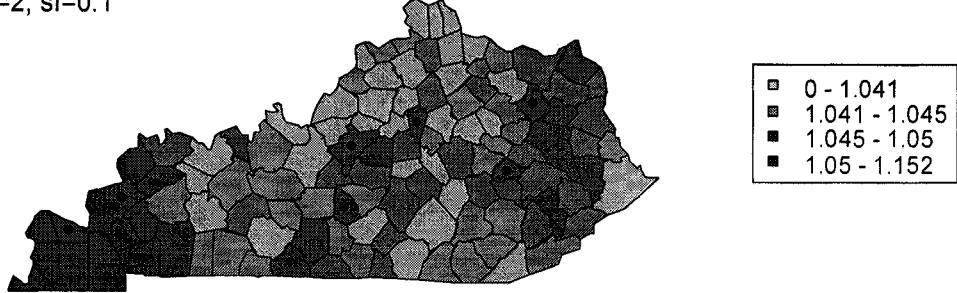
RR=2, sf=0.5



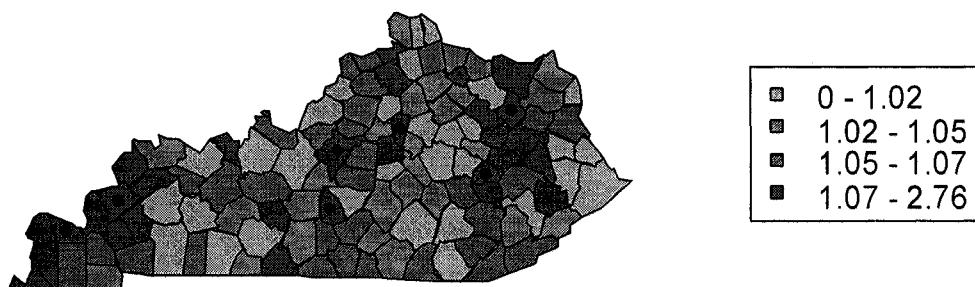
RR=2, sf=0.3



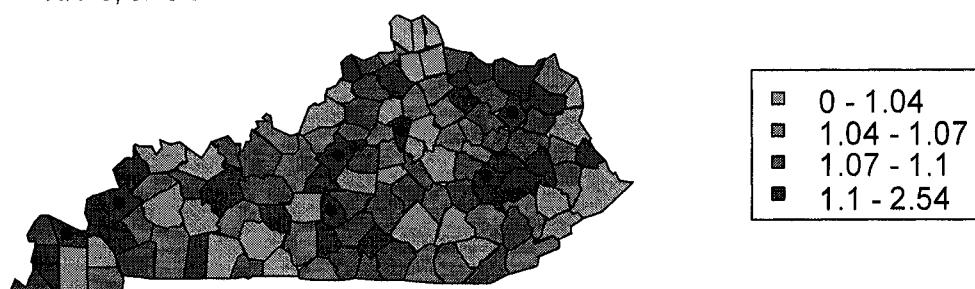
RR=2, sf=0.1



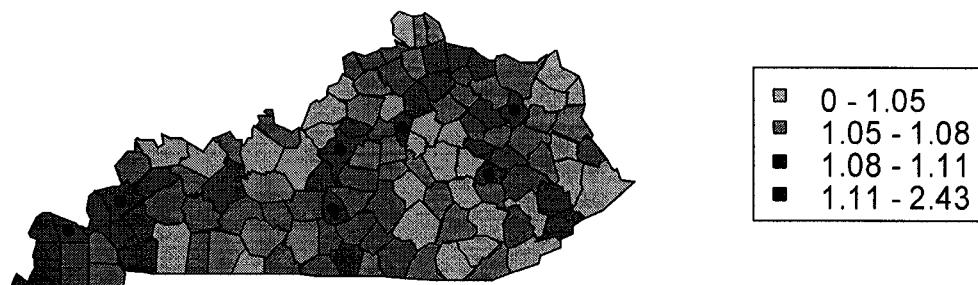
RR=3, sf=1



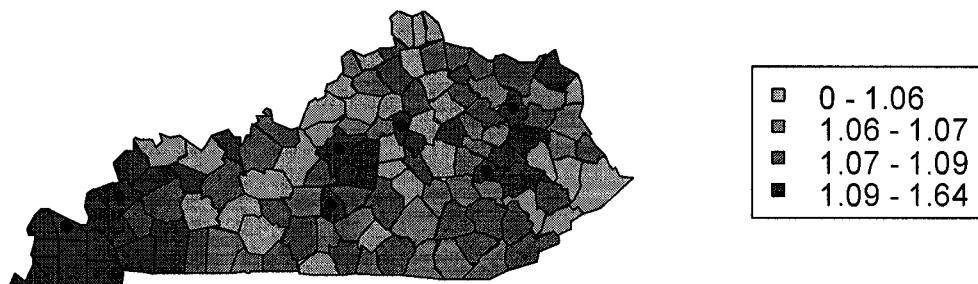
RR=3, sf=0.5



RR=3, sf=0.3



RR=3, sf=0.1



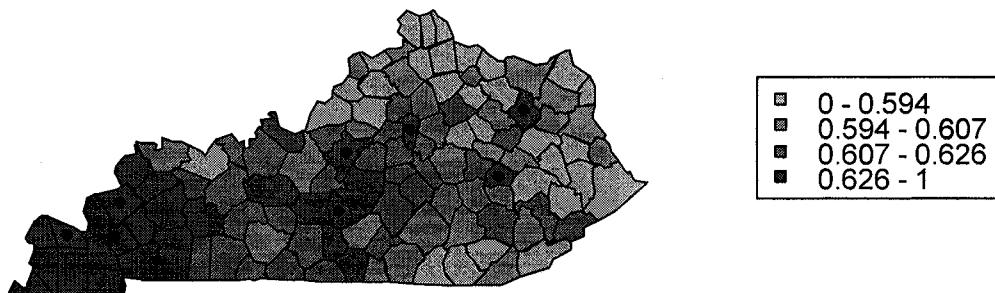
### A3. Disease maps of spatial and spatio-temporal models, Scenario 1

Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 1.5, scale factor equal to 1, and for the temporal pattern of Scenario 1

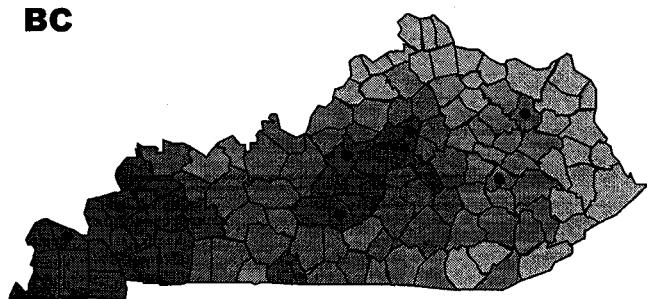
**BYM**



**KHB**

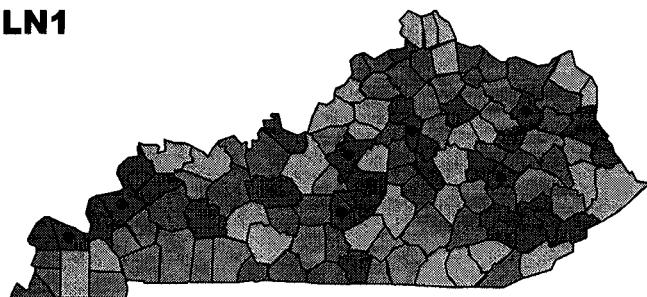


**BC**



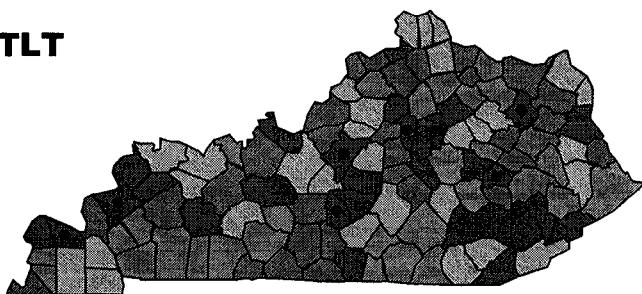
■	0 - 0.662
■	0.662 - 0.687
■	0.687 - 0.712
■	0.712 - 1

**LN1**



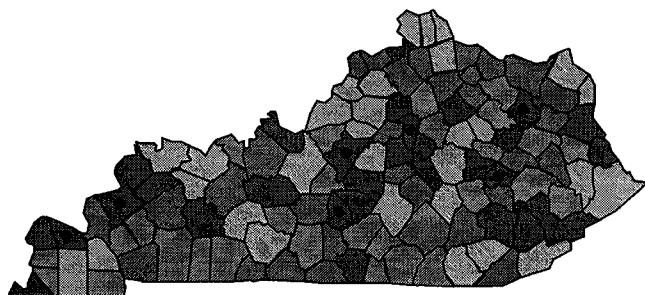
■	0 - 0.599
■	0.599 - 0.611
■	0.611 - 0.619
■	0.619 - 1

**TLT**



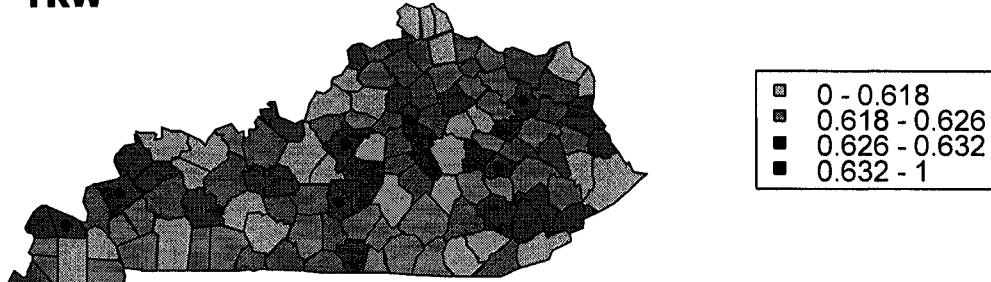
■	0 - 0.513
■	0.513 - 0.525
■	0.525 - 0.535
■	0.535 - 1

**LN2**



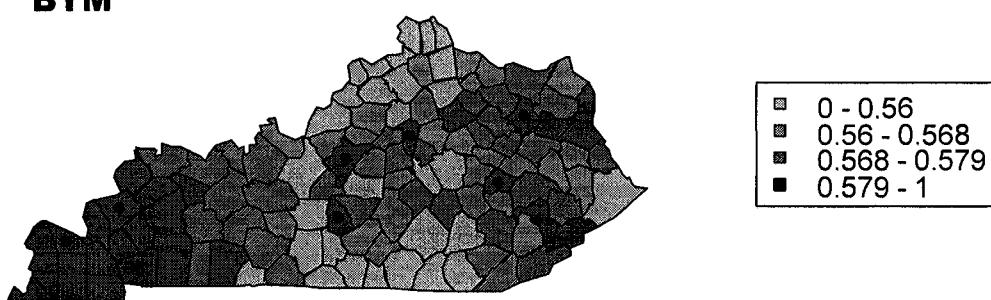
■	0 - 0.645
■	0.645 - 0.651
■	0.651 - 0.659
■	0.659 - 1

**TRW**

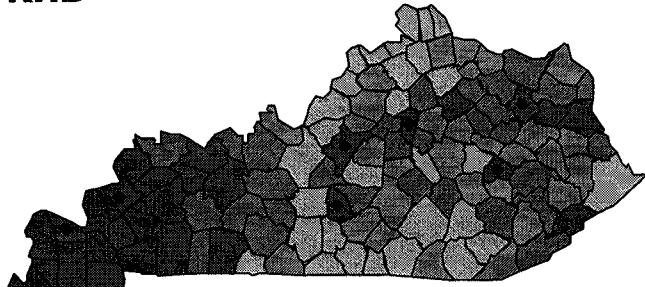


Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 1.5, scale factor equal to 0.5, and for the temporal pattern of Scenario 1

**BYM**

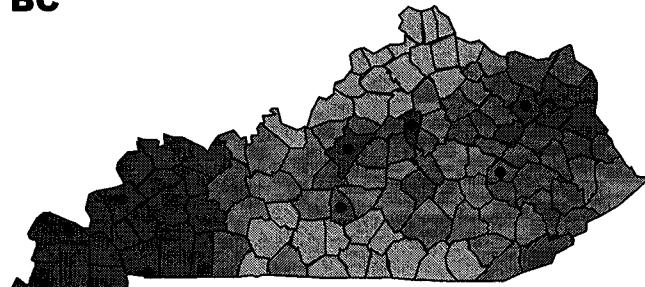


**KHB**



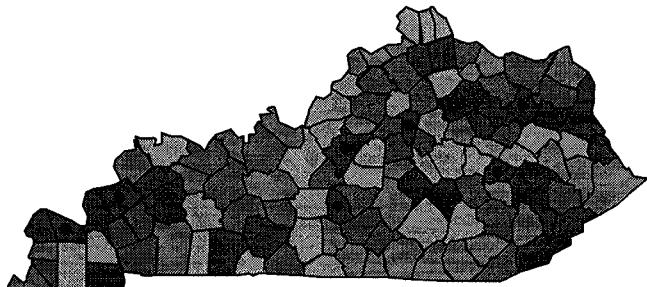
■	0 - 0.551
■	0.551 - 0.557
■	0.557 - 0.57
■	0.57 - 1

**BC**



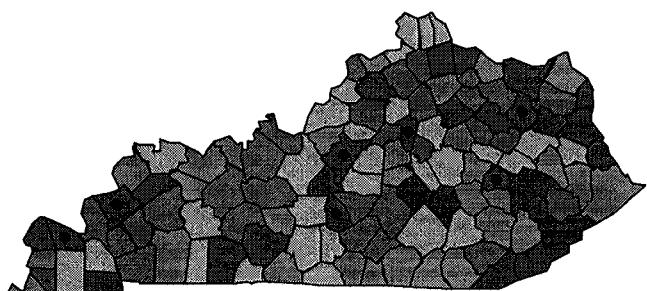
■	0 - 0.595
■	0.595 - 0.602
■	0.602 - 0.609
■	0.609 - 1

**LN1**



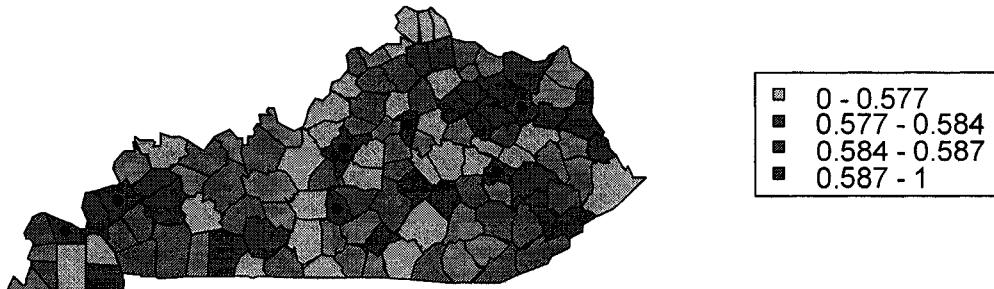
■	0 - 0.547
■	0.547 - 0.555
■	0.555 - 0.562
■	0.562 - 1

**TLT**

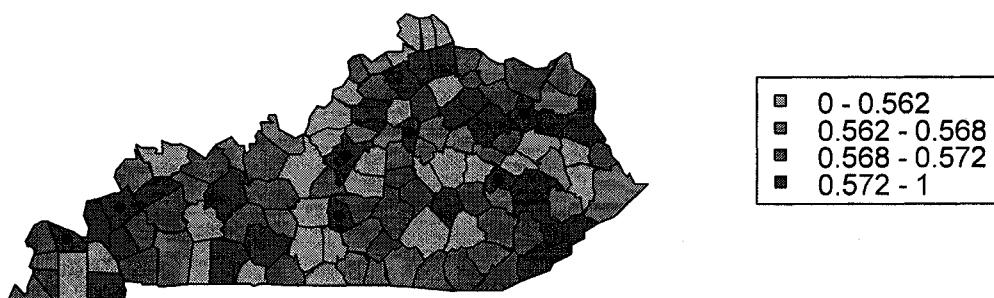


■	0 - 0.521
■	0.521 - 0.531
■	0.531 - 0.538
■	0.538 - 1

**LN2**

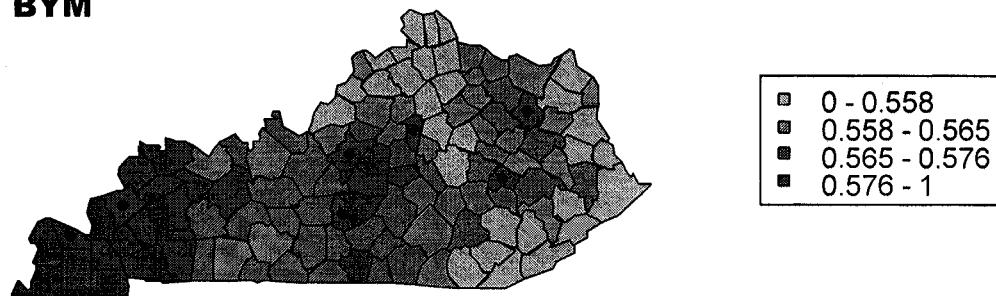


**TRW**

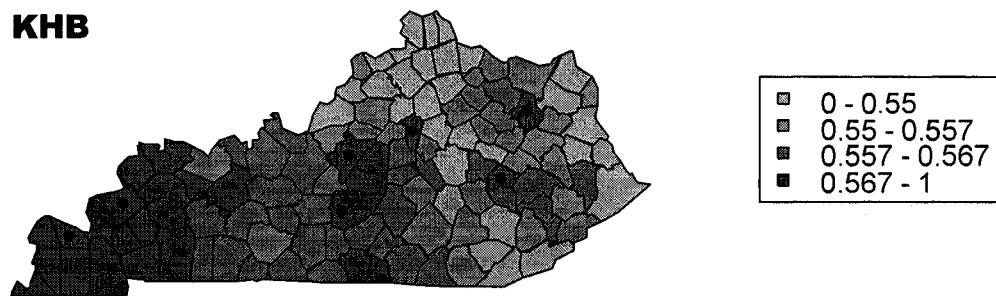


Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 1.5, scale factor equal to 0.3, and for the temporal pattern of Scenario 1

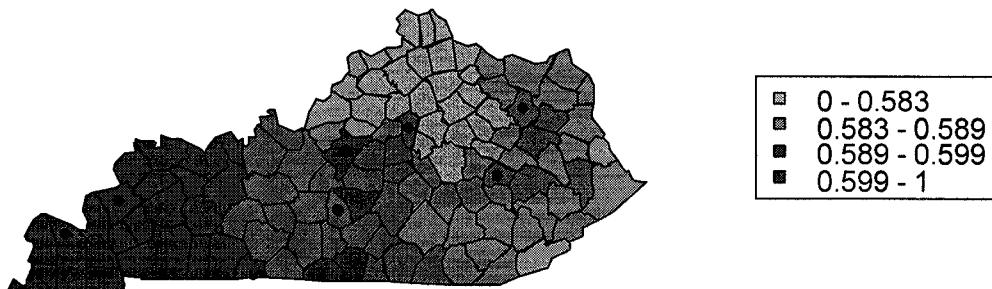
**BYM**



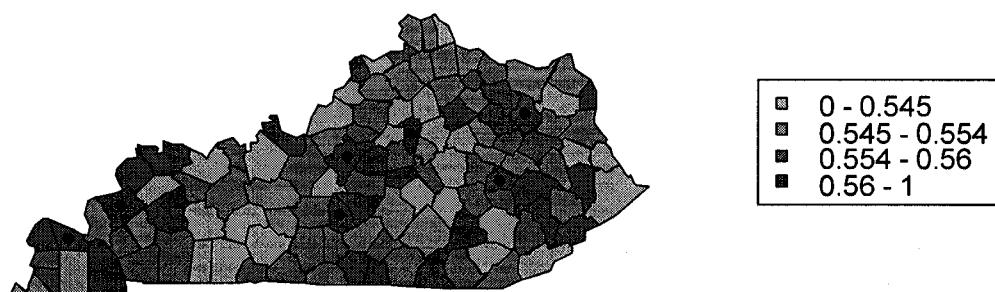
**KHB**



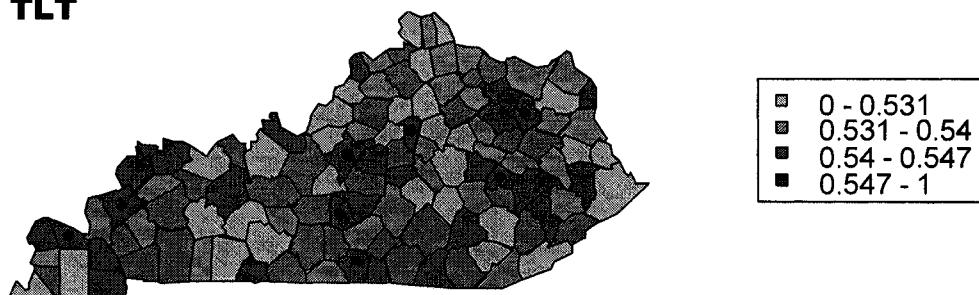
**BC**



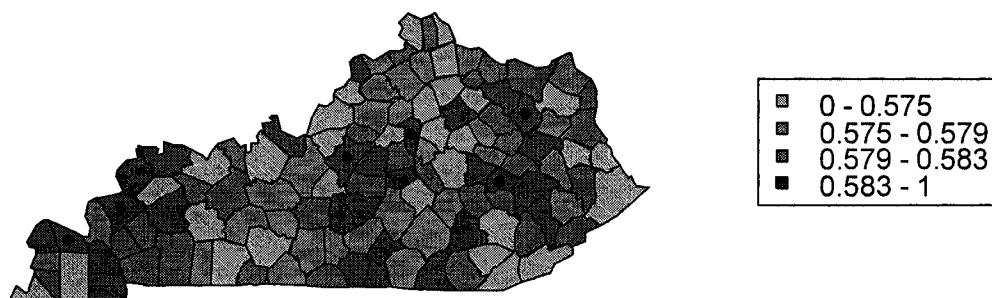
**LN1**



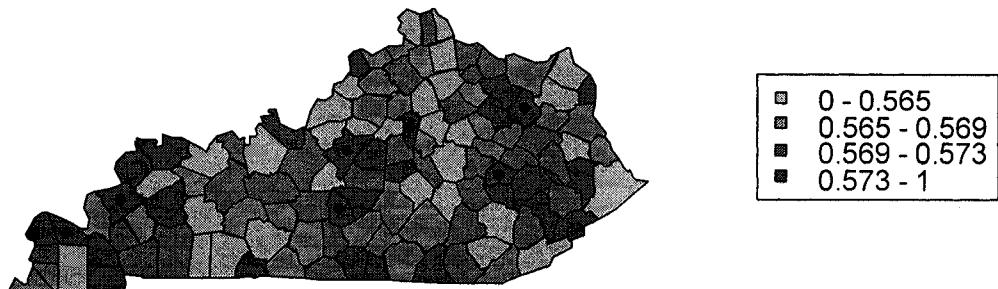
**TLT**



**LN2**

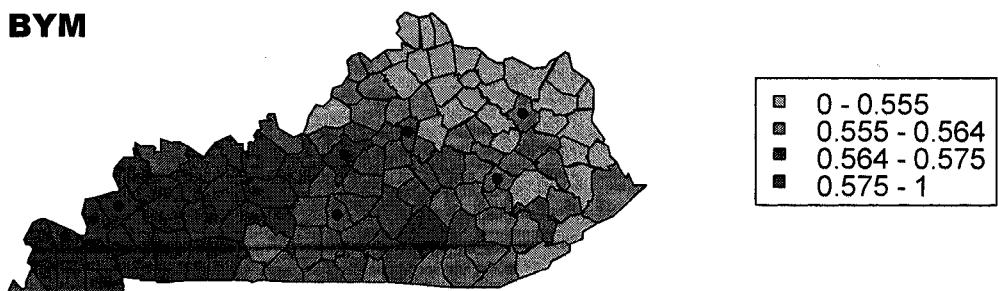


**TRW**

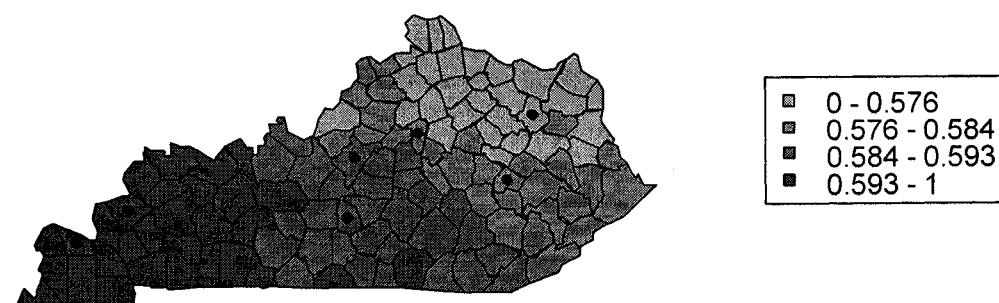
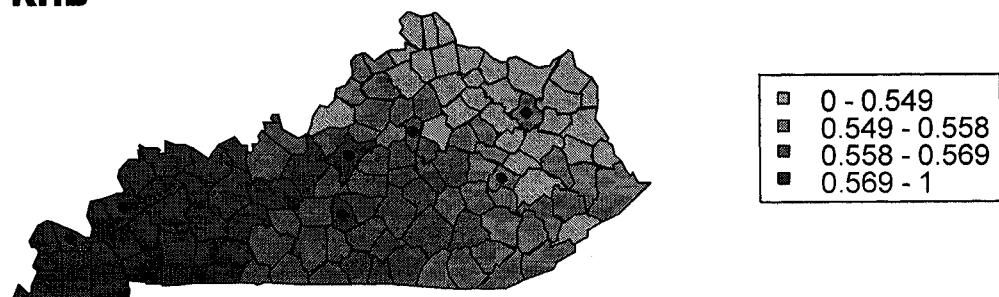


Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 1.5, scale factor equal to 0.1, and for the temporal pattern of Scenario 1

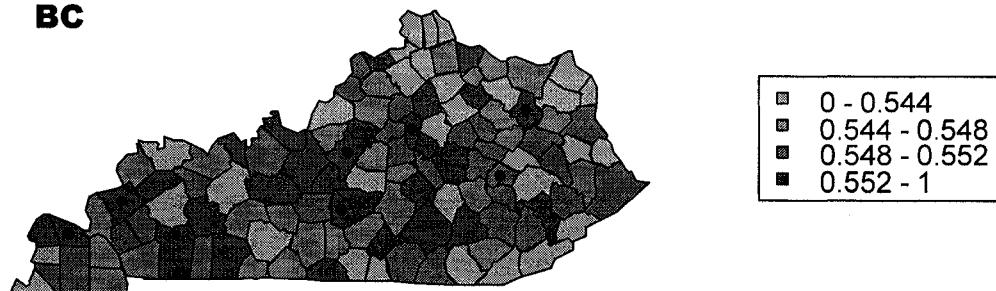
**BYM**



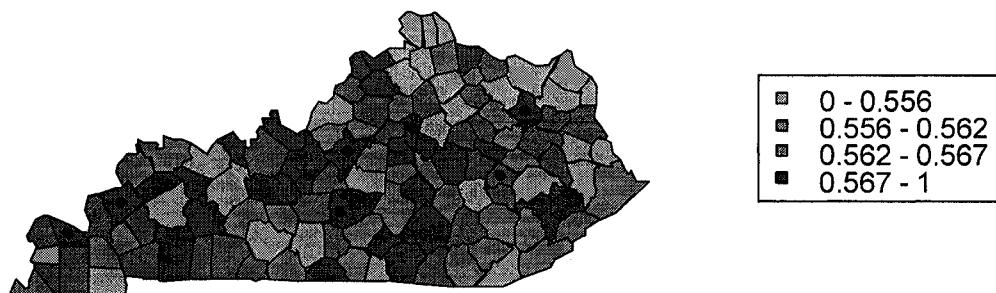
**KHB**



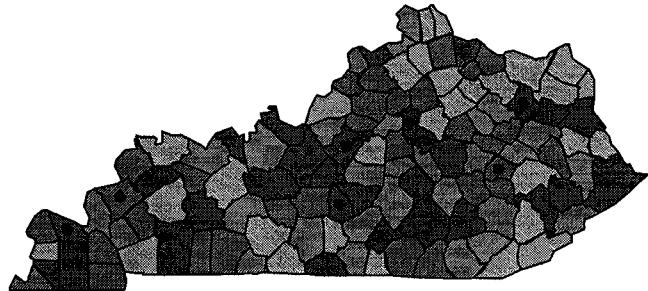
**BC**



**LN1**

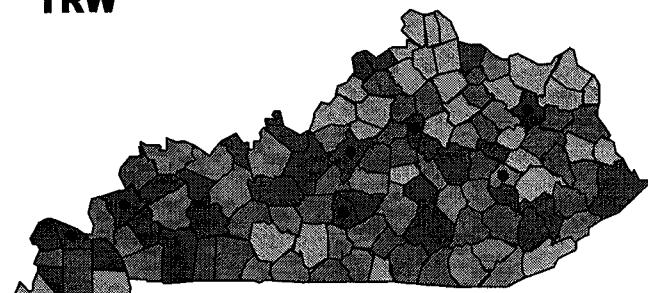


**TLT**



■	0 - 0.578
■	0.578 - 0.58
■	0.58 - 0.582
■	0.582 - 1

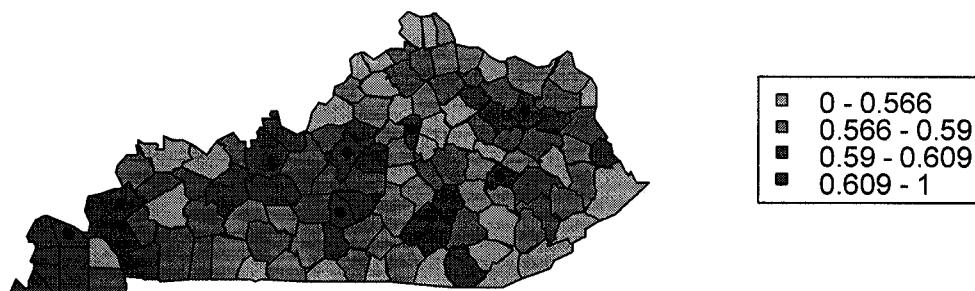
**TRW**



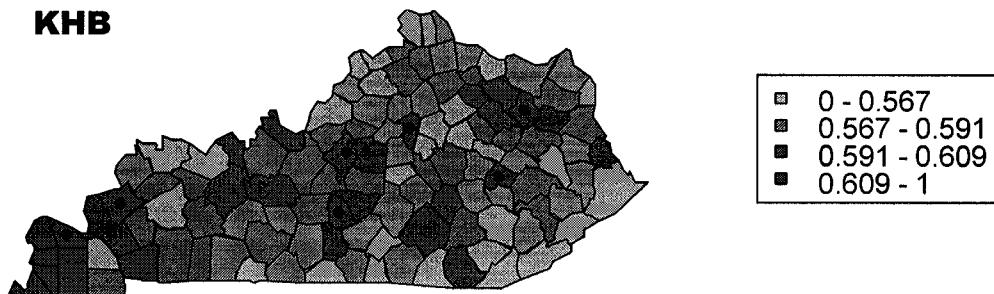
■	0 - 0.563
■	0.563 - 0.565
■	0.565 - 0.567
■	0.567 - 1

Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 2, scale factor equal to 1, and for the temporal pattern of Scenario 1

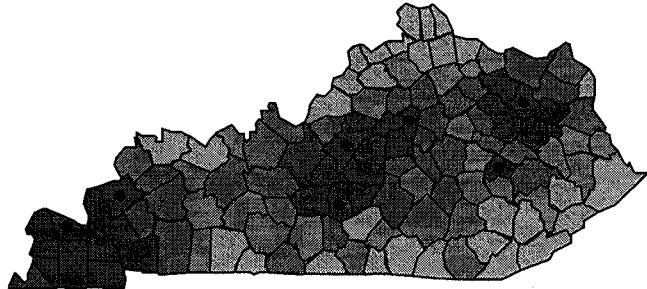
**BYM**



**KHB**

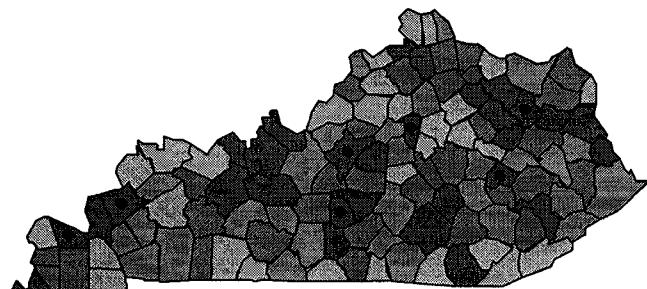


**BC**



■	0 - 0.63
■	0.63 - 0.662
■	0.662 - 0.731
■	0.731 - 1

**LN1**

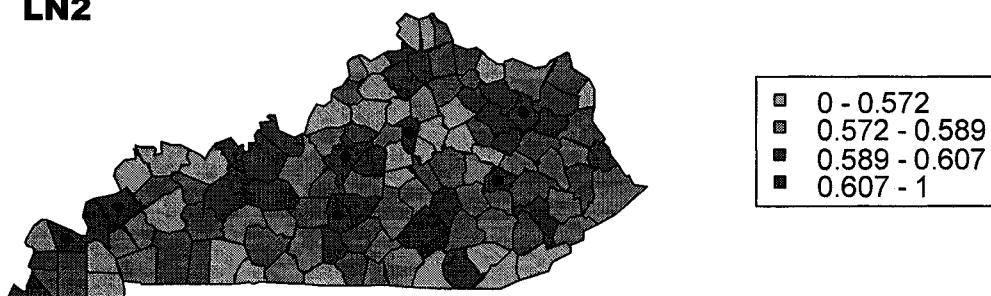


■	0 - 0.554
■	0.554 - 0.574
■	0.574 - 0.591
■	0.591 - 1

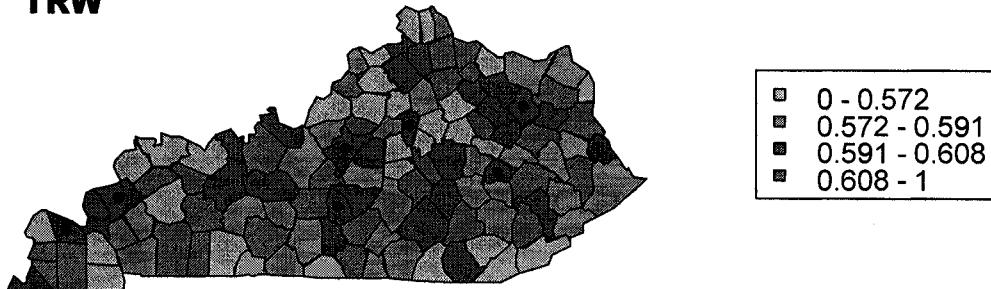
**TLT**



**LN2**

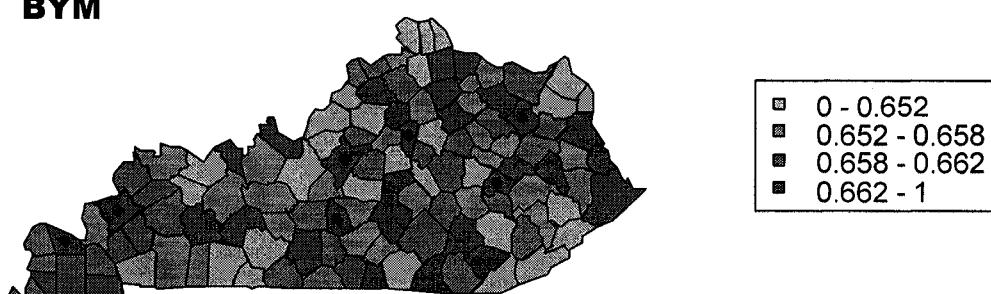


**TRW**



Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 2, scale factor equal to 0.5, and for the temporal pattern of Scenario 1

**BYM**

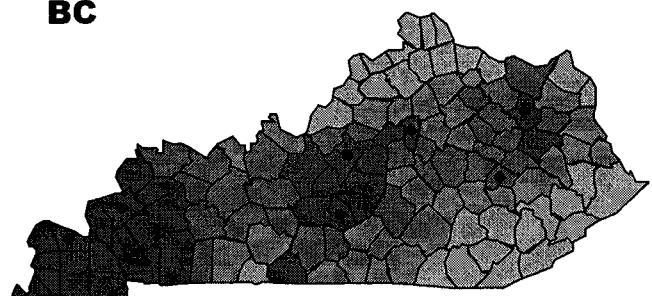


**KHB**



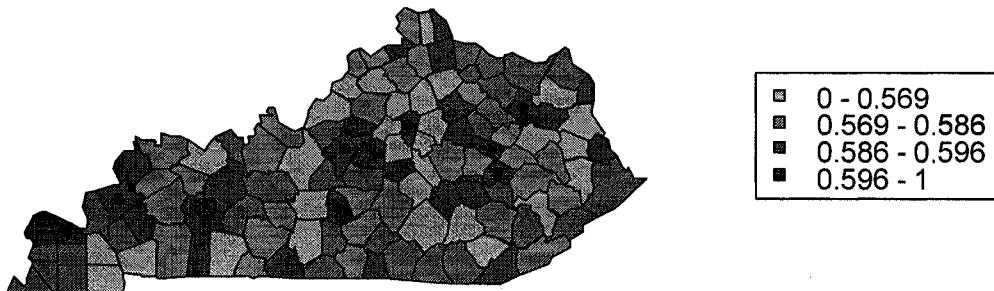
■	0 - 0.624
■	0.624 - 0.64
■	0.64 - 0.661
■	0.661 - 1

**BC**

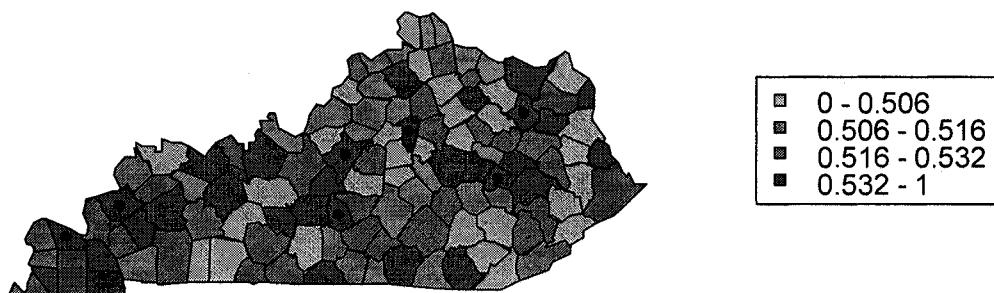


■	0 - 0.714
■	0.714 - 0.735
■	0.735 - 0.757
■	0.757 - 1

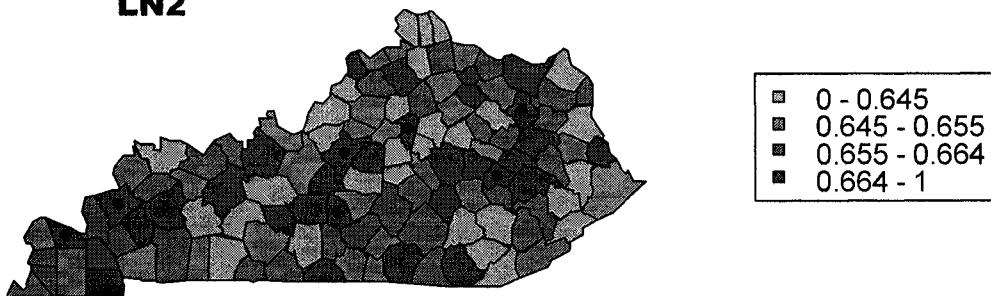
**LN1**



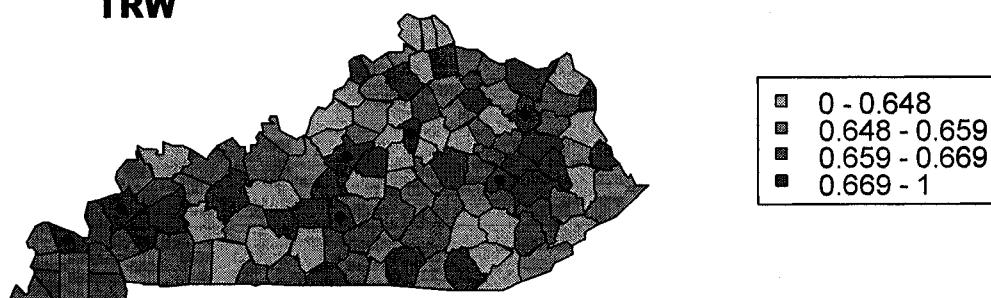
**TLT**



**LN2**

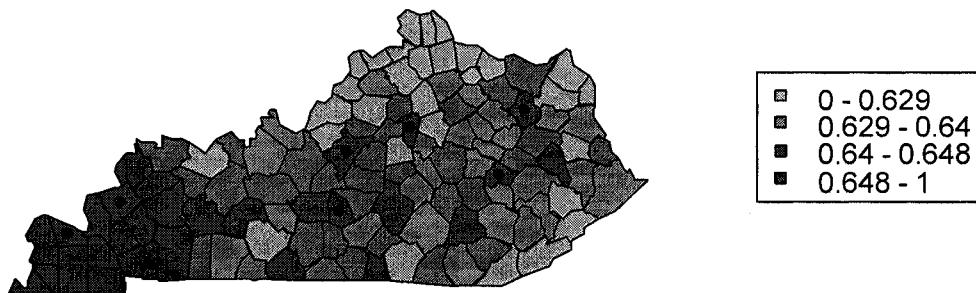


**TRW**

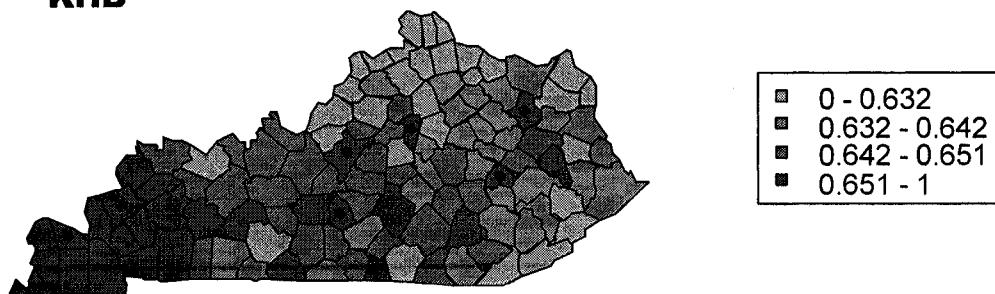


Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 2, scale factor equal to 0.3, and for the temporal pattern of Scenario 1

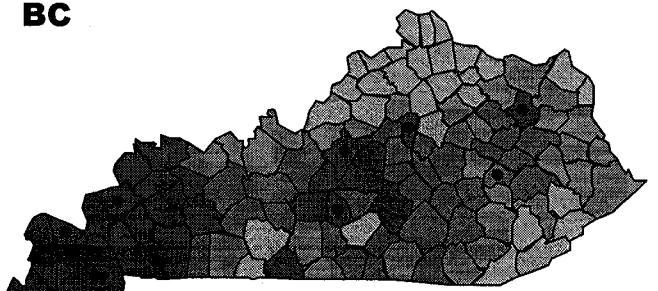
**BYM**



**KHB**



**BC**



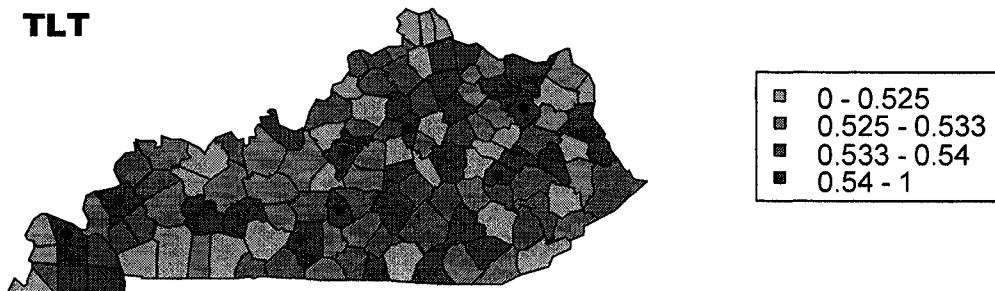
■	0 - 0.691
■	0.691 - 0.7
■	0.7 - 0.708
■	0.708 - 1

**LN1**

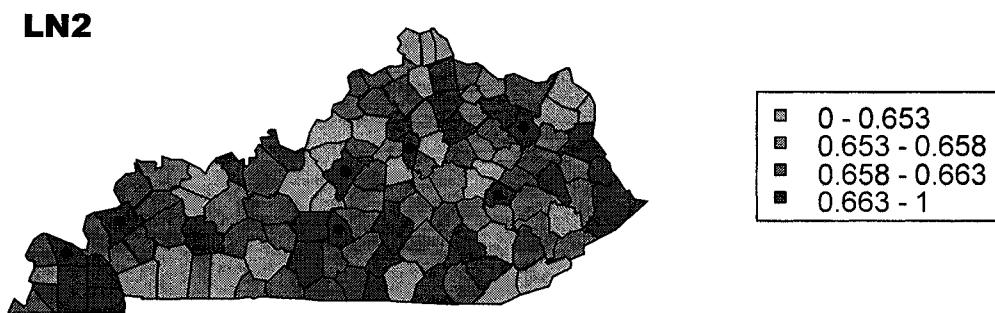


■	0 - 0.597
■	0.597 - 0.605
■	0.605 - 0.612
■	0.612 - 1

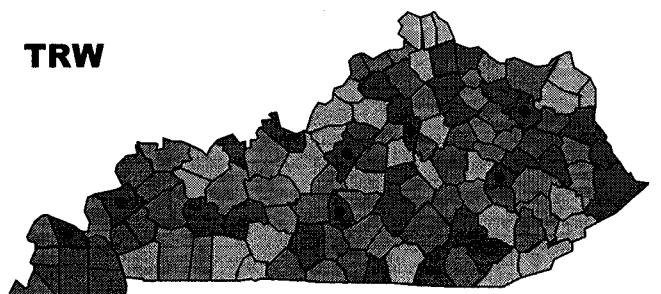
**TLT**



**LN2**



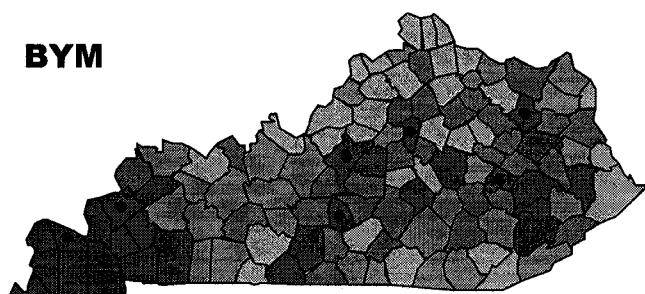
**TRW**



■	0 - 0.652
■	0.652 - 0.658
■	0.658 - 0.662
■	0.662 - 1

Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 2, scale factor equal to 0.1, and for the temporal pattern of Scenario 1

**BYM**



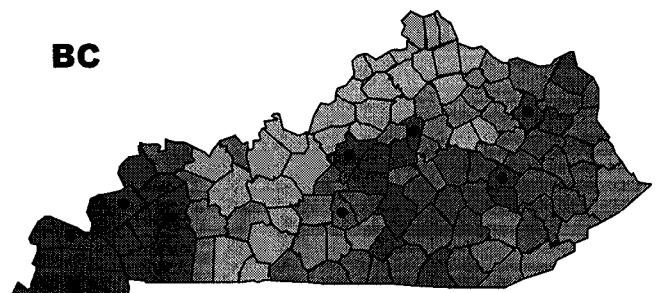
■	0 - 0.585
■	0.585 - 0.591
■	0.591 - 0.596
■	0.596 - 1

**KHB**



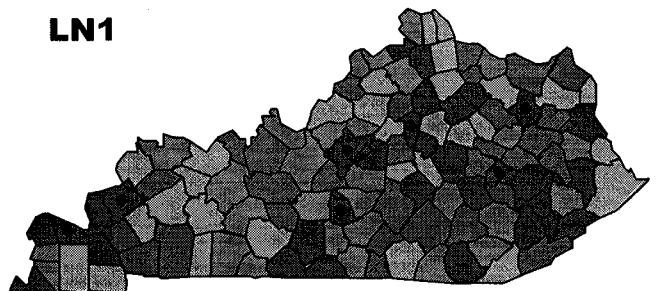
■	0 - 0.59
■	0.59 - 0.595
■	0.595 - 0.601
■	0.601 - 1

**BC**



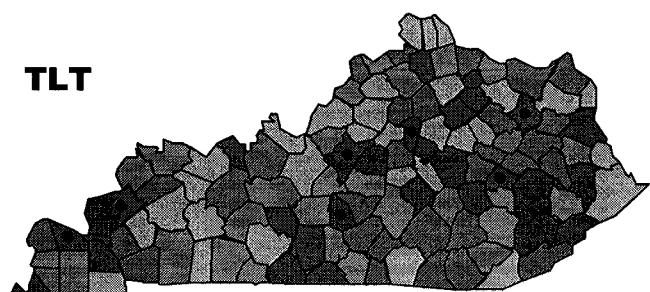
■	0 - 0.609
■	0.609 - 0.616
■	0.616 - 0.619
■	0.619 - 1

**LN1**



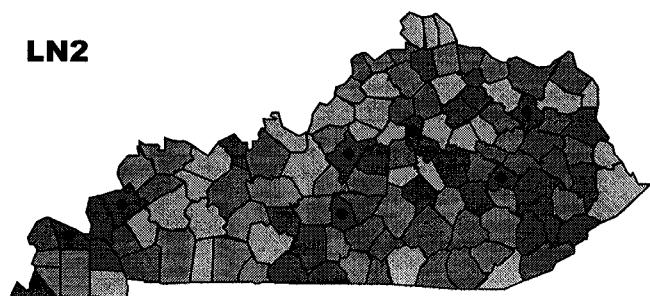
■	0 - 0.563
■	0.563 - 0.569
■	0.569 - 0.575
■	0.575 - 1

**TLT**



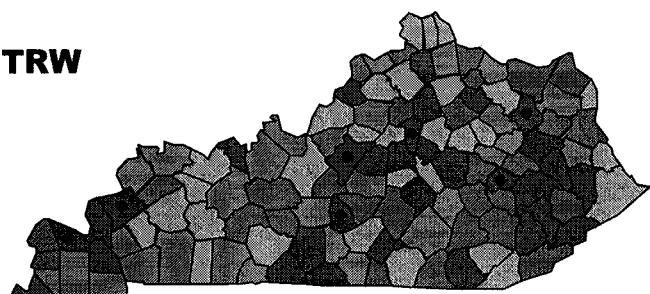
■	0 - 0.541
■	0.541 - 0.548
■	0.548 - 0.554
■	0.554 - 1

**LN2**



■	0 - 0.541
■	0.541 - 0.548
■	0.548 - 0.554
■	0.554 - 1

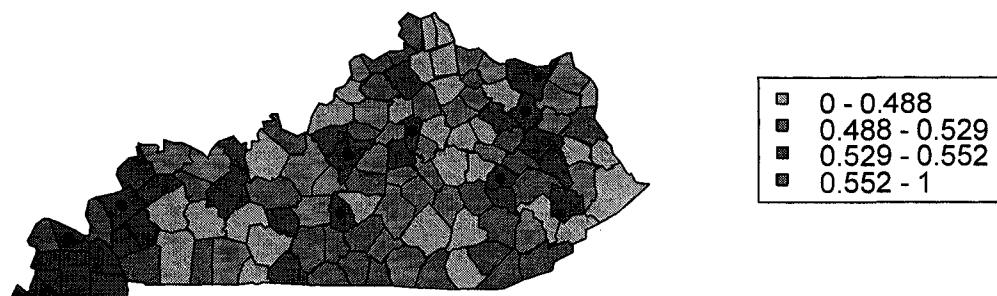
**TRW**



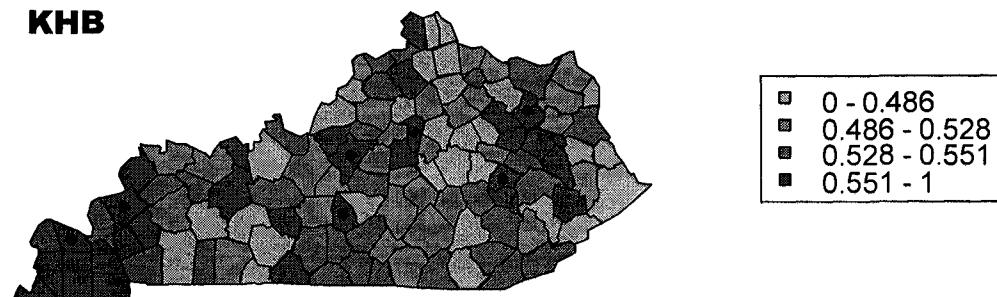
■	0 - 0.603
■	0.603 - 0.608
■	0.608 - 0.612
■	0.612 - 1

Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 3, scale factor equal to 1, and for the temporal pattern of Scenario 1

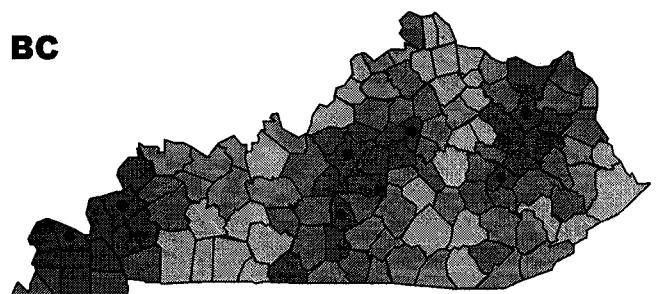
**BYM**



**KHB**

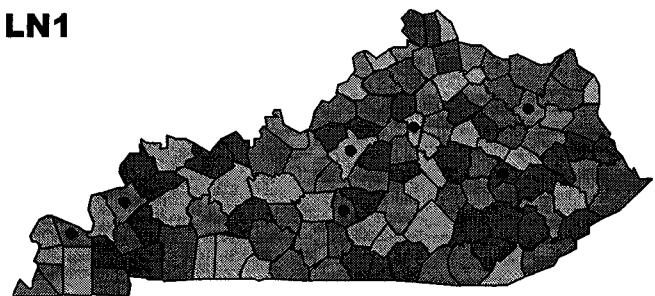


**BC**



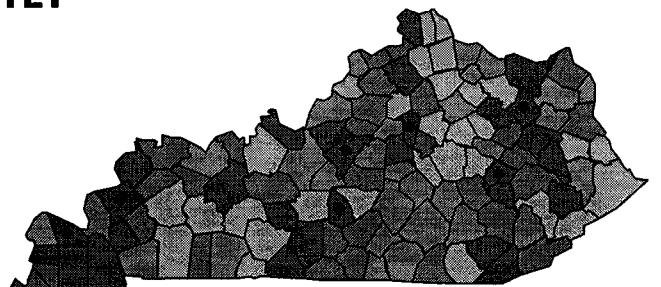
■	0 - 0.472
■	0.472 - 0.513
■	0.513 - 0.613
■	0.613 - 1

**LN1**



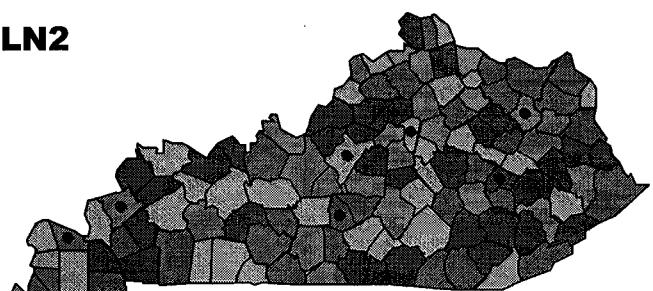
■	0 - 0.515
■	0.515 - 0.545
■	0.545 - 0.571
■	0.571 - 1

**TLT**



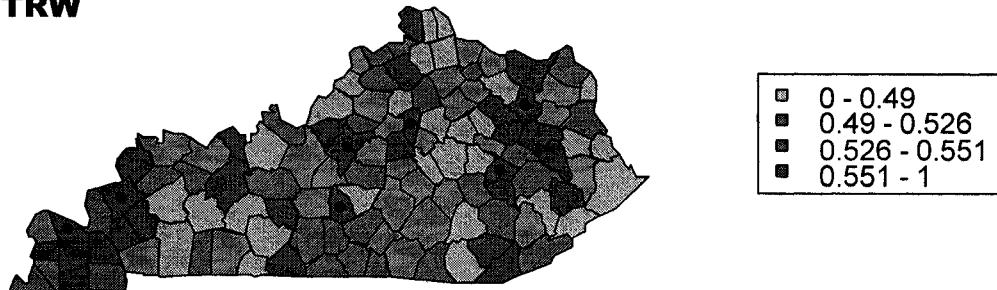
■	0 - 0.459
■	0.459 - 0.49
■	0.49 - 0.515
■	0.515 - 1

**LN2**



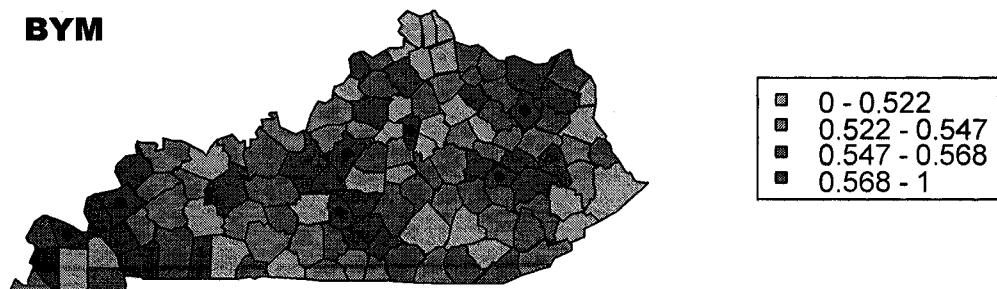
■	0 - 0.515
■	0.515 - 0.546
■	0.546 - 0.574
■	0.574 - 1

**TRW**

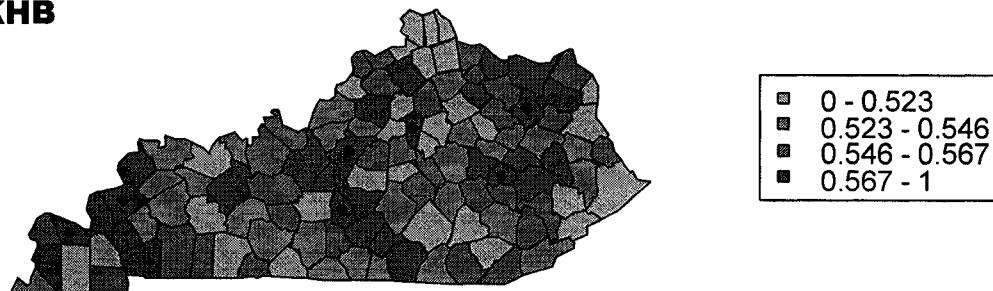


Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 3, scale factor equal to 0.5, and for the temporal pattern of Scenario 1

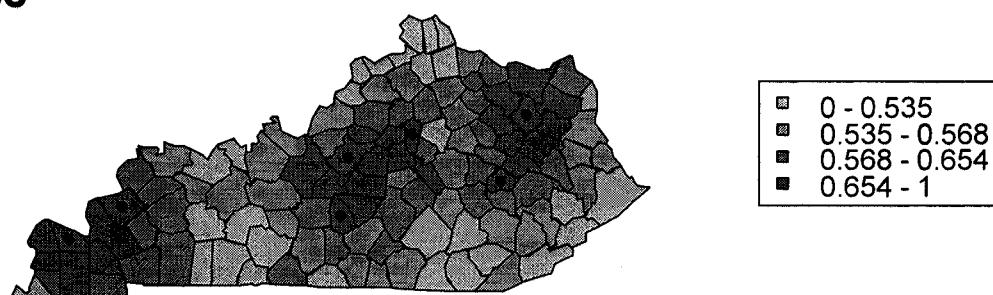
**BYM**



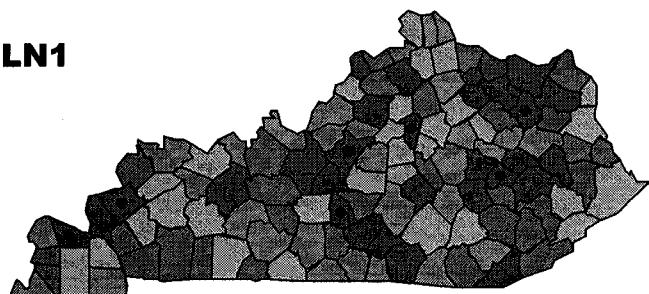
**KHB**



**BC**

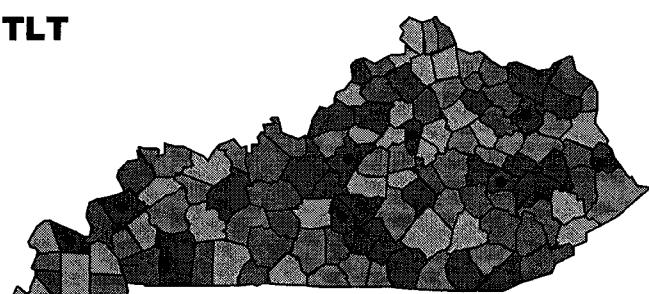


**LN1**



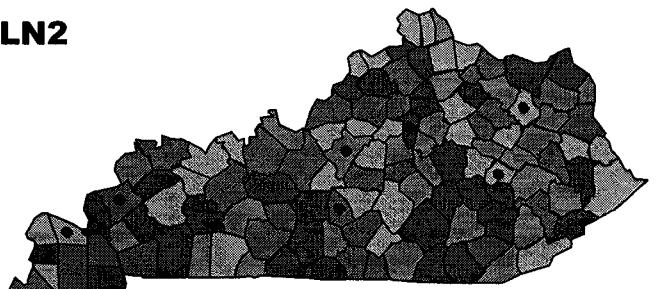
■	0 - 0.517
■	0.517 - 0.545
■	0.545 - 0.564
■	0.564 - 1

**TLT**



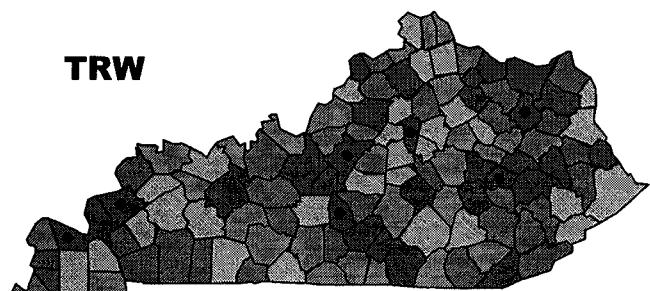
■	0 - 0.47
■	0.47 - 0.499
■	0.499 - 0.518
■	0.518 - 1

**LN2**



■	0 - 0.526
■	0.526 - 0.564
■	0.564 - 0.591
■	0.591 - 1

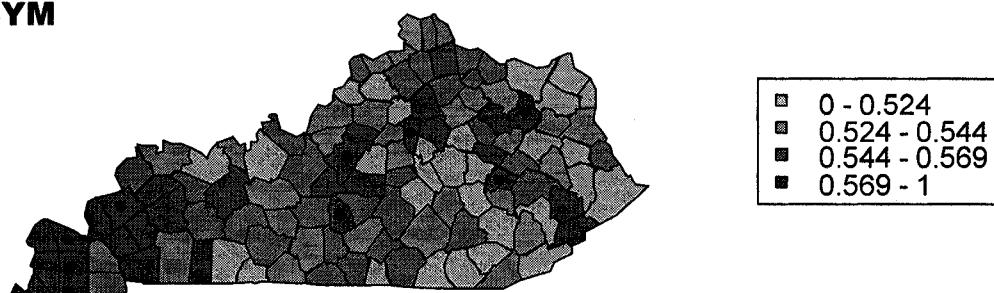
**TRW**



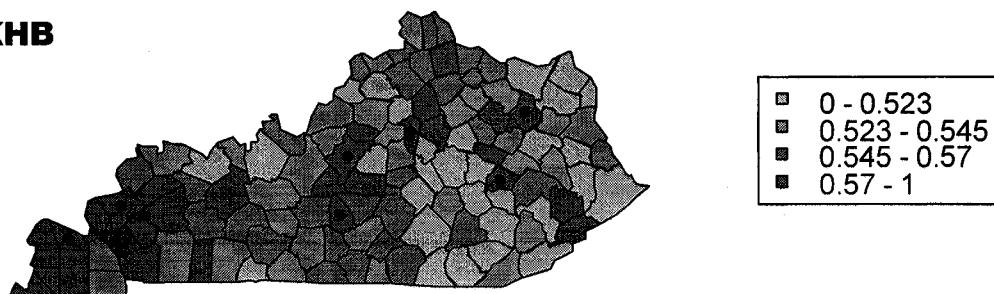
■	0 - 0.519
■	0.519 - 0.546
■	0.546 - 0.565
■	0.565 - 1

Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 3, scale factor equal to 0.3, and for the temporal pattern of Scenario 1

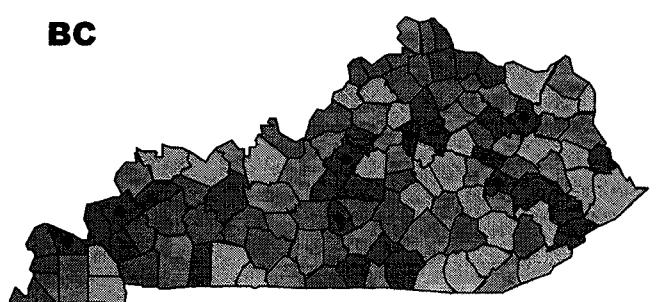
**BYM**



**KHB**

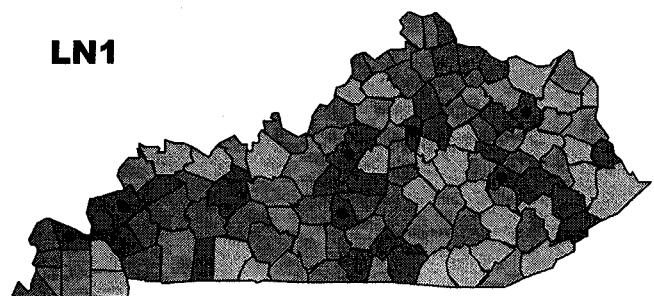


**BC**



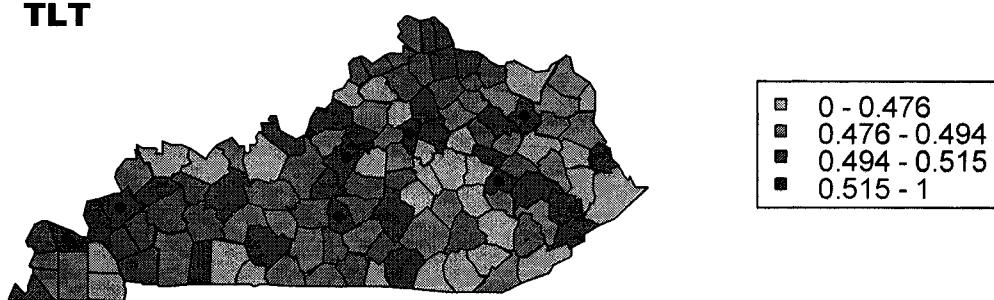
■	0 - 0.514
■	0.514 - 0.532
■	0.532 - 0.553
■	0.553 - 1

**LN1**

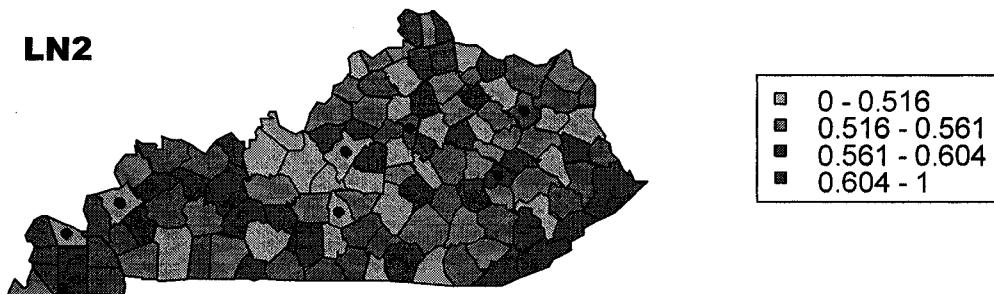


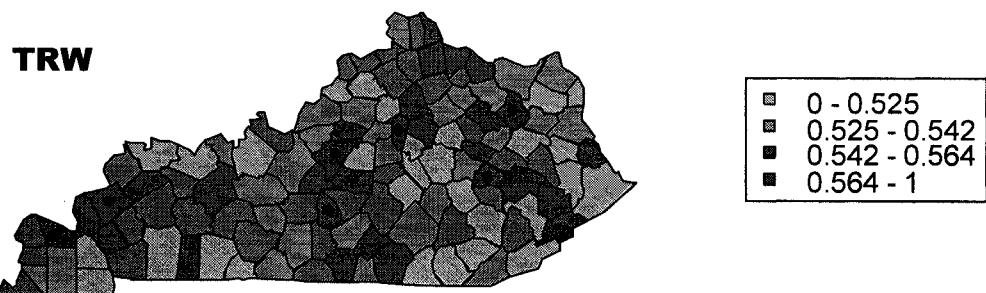
■	0 - 0.514
■	0.514 - 0.532
■	0.532 - 0.553
■	0.553 - 1

**TLT**

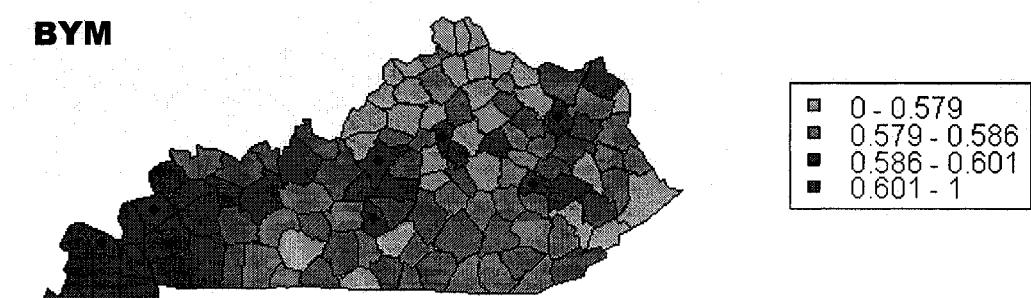


**LN2**

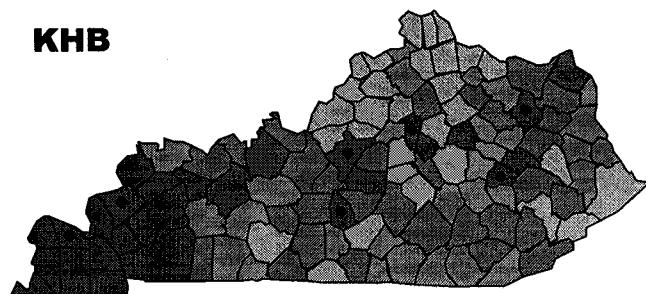




Disease maps of the posterior proportion greater than 1 for simulations with true hot spot relative risk equal to 3, scale factor equal to 0.1, and for the temporal pattern of Scenario 1

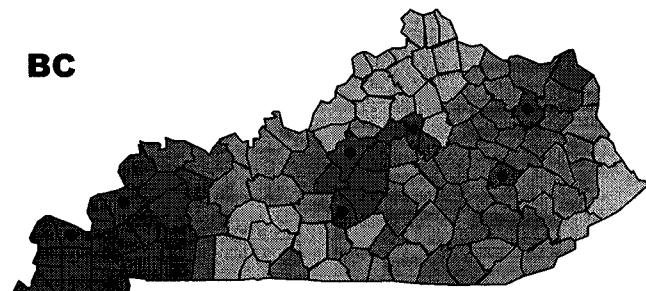


**KHB**



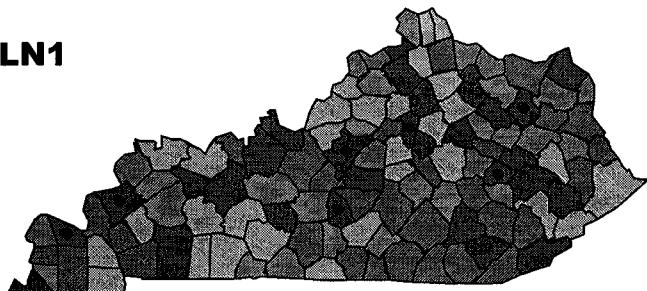
■	0 - 0.586
■	0.586 - 0.595
■	0.595 - 0.608
■	0.608 - 1

**BC**



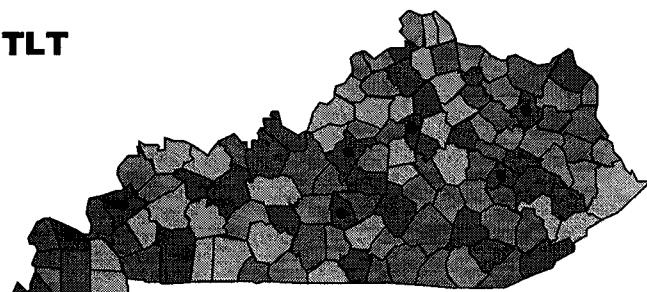
■	0 - 0.645
■	0.645 - 0.658
■	0.658 - 0.671
■	0.671 - 1

**LN1**



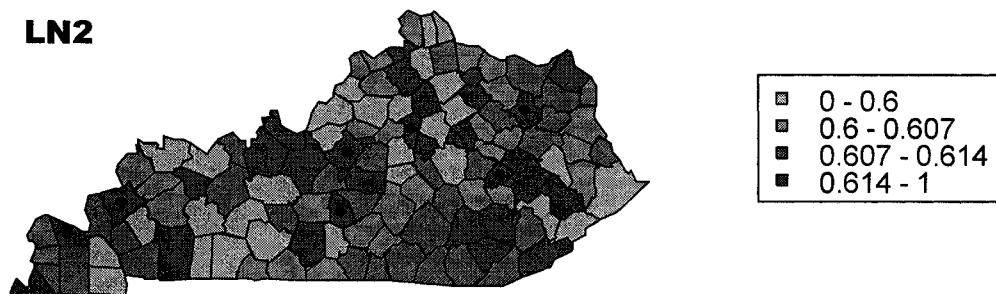
■	0 - 0.538
■	0.538 - 0.55
■	0.55 - 0.558
■	0.558 - 1

**TLT**

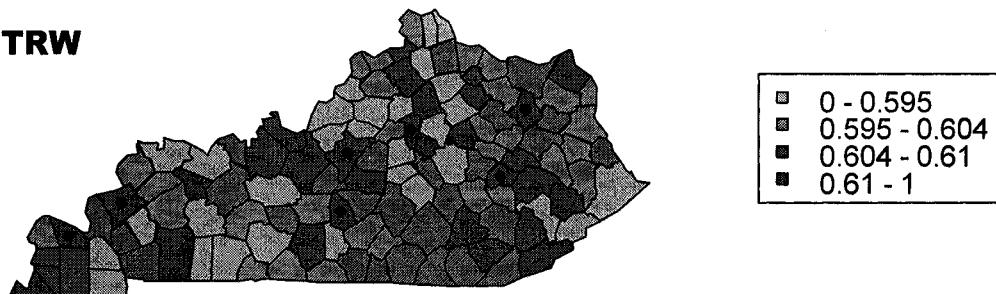


■	0 - 0.516
■	0.516 - 0.53
■	0.53 - 0.539
■	0.539 - 1

**LN2**

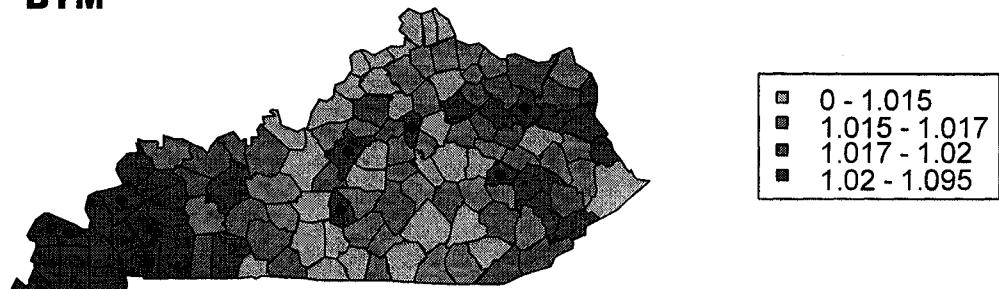


**TRW**

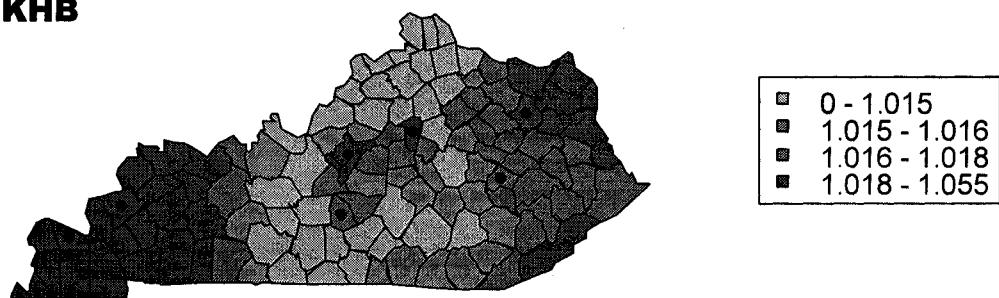


Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 1, scale factor equal to 0.5, and for the temporal pattern of Scenario 1

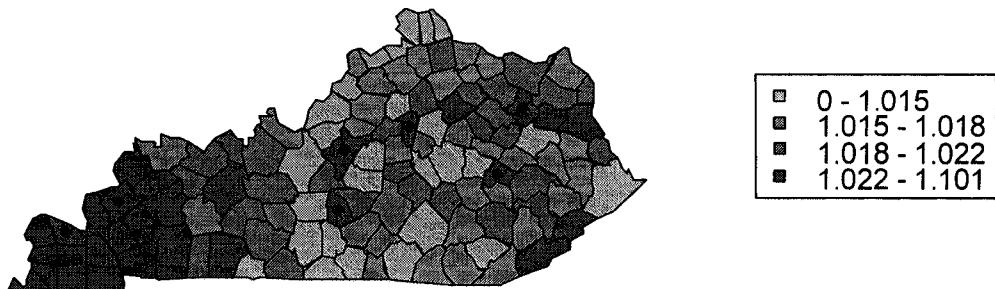
**BYM**



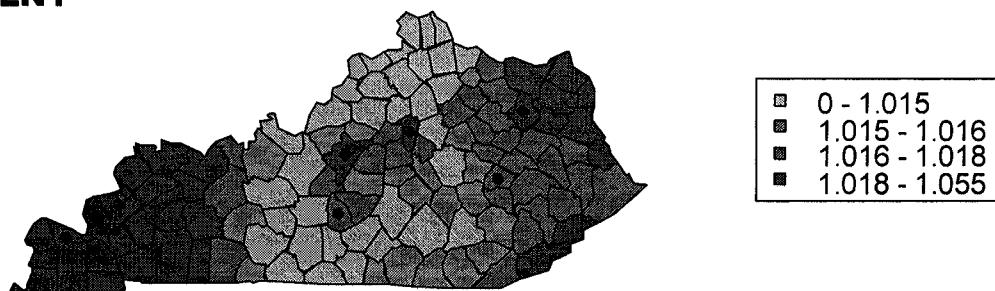
**KHB**



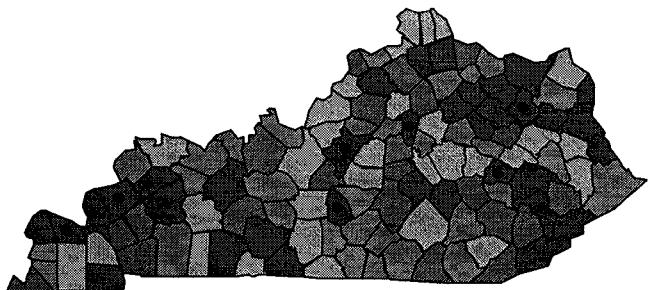
**BC**



**LN1**

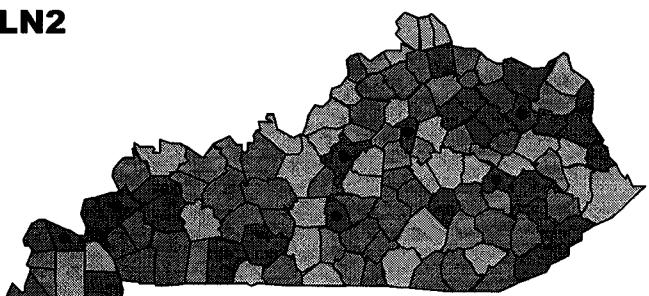


**TLT**



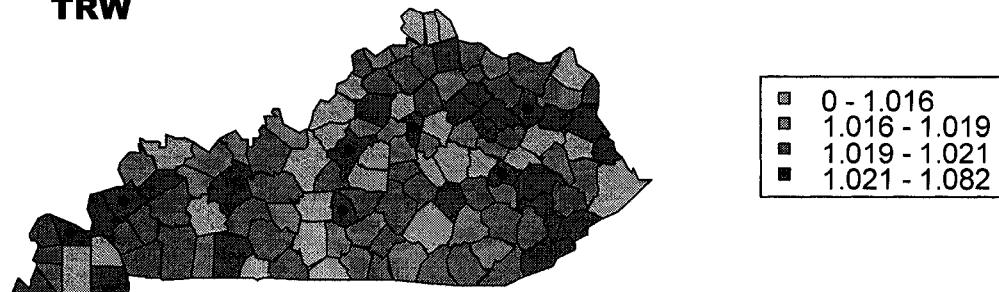
■	0 - 1.015
■	1.015 - 1.019
■	1.019 - 1.021
■	1.021 - 1.119

**LN2**



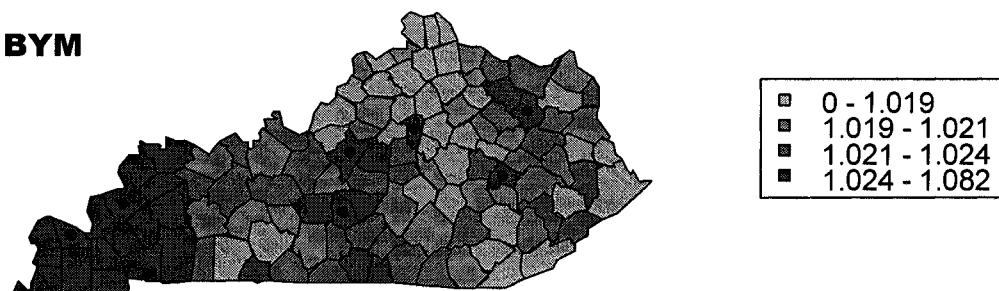
■	0 - 1.016
■	1.016 - 1.018
■	1.018 - 1.02
■	1.02 - 1.074

**TRW**

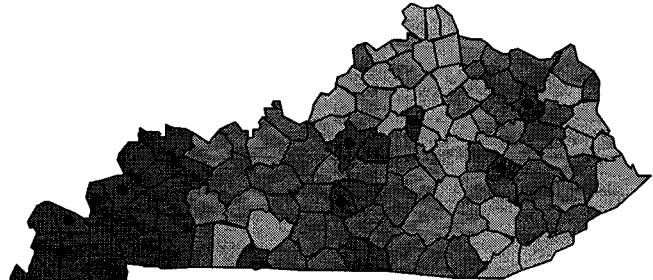


Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 1, scale factor equal to 0.3, and for the temporal pattern of Scenario 1

**BYM**

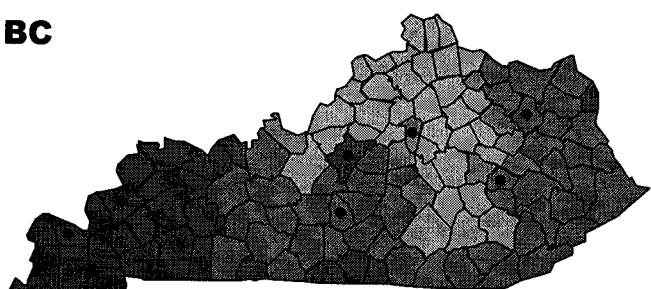


**KHB**



■	0 - 1.02
■	1.02 - 1.022
■	1.022 - 1.026
■	1.026 - 1.09

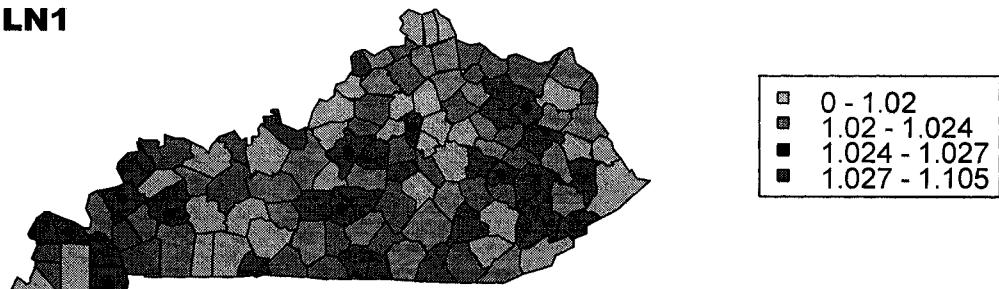
**BC**



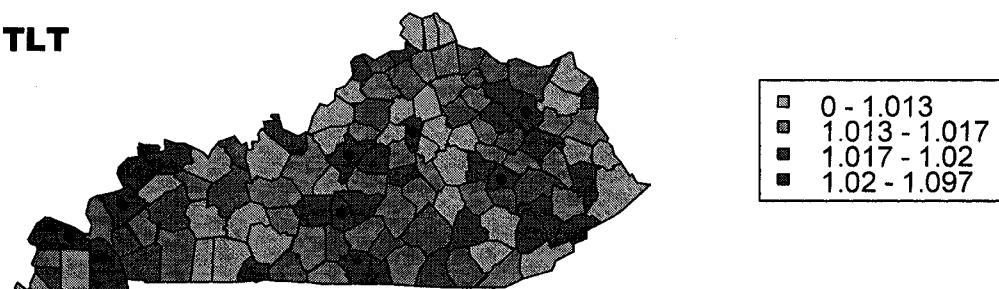
■	0 - 1.017
■	1.017 - 1.018
■	1.018 - 1.021
■	1.021 - 1.059

**200**

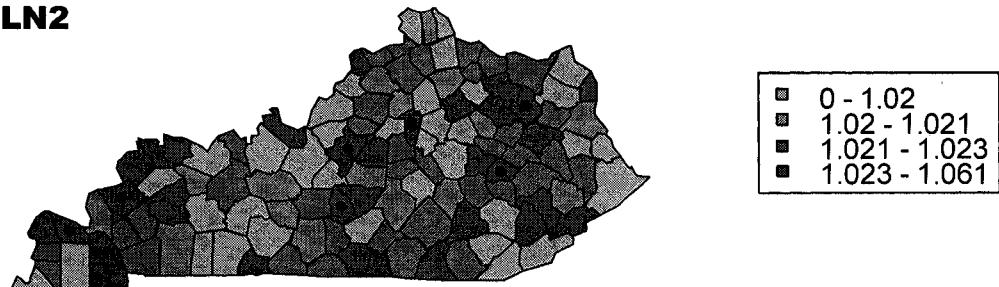
**LN1**



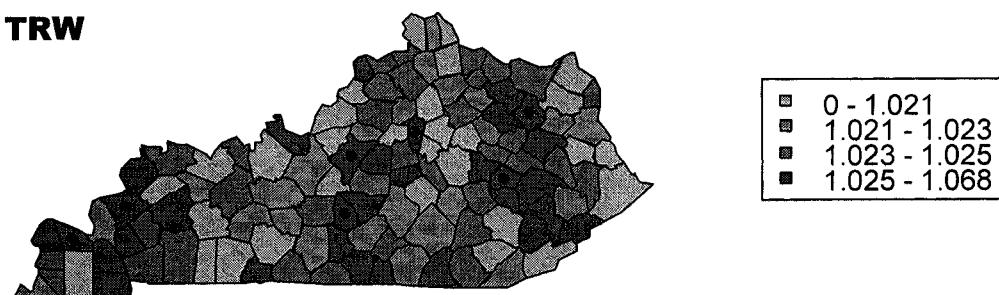
**TLT**



**LN2**



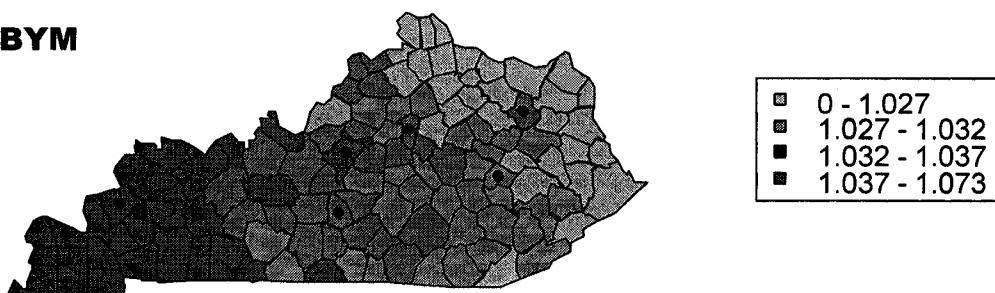
**TRW**



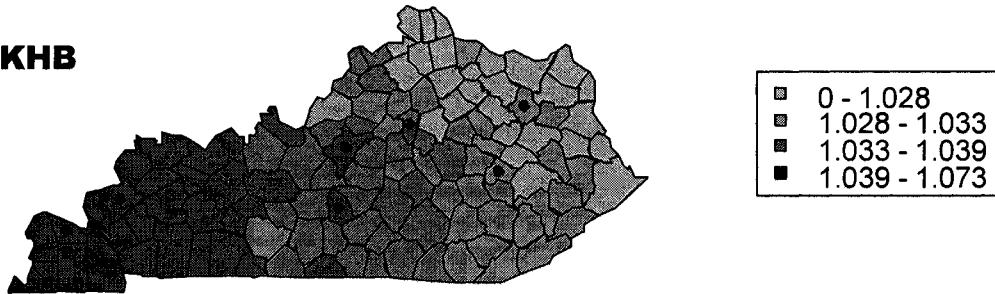
**202**

Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 1, scale factor equal to 0.1, and for the temporal pattern of Scenario 1

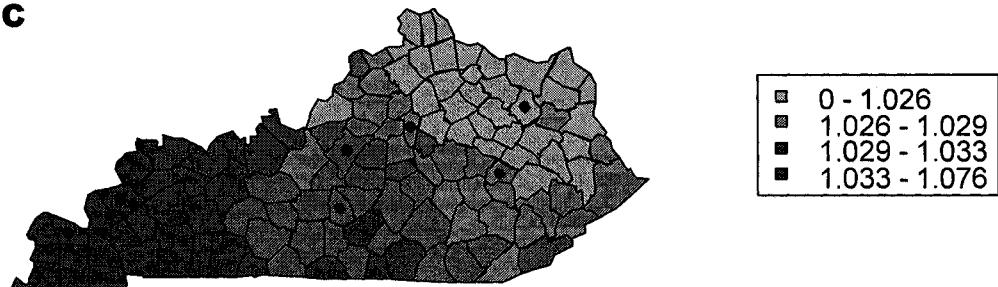
**BYM**



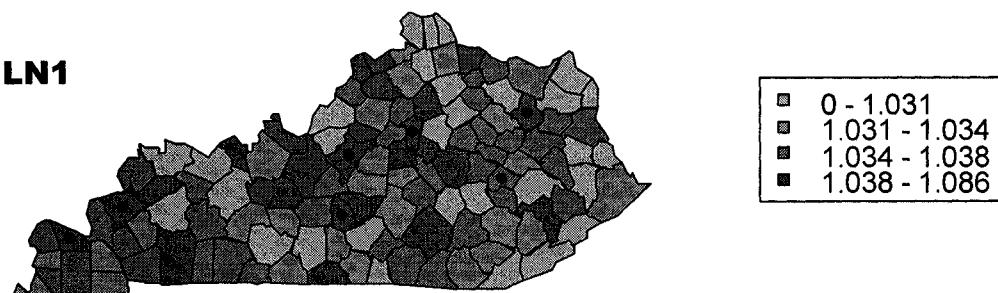
**KHB**



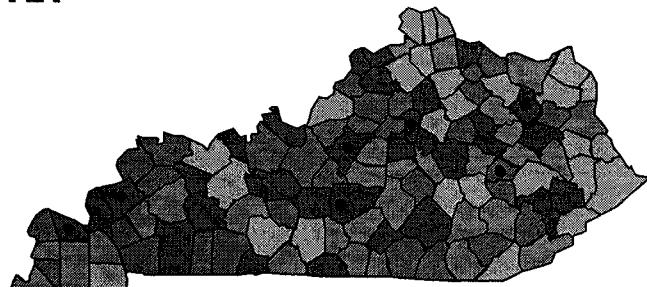
**BC**



**LN1**

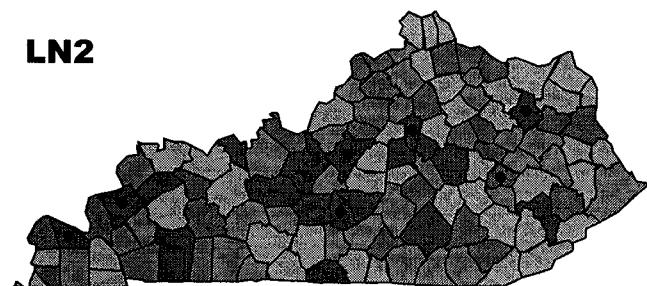


**TLT**



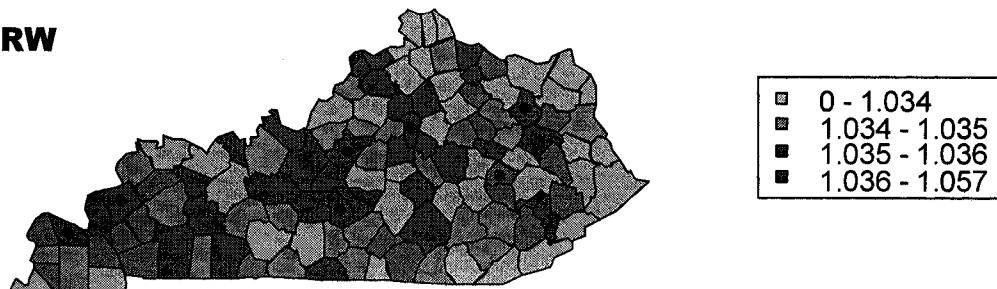
■	0 - 1.028
■	1.028 - 1.032
■	1.032 - 1.036
■	1.036 - 1.088

**LN2**



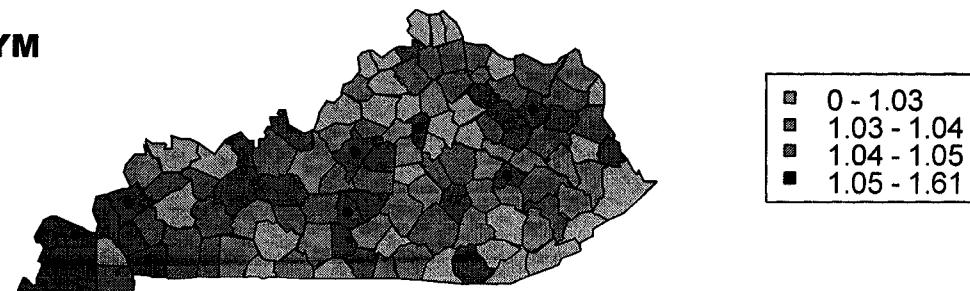
■	0 - 1.031
■	1.031 - 1.032
■	1.032 - 1.033
■	1.033 - 1.048

**TRW**

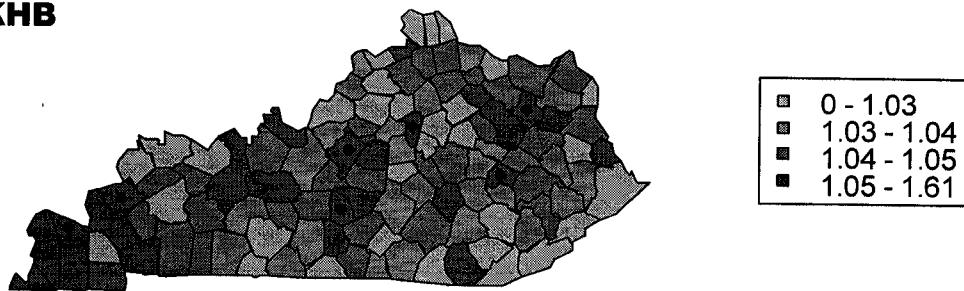


Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 2, scale factor equal to 1, and for the temporal pattern of Scenario 1

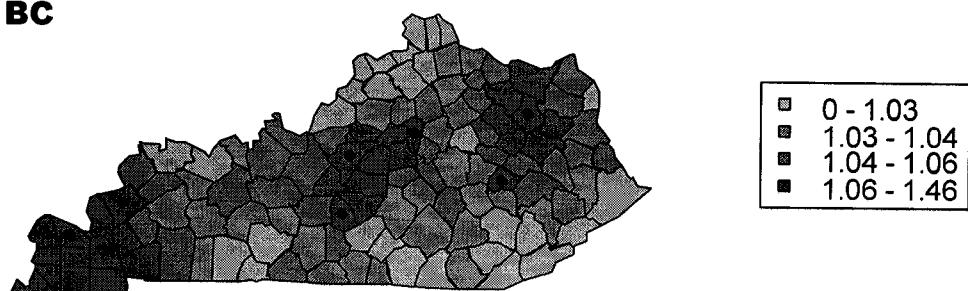
**BYM**



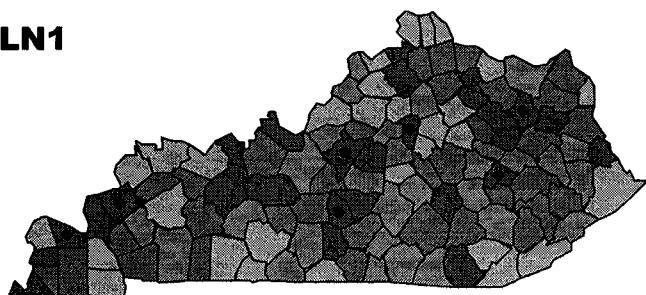
**KHB**



**BC**

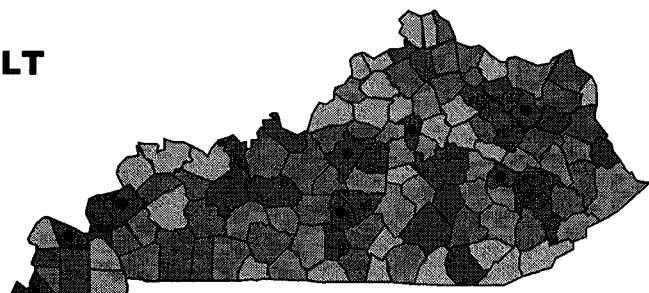


**LN1**



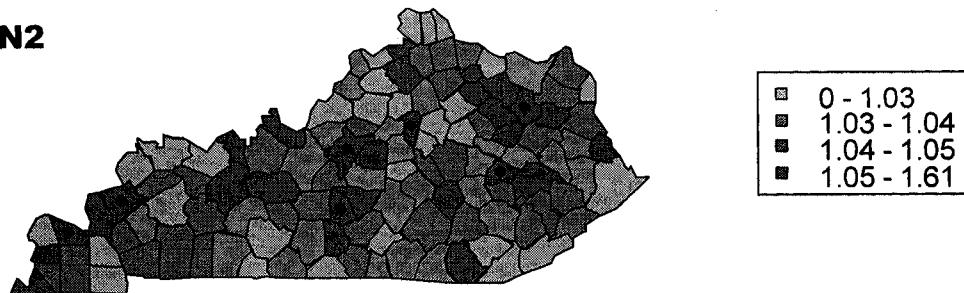
■	0 - 1.03
■	1.03 - 1.04
■	1.04 - 1.05
■	1.05 - 1.64

**TLT**

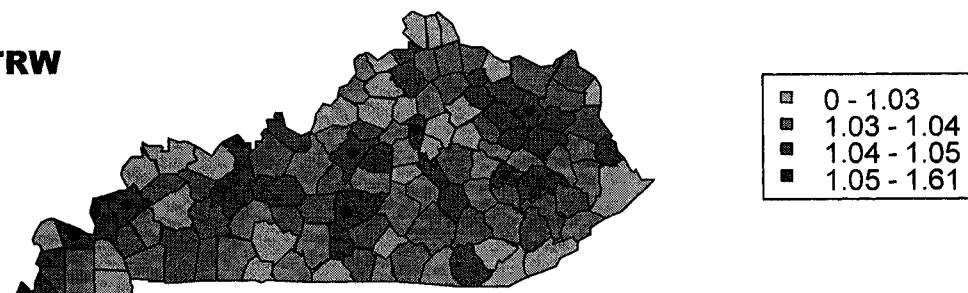


■	0 - 1.006
■	1.006 - 1.015
■	1.015 - 1.025
■	1.025 - 1.64

**LN2**

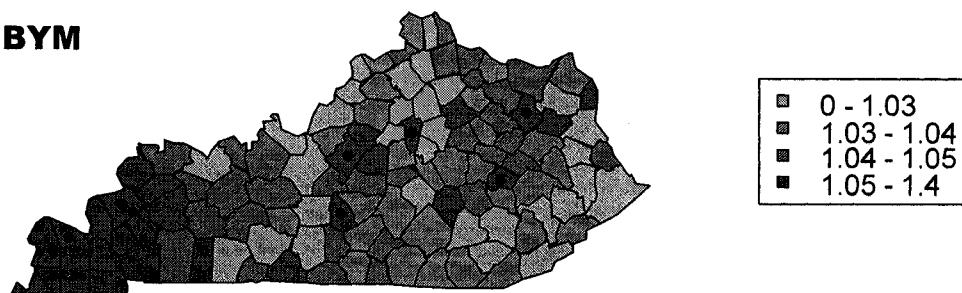


**TRW**

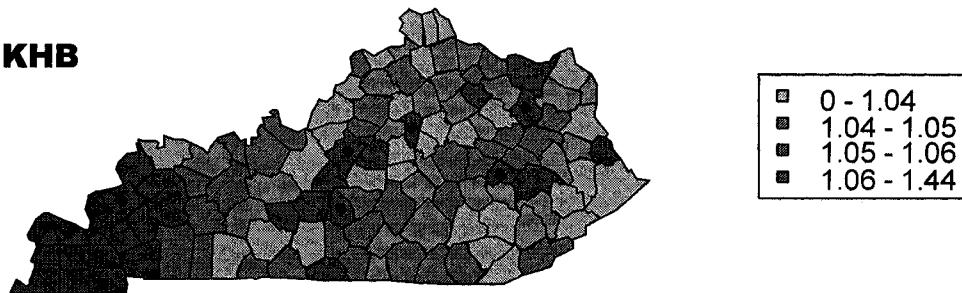


Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 2, scale factor equal to 0.5, and for the temporal pattern of Scenario 1

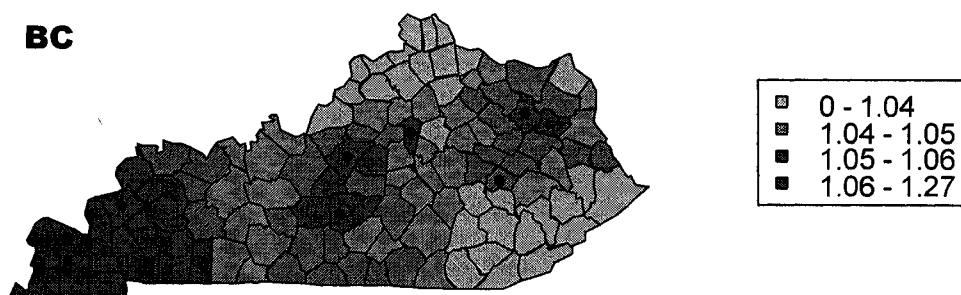
**BYM**



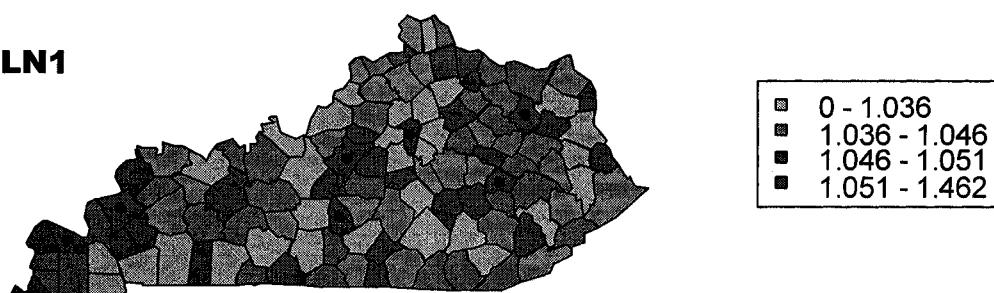
**KHB**



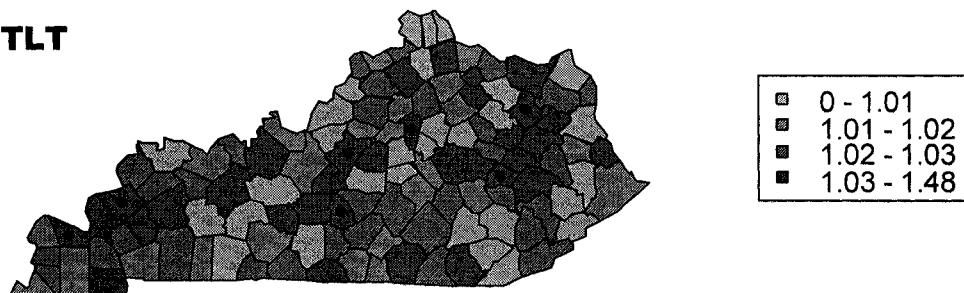
**BC**



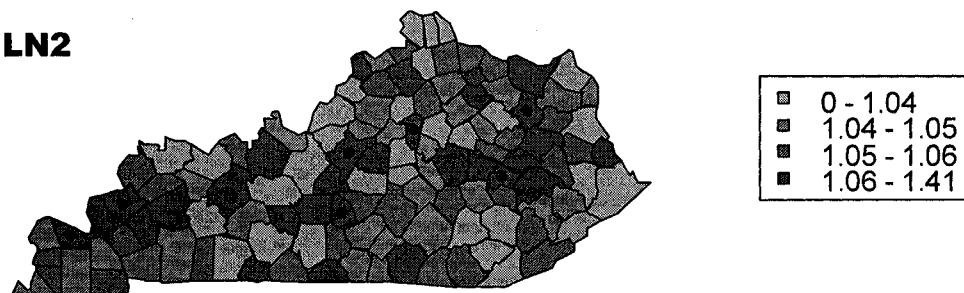
**LN1**



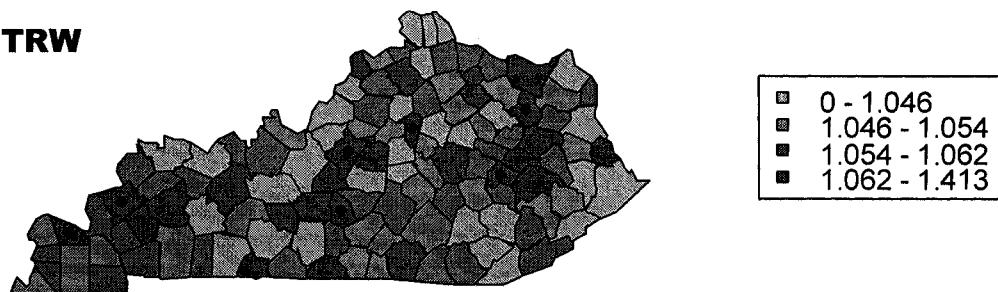
**TLT**



**LN2**

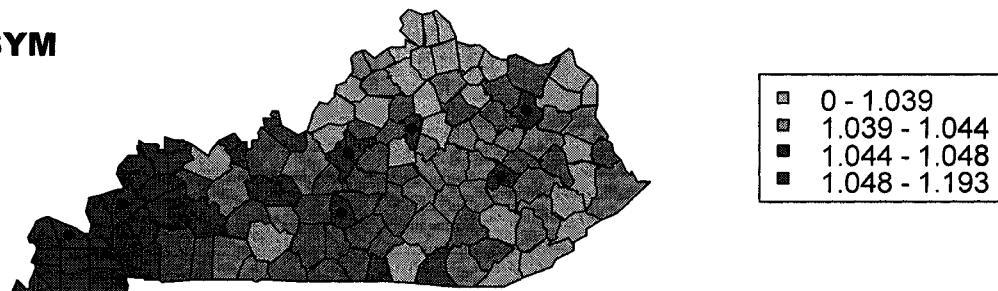


**TRW**

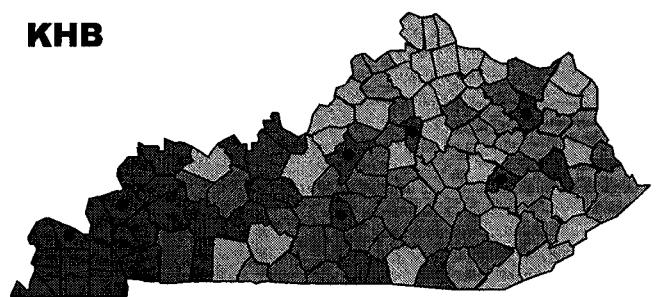


Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 2, scale factor equal to 0.3, and for the temporal pattern of Scenario 1

**BYM**

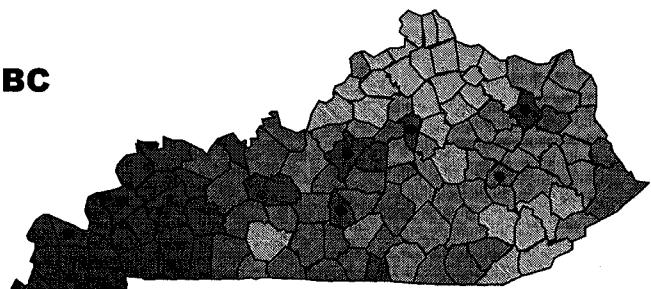


**KHB**

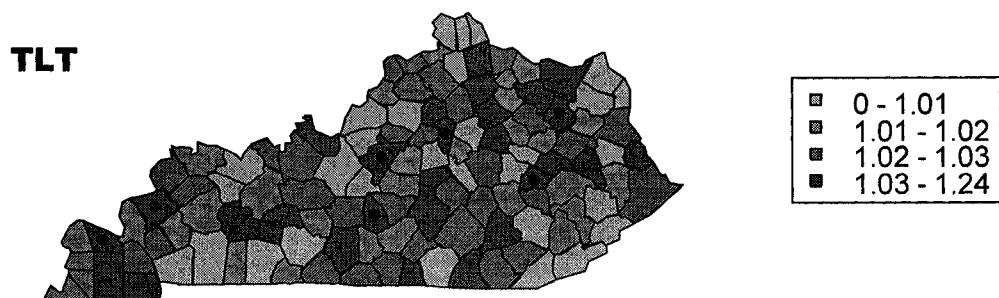
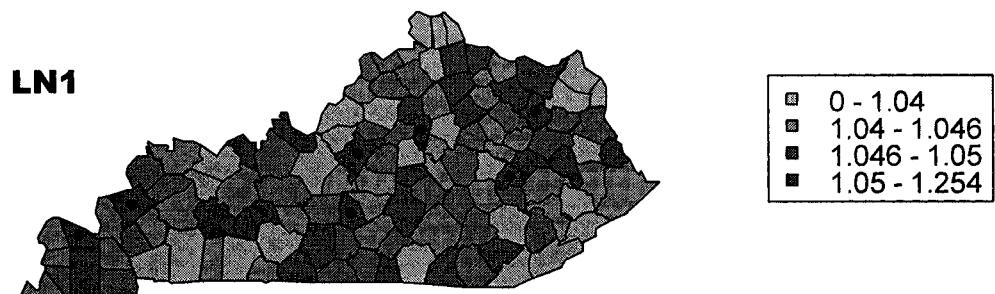


■	0 - 1.039
■	1.039 - 1.043
■	1.043 - 1.047
■	1.047 - 1.191

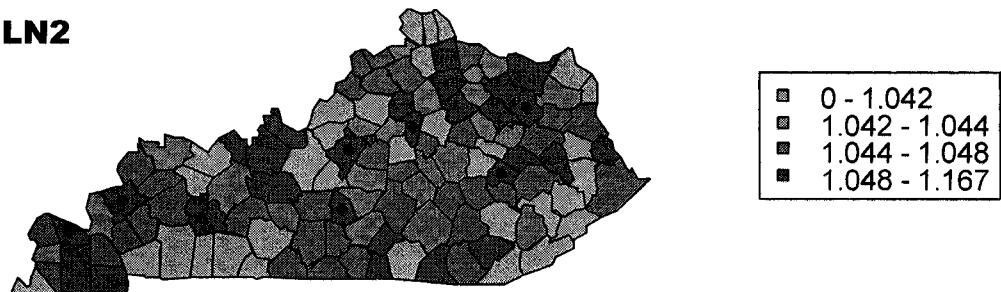
**BC**



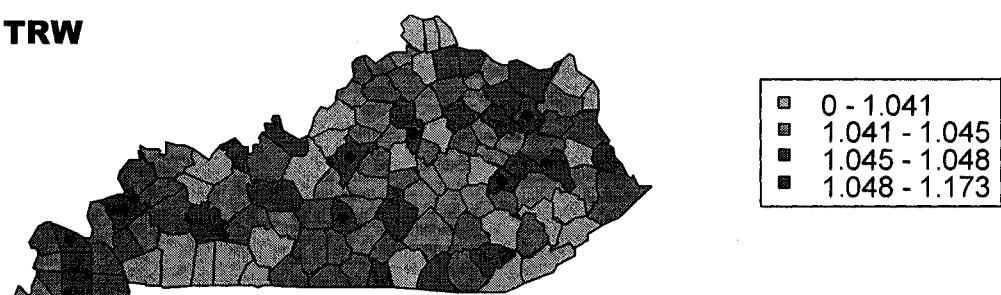
■	0 - 1.038
■	1.038 - 1.041
■	1.041 - 1.044
■	1.044 - 1.116



**LN2**

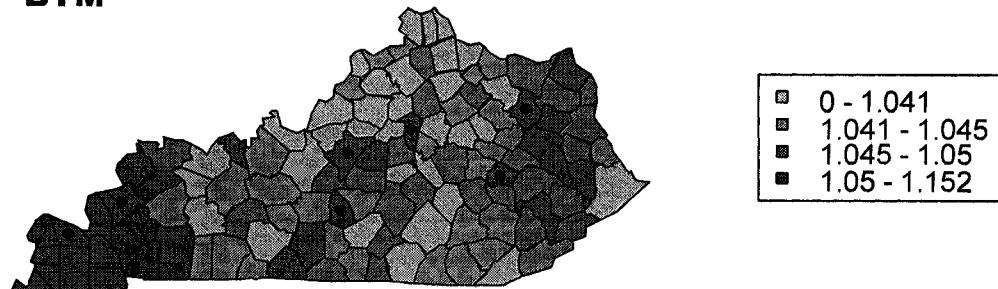


**TRW**

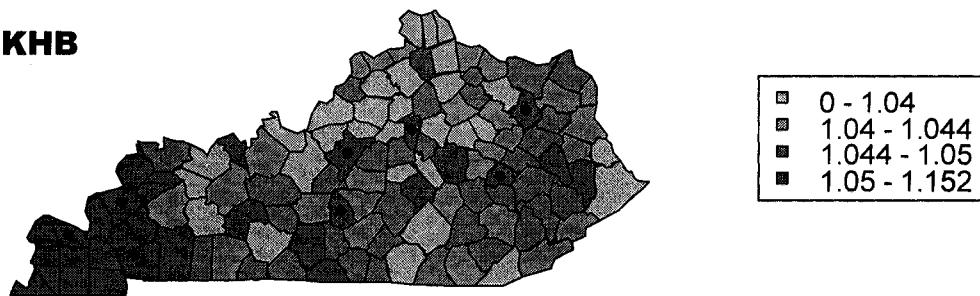


Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 2, scale factor equal to 0.1, and for the temporal pattern of Scenario 1

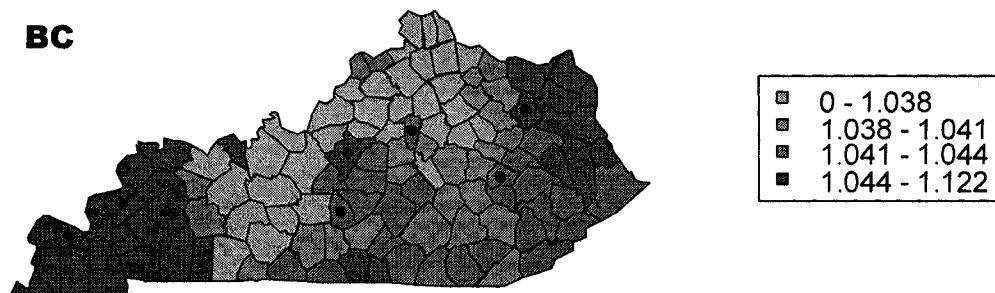
**BYM**



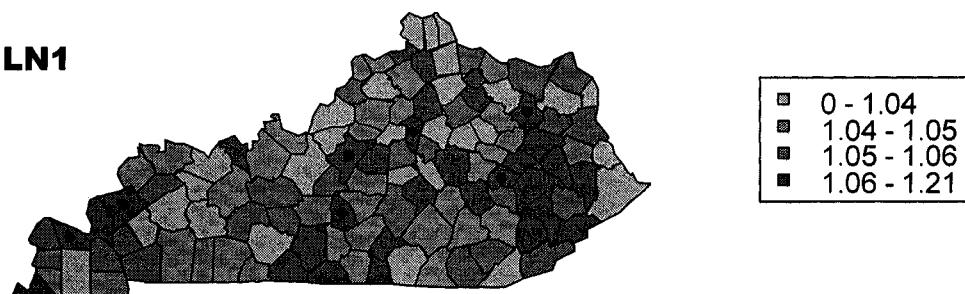
**KHB**



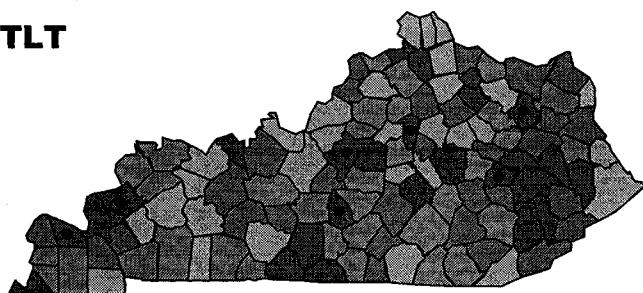
**BC**



**LN1**

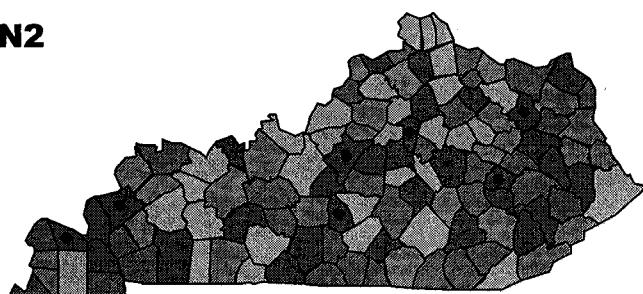


**TLT**



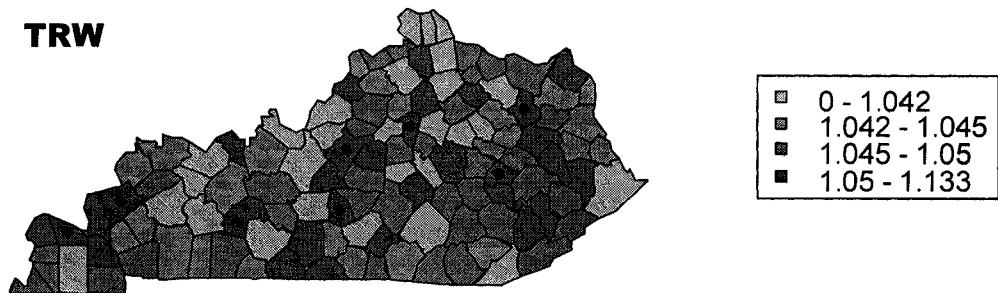
■	0 - 1.031
■	1.031 - 1.038
■	1.038 - 1.043
■	1.043 - 1.192

**LN2**



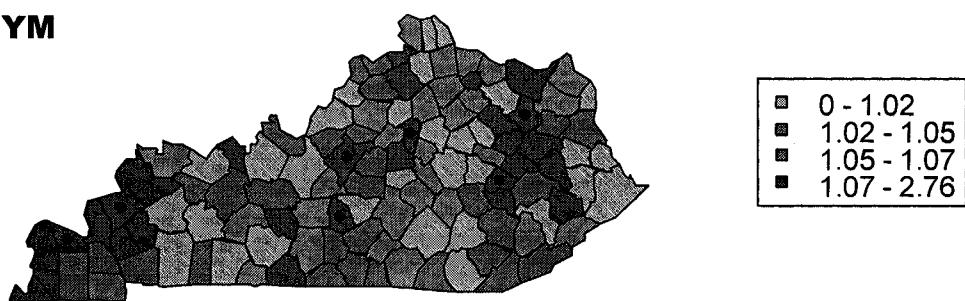
■	0 - 1.041
■	1.041 - 1.044
■	1.044 - 1.047
■	1.047 - 1.122

**TRW**

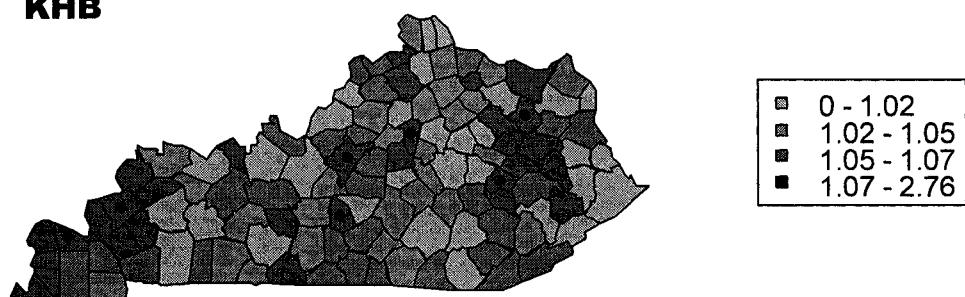


Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 3, scale factor equal to 1, and for the temporal pattern of Scenario 1

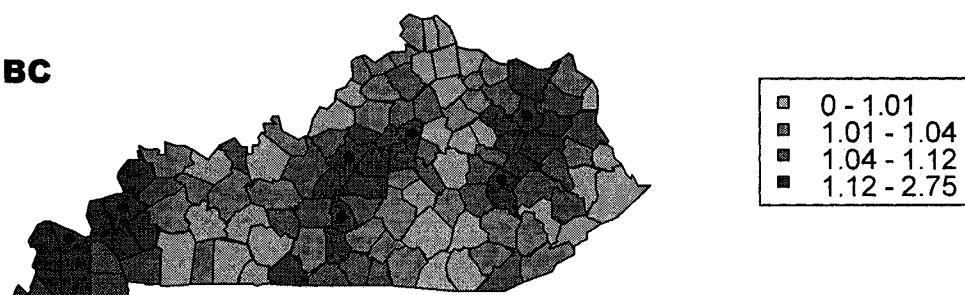
**BYM**



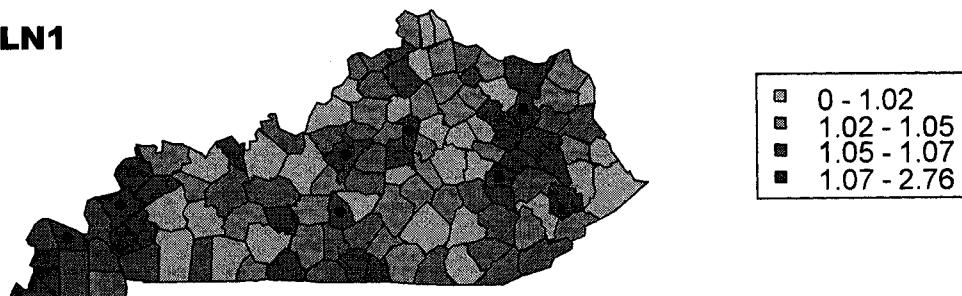
**KHB**



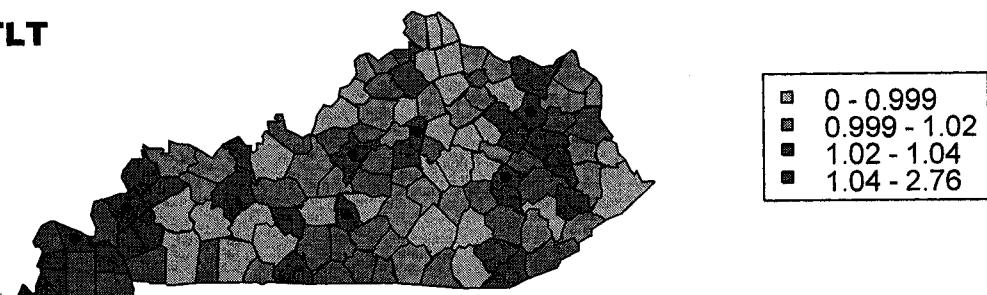
**BC**



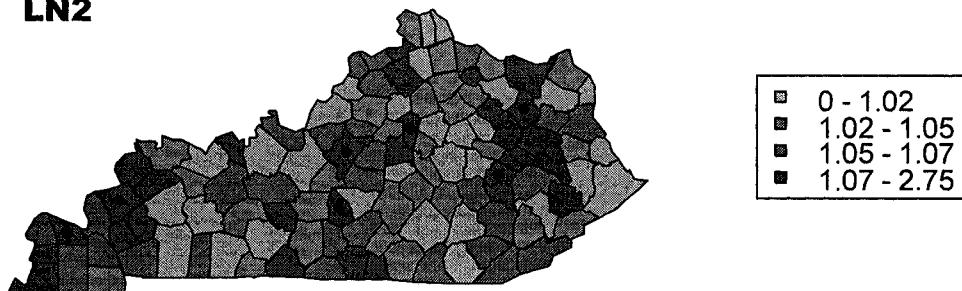
**LN1**



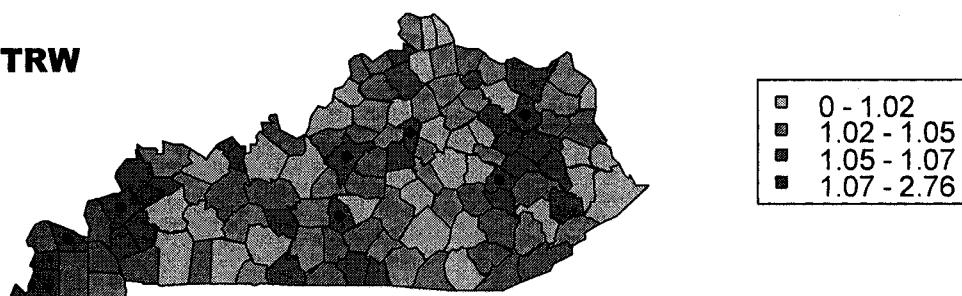
**TLT**



**LN2**

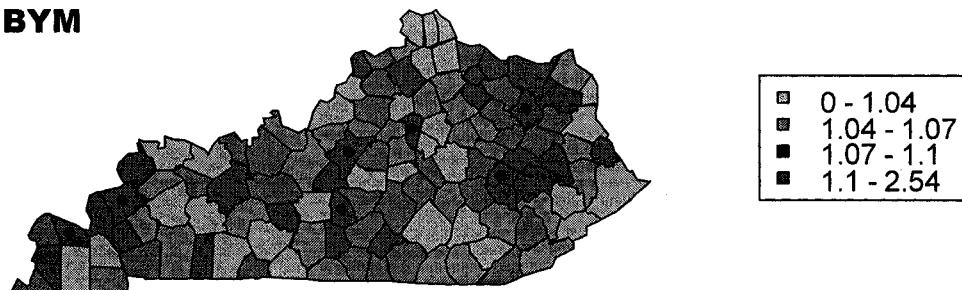


**TRW**

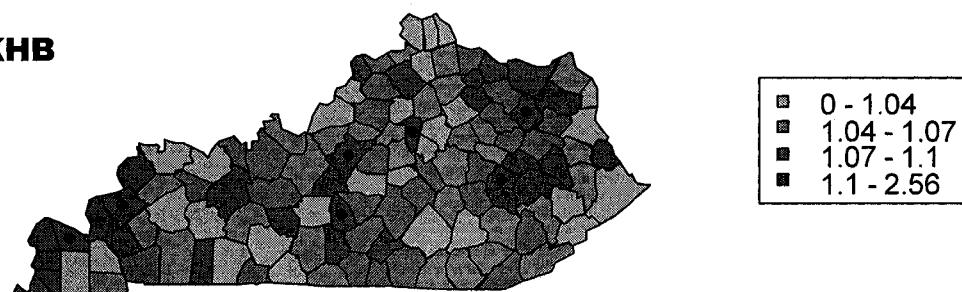


Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 3, scale factor equal to 0.5, and for the temporal pattern of Scenario 1

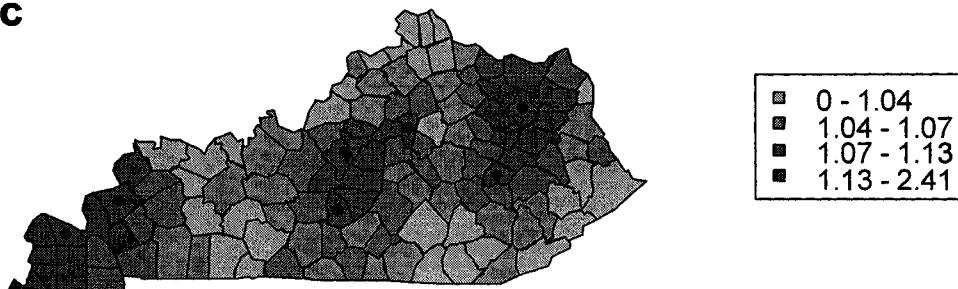
**BYM**



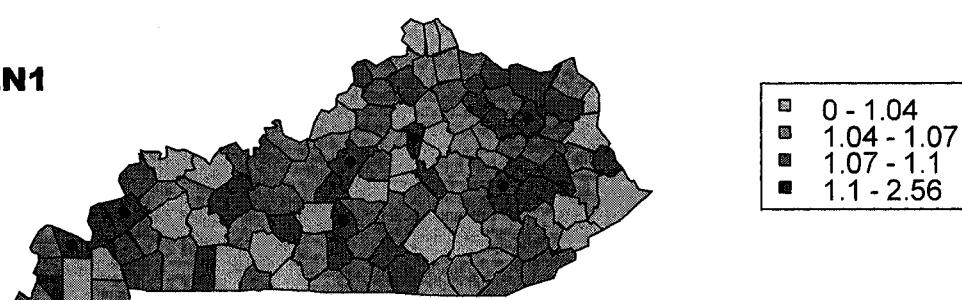
**KHB**



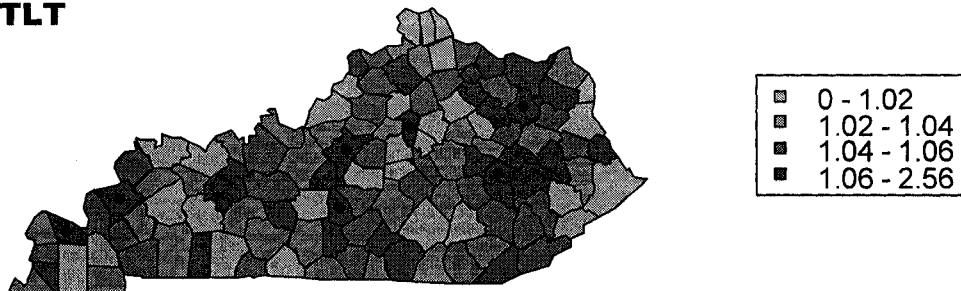
**BC**



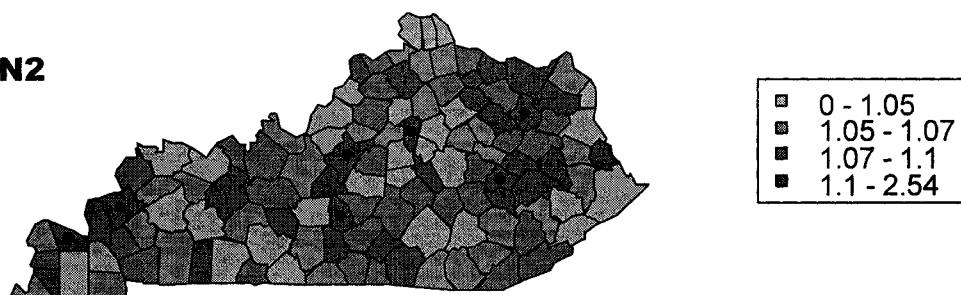
**LN1**



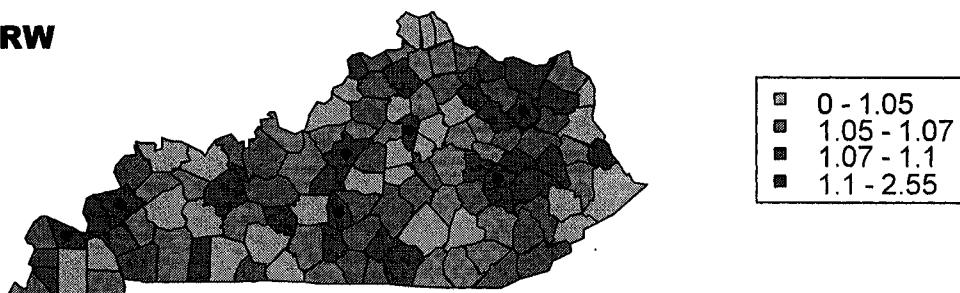
**TLT**



**LN2**

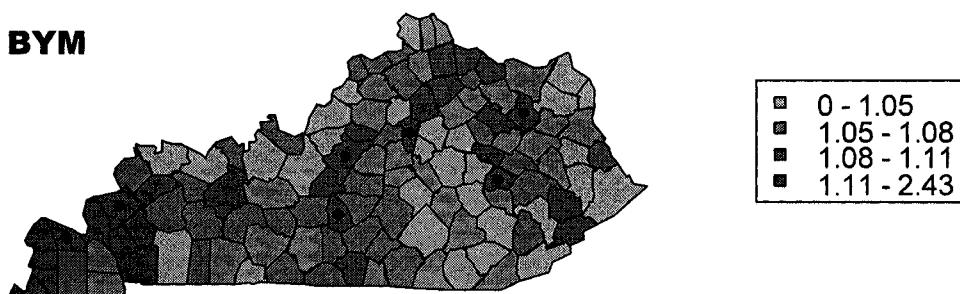


**TRW**

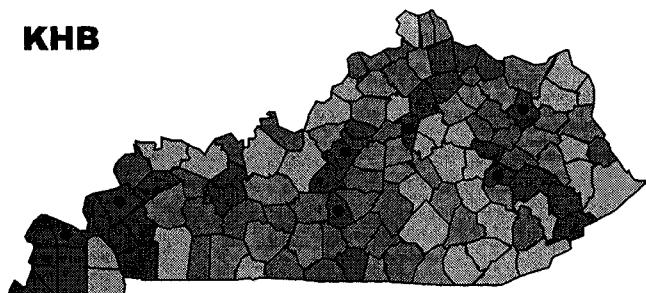


Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 3, scale factor equal to 0.3, and for the temporal pattern of Scenario 1

**BYM**

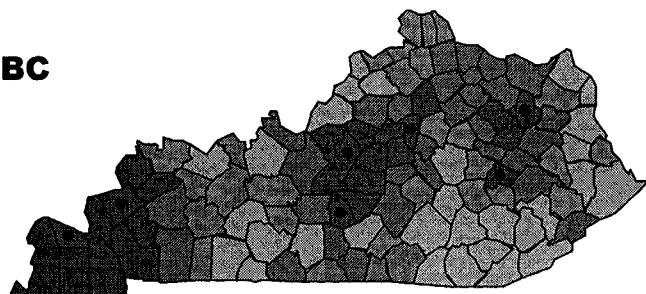


**KHB**



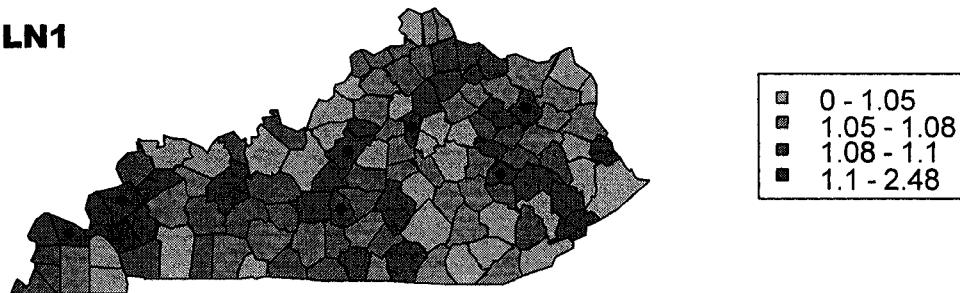
■	0 - 1.05
■	1.05 - 1.08
■	1.08 - 1.11
■	1.11 - 2.43

**BC**

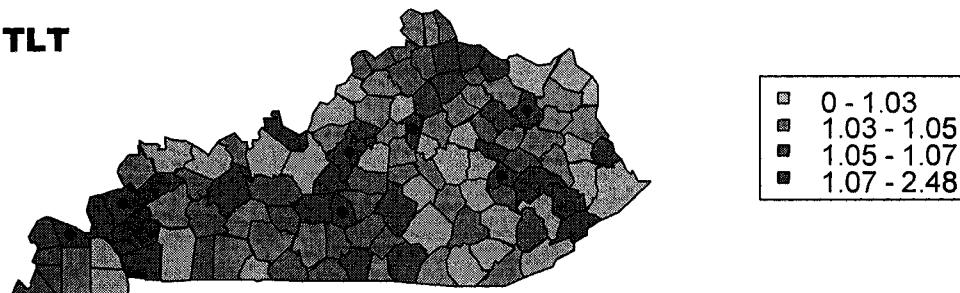


■	0 - 1.05
■	1.05 - 1.07
■	1.07 - 1.12
■	1.12 - 2.12

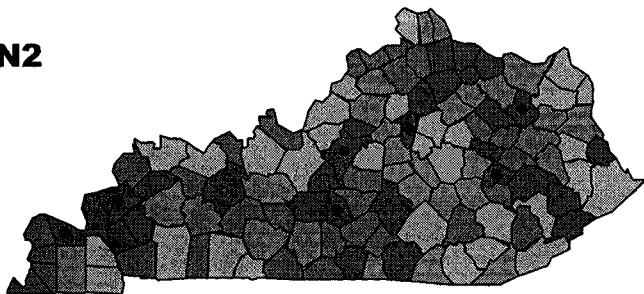
**LN1**



**TLT**

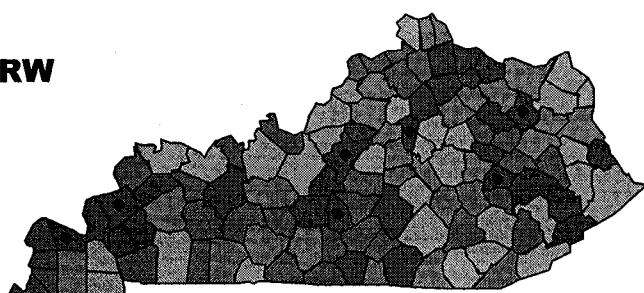


**LN2**



■	0 - 1.05
■	1.05 - 1.08
■	1.08 - 1.1
■	1.1 - 2.43

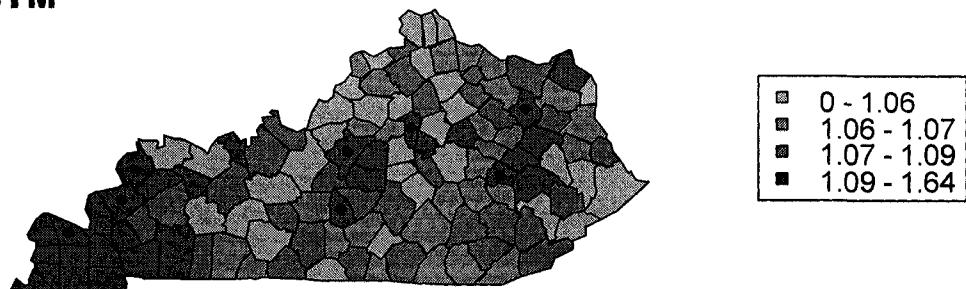
**TRW**



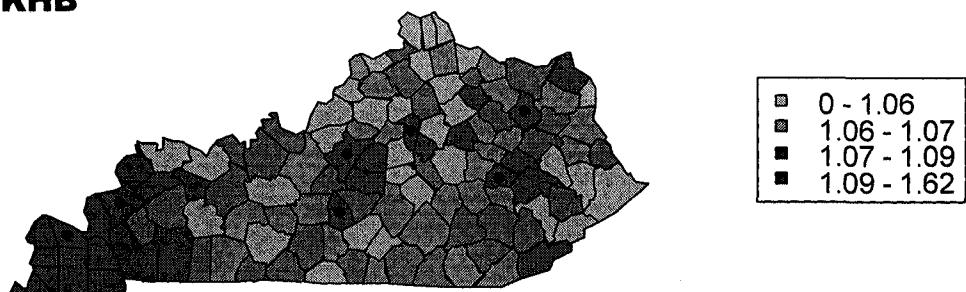
■	0 - 1.05
■	1.05 - 1.08
■	1.08 - 1.1
■	1.1 - 2.43

Disease maps of the posterior mean relative risk for simulations with true hot spot relative risk equal to 3, scale factor equal to 0.1, and for the temporal pattern of Scenario 1

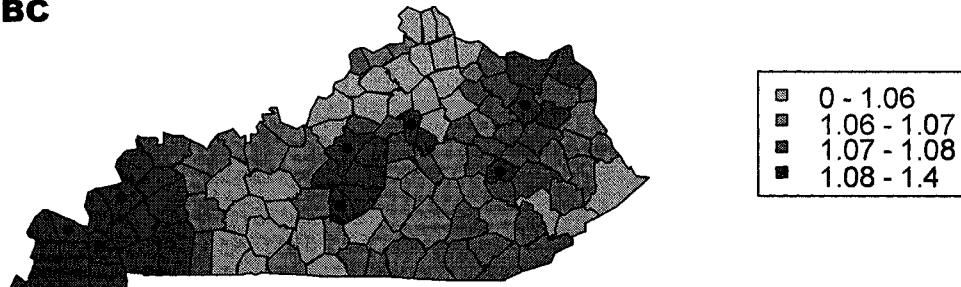
**BYM**



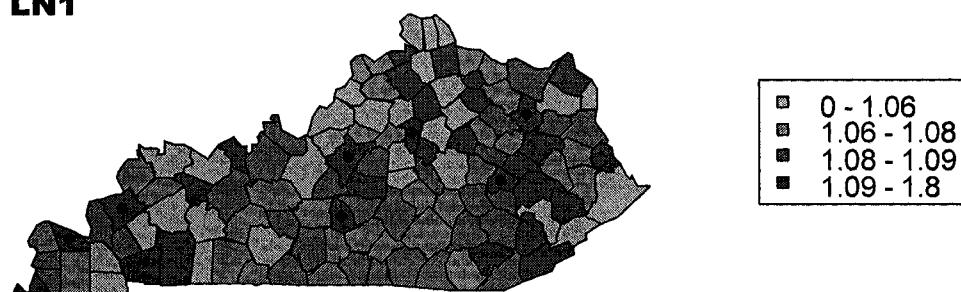
**KHB**



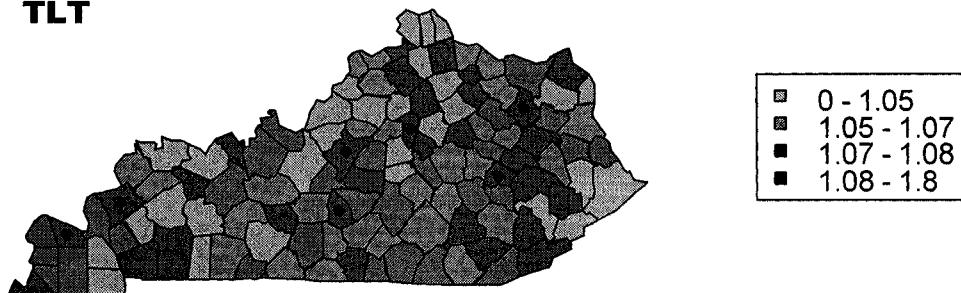
**BC**



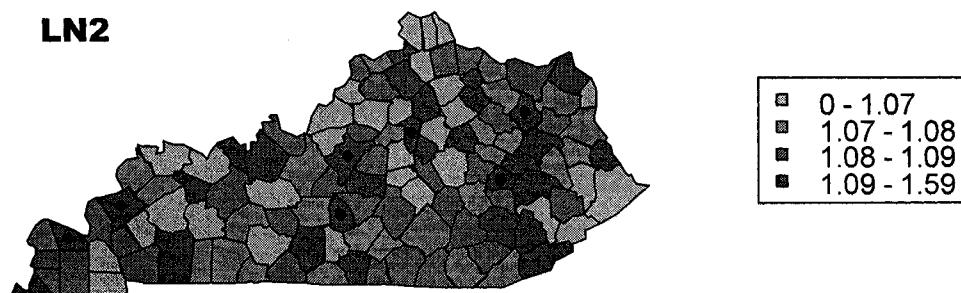
**LN1**



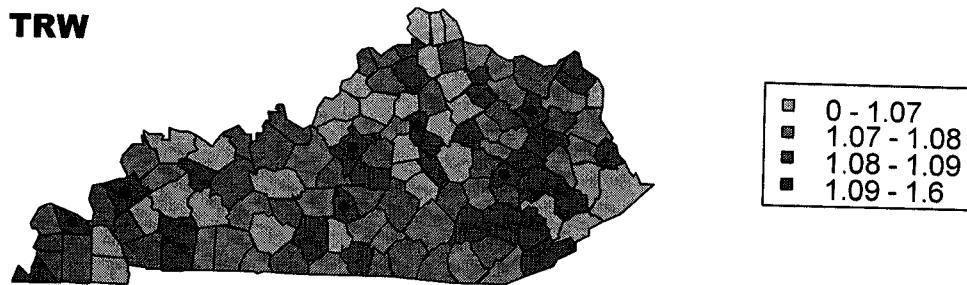
**TLT**



**LN2**



**TRW**



#### A4. Confidence Limits for differences in AUCs

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 1.5, scale factor equal to 0.5, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0392	0.0621	-0.0797	-0.1774	-0.0106	-0.0265
	<b>97.5%</b>	0	0.0196	0.1243	-0.0227	-0.1161	0.0493	0.0328
<b>KHB</b>	<b>2.5%</b>	-0.0196	0	0.0721	-0.0697	-0.1669	-0.0006	-0.0165
	<b>97.5%</b>	0.0392	0	0.1339	-0.0131	-0.1069	0.0589	0.0424
<b>BC</b>	<b>2.5%</b>	-0.1243	-0.1339	0	-0.1759	-0.2751	-0.1053	-0.1212
	<b>97.5%</b>	-0.0621	-0.0721	0	-0.1129	-0.2047	-0.0424	-0.0589
<b>LN1</b>	<b>2.5%</b>	0.0227	0.0131	0.1129	0	-0.1232	0.0417	0.0258
	<b>97.5%</b>	0.0797	0.0697	0.1759	0	-0.0678	0.0995	0.083
<b>TLT</b>	<b>2.5%</b>	0.1161	0.1069	0.2047	0.0678	0	0.1343	0.1191
	<b>97.5%</b>	0.1774	0.1669	0.2751	0.1232	0	0.1979	0.1807
<b>LN2</b>	<b>2.5%</b>	-0.0493	-0.0589	0.0424	-0.0995	-0.1979	0	-0.0462
	<b>97.5%</b>	0.0106	0.0006	0.1053	-0.0417	-0.1343	0	0.0138
<b>TRW</b>	<b>2.5%</b>	-0.0328	-0.0424	0.0589	-0.083	-0.1807	-0.0138	0
	<b>97.5%</b>	0.0265	0.0165	0.1212	-0.0258	-0.1191	0.0462	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 1.5, scale factor equal to 0.3, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0383	-0.001	-0.0687	-0.1553	-0.0162	-0.0285
	<b>97.5%</b>	0	0.0239	0.062	-0.0075	-0.086	0.0467	0.0341
<b>KHB</b>	<b>2.5%</b>	-0.0239	0	0.006	-0.0614	-0.1477	-0.0089	-0.0212
	<b>97.5%</b>	0.0383	0	0.0691	-0.0005	-0.0793	0.0538	0.0412
<b>BC</b>	<b>2.5%</b>	-0.062	-0.0691	0	-0.1003	-0.1882	-0.047	-0.0592
	<b>97.5%</b>	0.0013	-0.006	0	-0.0366	-0.1138	0.0168	0.0042
<b>LN1</b>	<b>2.5%</b>	0.0075	0.0005	0.0366	0	-0.1144	0.0226	0.0103
	<b>97.5%</b>	0.0687	0.0614	0.1003	0	-0.0507	0.0842	0.0716
<b>TLT</b>	<b>2.5%</b>	0.086	0.0793	0.1138	0.0507	0	0.1003	0.0887
	<b>97.5%</b>	0.1553	0.1477	0.1882	0.1144	0	0.1717	0.1583
<b>LN2</b>	<b>2.5%</b>	-0.0467	-0.0538	-0.017	-0.0842	-0.1717	0	-0.044
	<b>97.5%</b>	0.0162	0.0089	0.047	-0.0226	-0.1003	0	0.019
<b>TRW</b>	<b>2.5%</b>	-0.0341	-0.0412	-0.004	-0.0716	-0.1583	-0.019	0
	<b>97.5%</b>	0.0285	0.0212	0.0592	-0.0103	-0.0887	0.044	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollié (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 1.5, scale factor equal to 0.1, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0406	0.0696	-0.0794	-0.1536	-0.0191	-0.0379
	<b>97.5%</b>	0	0.0257	0.1369	-0.014	-0.0844	0.0476	0.0284
<b>KHB</b>	<b>2.5%</b>	-0.0257	0	0.0768	-0.0719	-0.1457	-0.0117	-0.0305
	<b>97.5%</b>	0.0406	0	0.1445	-0.0067	-0.0774	0.0549	0.0358
<b>BC</b>	<b>2.5%</b>	-0.1369	-0.1445	0	-0.1853	-0.2609	-0.1236	-0.1429
	<b>97.5%</b>	-0.0696	-0.0768	0	-0.1146	-0.1837	-0.0546	-0.0731
<b>LN1</b>	<b>2.5%</b>	0.014	0.0067	0.1146	0	-0.1046	0.0281	0.0093
	<b>97.5%</b>	0.0794	0.0719	0.1853	0	-0.04	0.0937	0.0745
<b>TLT</b>	<b>2.5%</b>	0.0844	0.0774	0.1837	0.04	0	0.0979	0.0801
	<b>97.5%</b>	0.1536	0.1457	0.2609	0.1046	0	0.1685	0.1484
<b>LN2</b>	<b>2.5%</b>	-0.0476	-0.0549	0.0546	-0.0937	-0.1685	0	-0.0523
	<b>97.5%</b>	0.0191	0.0117	0.1236	-0.0281	-0.0979	0	0.0143
<b>TRW</b>	<b>2.5%</b>	-0.0284	-0.0358	0.0731	-0.0745	-0.1484	-0.0143	0
	<b>97.5%</b>	0.0379	0.0305	0.1429	-0.0093	-0.0801	0.0523	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 2, scale factor equal to 1, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0105	0.0925	-0.0213	-0.0267	-0.0141	-0.0162
	<b>97.5%</b>	0	0.0091	0.1253	-0.0027	-0.0086	0.0051	0.0029
<b>KHB</b>	<b>2.5%</b>	-0.0091	0	0.0932	-0.0206	-0.026	-0.0134	-0.0155
	<b>97.5%</b>	0.0105	0	0.126	-0.0021	-0.0079	0.0057	0.0035
<b>BC</b>	<b>2.5%</b>	-0.1253	-0.126	0	-0.1377	-0.1437	-0.1301	-0.1323
	<b>97.5%</b>	-0.0925	-0.0932	0	-0.1041	-0.1094	-0.0968	-0.0988
<b>LN1</b>	<b>2.5%</b>	0.0027	0.0021	0.1041	0	-0.014	-0.0015	-0.0036
	<b>97.5%</b>	0.0213	0.0206	0.1377	0	0.0028	0.0166	0.0143
<b>TLT</b>	<b>2.5%</b>	0.0086	0.0079	0.1094	-0.0028	0	0.0043	0.0023
	<b>97.5%</b>	0.0267	0.026	0.1437	0.014	0	0.0219	0.0197
<b>LN2</b>	<b>2.5%</b>	-0.0051	-0.0057	0.0968	-0.0166	-0.0219	0	-0.0114
	<b>97.5%</b>	0.0141	0.0134	0.1301	0.0015	-0.0043	0	0.0072
<b>TRW</b>	<b>2.5%</b>	-0.0029	-0.0035	0.0988	-0.0143	-0.0197	-0.0072	0
	<b>97.5%</b>	0.0162	0.0155	0.1323	0.0036	-0.0023	0.0114	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 2, scale factor equal to 0.5, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.025	0.1017	-0.0687	-0.1122	-0.0286	-0.0294
	<b>97.5%</b>	0	0.0278	0.1662	-0.0315	-0.0682	0.024	0.023
<b>KHB</b>	<b>2.5%</b>	-0.0278	0	0.1058	-0.0755	-0.1105	-0.0238	-0.0247
	<b>97.5%</b>	0.025	0	0.1593	-0.0274	-0.0728	0.0165	0.0156
<b>BC</b>	<b>2.5%</b>	-0.1662	-0.1593	0	-0.2144	-0.2512	-0.1624	-0.1633
	<b>97.5%</b>	-0.1017	-0.1058	0	-0.1536	-0.1971	-0.11	-0.1109
<b>LN1</b>	<b>2.5%</b>	0.0315	0.0274	0.1536	0	-0.0593	0.024	0.0232
	<b>97.5%</b>	0.0687	0.0755	0.2144	0	-0.021	0.0716	0.0706
<b>TLT</b>	<b>2.5%</b>	0.0682	0.0728	0.1971	0.021	0	0.0692	0.0684
	<b>97.5%</b>	0.1122	0.1105	0.2512	0.0593	0	0.1067	0.1057
<b>LN2</b>	<b>2.5%</b>	-0.024	-0.0165	0.11	-0.0716	-0.1067	0	-0.0209
	<b>97.5%</b>	0.0286	0.0238	0.1624	-0.024	-0.0692	0	0.0191
<b>TRW</b>	<b>2.5%</b>	-0.023	-0.0156	0.1109	-0.0706	-0.1057	-0.0191	0
	<b>97.5%</b>	0.0294	0.0247	0.1633	-0.0232	-0.0684	0.0209	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 2, scale factor equal to 0.3, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0271	0.0631	-0.0985	-0.1967	-0.0165	-0.023
	<b>97.5%</b>	0	0.0296	0.1234	-0.0451	-0.1398	0.0408	0.0339
<b>KHB</b>	<b>2.5%</b>	-0.0296	0	0.0618	-0.0998	-0.198	-0.0177	-0.0243
	<b>97.5%</b>	0.0271	0	0.1222	-0.0463	-0.141	0.0396	0.0327
<b>BC</b>	<b>2.5%</b>	-0.1234	-0.1222	0	-0.1954	-0.2944	-0.1115	-0.1181
	<b>97.5%</b>	-0.0631	-0.0618	0	-0.1347	-0.2286	-0.0507	-0.0576
<b>LN1</b>	<b>2.5%</b>	0.0451	0.0463	0.1347	0	-0.1206	0.0569	0.0504
	<b>97.5%</b>	0.0985	0.0998	0.1954	0	-0.0723	0.111	0.1041
<b>TLT</b>	<b>2.5%</b>	0.1398	0.141	0.2286	0.0723	0	0.1512	0.1449
	<b>97.5%</b>	0.1967	0.198	0.2944	0.1206	0	0.2096	0.2025
<b>LN2</b>	<b>2.5%</b>	-0.0408	-0.0396	0.0507	-0.111	-0.2096	0	-0.0355
	<b>97.5%</b>	0.0165	0.0177	0.1115	-0.0569	-0.1512	0	0.022
<b>TRW</b>	<b>2.5%</b>	-0.0339	-0.0327	0.0576	-0.1041	-0.2025	-0.022	0
	<b>97.5%</b>	0.023	0.0243	0.1181	-0.0504	-0.1449	0.0355	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollié (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 2, scale factor equal to 0.1, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0284	0.0047	-0.0747	-0.2024	-0.0186	-0.0241
	<b>97.5%</b>	0	0.035	0.0689	-0.0128	-0.1331	0.0451	0.0394
<b>KHB</b>	<b>2.5%</b>	-0.035	0	0.0013	-0.078	-0.2059	-0.0219	-0.0274
	<b>97.5%</b>	0.0284	0	0.0656	-0.016	-0.1363	0.0418	0.0362
<b>BC</b>	<b>2.5%</b>	-0.0689	-0.0656	0	-0.112	-0.2411	-0.0558	-0.0613
	<b>97.5%</b>	-0.0047	-0.0013	0	-0.049	-0.168	0.0088	0.0031
<b>LN1</b>	<b>2.5%</b>	0.0128	0.016	0.049	0	-0.156	0.0258	0.0203
	<b>97.5%</b>	0.0747	0.078	0.112	0	-0.0921	0.0881	0.0825
<b>TLT</b>	<b>2.5%</b>	0.1331	0.1363	0.168	0.0921	0	0.1455	0.1403
	<b>97.5%</b>	0.2024	0.2059	0.2411	0.156	0	0.2165	0.2106
<b>LN2</b>	<b>2.5%</b>	-0.0451	-0.0418	-0.009	-0.0881	-0.2165	0	-0.0375
	<b>97.5%</b>	0.0186	0.0219	0.0558	-0.0258	-0.1455	0	0.0263
<b>TRW</b>	<b>2.5%</b>	-0.0394	-0.0362	-0.003	-0.0825	-0.2106	-0.0263	0
	<b>97.5%</b>	0.0241	0.0274	0.0613	-0.0203	-0.1403	0.0375	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 3, scale factor equal to 1, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.003	0.0066	-0.0033	-0.0033	-0.003	-0.0033
	<b>97.5%</b>	0	0.0032	0.0143	0.003	0.0029	0.0032	0.0029
<b>KHB</b>	<b>2.5%</b>	-0.0032	0	0.0064	-0.0034	-0.0034	-0.0031	-0.0035
	<b>97.5%</b>	0.003	0	0.0142	0.0029	0.0028	0.0032	0.0028
<b>BC</b>	<b>2.5%</b>	-0.0143	-0.0142	0	-0.0144	-0.0145	-0.0142	-0.0145
	<b>97.5%</b>	-0.0066	-0.0064	0	-0.0067	-0.0068	-0.0064	-0.0068
<b>LN1</b>	<b>2.5%</b>	-0.003	-0.0029	0.0067	0	-0.0031	-0.0028	-0.0032
	<b>97.5%</b>	0.0033	0.0034	0.0144	0	0.003	0.0034	0.003
<b>TLT</b>	<b>2.5%</b>	-0.0029	-0.0028	0.0068	-0.003	0	-0.0028	-0.0031
	<b>97.5%</b>	0.0033	0.0034	0.0145	0.0031	0	0.0034	0.0031
<b>LN2</b>	<b>2.5%</b>	-0.0032	-0.0032	0.0064	-0.0034	-0.0034	0	-0.0035
	<b>97.5%</b>	0.003	0.0031	0.0142	0.0028	0.0028	0	0.0027
<b>TRW</b>	<b>2.5%</b>	-0.0029	-0.0028	0.0068	-0.003	-0.0031	-0.0027	0
	<b>97.5%</b>	0.0033	0.0035	0.0145	0.0032	0.0031	0.0035	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollié (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 3, scale factor equal to 0.5, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0045	0.0305	-0.0056	-0.0053	-0.0057	-0.0057
	<b>97.5%</b>	0	0.0041	0.0448	0.0029	0.0032	0.0029	0.0028
<b>KHB</b>	<b>2.5%</b>	-0.0041	0	0.0307	-0.0054	-0.0051	-0.0055	-0.0055
	<b>97.5%</b>	0.0045	0	0.045	0.0031	0.0034	0.003	0.003
<b>BC</b>	<b>2.5%</b>	-0.0448	-0.045	0	-0.0462	-0.046	-0.0462	-0.0463
	<b>97.5%</b>	-0.0305	-0.0307	0	-0.0318	-0.0314	-0.0319	-0.0319
<b>LN1</b>	<b>2.5%</b>	-0.0029	-0.0031	0.0318	0	-0.0039	-0.0043	-0.0043
	<b>97.5%</b>	0.0056	0.0054	0.0462	0	0.0045	0.0041	0.0041
<b>TLT</b>	<b>2.5%</b>	-0.0032	-0.0034	0.0314	-0.0045	0	-0.0046	-0.0046
	<b>97.5%</b>	0.0053	0.0051	0.046	0.0039	0	0.0038	0.0038
<b>LN2</b>	<b>2.5%</b>	-0.0029	-0.003	0.0319	-0.0041	-0.0038	0	-0.0042
	<b>97.5%</b>	0.0057	0.0055	0.0462	0.0043	0.0046	0	0.0041
<b>TRW</b>	<b>2.5%</b>	-0.0028	-0.003	0.0319	-0.0041	-0.0038	-0.0041	0
	<b>97.5%</b>	0.0057	0.0055	0.0463	0.0043	0.0046	0.0042	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 3, scale factor equal to 0.3, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0105	0.076	-0.0163	-0.0221	-0.0142	-0.013
	<b>97.5%</b>	0	0.0078	0.1063	0.0015	-0.0049	0.0037	0.005
<b>KHB</b>	<b>2.5%</b>	-0.0078	0	0.0774	-0.0148	-0.0207	-0.0128	-0.0116
	<b>97.5%</b>	0.0105	0	0.1076	0.0028	-0.0036	0.005	0.0063
<b>BC</b>	<b>2.5%</b>	-0.1063	-0.1076	0	-0.1138	-0.1201	-0.1116	-0.1104
	<b>97.5%</b>	-0.076	-0.0774	0	-0.0833	-0.0892	-0.0812	-0.0799
<b>LN1</b>	<b>2.5%</b>	-0.0015	-0.0028	0.0833	0	-0.0143	-0.0064	-0.0052
	<b>97.5%</b>	0.0163	0.0148	0.1138	0	0.0021	0.0108	0.0121
<b>TLT</b>	<b>2.5%</b>	0.0049	0.0036	0.0892	-0.0021	0	0	0.0011
	<b>97.5%</b>	0.0221	0.0207	0.1201	0.0143	0	0.0166	0.0179
<b>LN2</b>	<b>2.5%</b>	-0.0037	-0.005	0.0812	-0.0108	-0.0166	0	-0.0075
	<b>97.5%</b>	0.0142	0.0128	0.1116	0.0064	0	0	0.01
<b>TRW</b>	<b>2.5%</b>	-0.005	-0.0063	0.0799	-0.0121	-0.0179	-0.01	0
	<b>97.5%</b>	0.013	0.0116	0.1104	0.0052	-0.0011	0.0075	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollié (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for simulations with true hot spot relative risk equal to 3, scale factor equal to 0.1, and for the temporal pattern of Scenario 1

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0167	0.0752	-0.0977	-0.1716	-0.0169	-0.0249
	<b>97.5%</b>	0	0.0374	0.134	-0.0478	-0.1197	0.0372	0.0288
<b>KHB</b>	<b>2.5%</b>	-0.0374	0	0.0649	-0.1083	-0.1825	-0.0275	-0.0355
	<b>97.5%</b>	0.0167	0	0.1236	-0.0578	-0.1295	0.0271	0.0187
<b>BC</b>	<b>2.5%</b>	-0.134	-0.1236	0	-0.2071	-0.2817	-0.1244	-0.1325
	<b>97.5%</b>	-0.0752	-0.0649	0	-0.1476	-0.2188	-0.0645	-0.0729
<b>LN1</b>	<b>2.5%</b>	0.0478	0.0578	0.1476	0	-0.0946	0.0576	0.0496
	<b>97.5%</b>	0.0977	0.1083	0.2071	0	-0.0513	0.1081	0.0997
<b>TLT</b>	<b>2.5%</b>	0.1197	0.1295	0.2188	0.0513	0	0.1292	0.1214
	<b>97.5%</b>	0.1716	0.1825	0.2817	0.0946	0	0.1824	0.1738
<b>LN2</b>	<b>2.5%</b>	-0.0372	-0.0271	0.0645	-0.1081	-0.1824	0	-0.0353
	<b>97.5%</b>	0.0169	0.0275	0.1244	-0.0576	-0.1292	0	0.0189
<b>TRW</b>	<b>2.5%</b>	-0.0288	-0.0187	0.0729	-0.0997	-0.1738	-0.0189	0
	<b>97.5%</b>	0.0249	0.0355	0.1325	-0.0496	-0.1214	0.0353	0

Entries refer to row heading – column heading.

BYM = Besag, York and Molie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 2, Relative Risk =  
1.5, sf=1 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0291	0.0623	-0.0794	-0.1631	-0.0191	-0.0256
	<b>97.5%</b>	0	0.0374	0.1296	-0.014	-0.0939	0.0476	0.041
<b>KHB</b>	<b>2.5%</b>	-0.0374	0	0.0581	-0.0836	-0.1673	-0.0234	-0.0298
	<b>97.5%</b>	0.0291	0	0.1255	-0.0182	-0.098	0.0434	0.0369
<b>BC</b>	<b>2.5%</b>	-0.1296	-0.1255	0	-0.1779	-0.263	-0.1163	-0.1226
	<b>97.5%</b>	-0.0623	-0.0581	0	-0.1074	-0.1859	-0.0473	-0.054
<b>LN1</b>	<b>2.5%</b>	0.014	0.0182	0.1074	0	-0.1142	0.0281	0.0217
	<b>97.5%</b>	0.0794	0.0836	0.1779	0	-0.0494	0.0937	0.0871
<b>TLT</b>	<b>2.5%</b>	0.0939	0.098	0.1859	0.0494	0	0.1073	0.1009
	<b>97.5%</b>	0.1631	0.1673	0.263	0.1142	0	0.1781	0.1715
<b>LN2</b>	<b>2.5%</b>	-0.0476	-0.0434	0.0473	-0.0937	-0.1781	0	-0.0399
	<b>97.5%</b>	0.0191	0.0234	0.1163	-0.0281	-0.1073	0	0.0269
<b>TRW</b>	<b>2.5%</b>	-0.041	-0.0369	0.054	-0.0871	-0.1715	-0.0269	0
	<b>97.5%</b>	0.0256	0.0298	0.1226	-0.0217	-0.1009	0.0399	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 3, Relative Risk =  
1.5, sf=1 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0325	0.0492	-0.0794	-0.1682	-0.0191	-0.0243
	<b>97.5%</b>	0	0.034	0.1165	-0.014	-0.0958	0.0476	0.0423
<b>KHB</b>	<b>2.5%</b>	-0.034	0	0.0485	-0.0801	-0.169	-0.0199	-0.0251
	<b>97.5%</b>	0.0325	0	0.1157	-0.0147	-0.0964	0.0468	0.0415
<b>BC</b>	<b>2.5%</b>	-0.1165	-0.1157	0	-0.1639	-0.2543	-0.1024	-0.1076
	<b>97.5%</b>	-0.0492	-0.0485	0	-0.0951	-0.1754	-0.0349	-0.0401
<b>LN1</b>	<b>2.5%</b>	0.0140	0.0147	0.0951	0	-0.1191	0.0281	0.0229
	<b>97.5%</b>	0.0794	0.0801	0.1639	0	-0.0515	0.0937	0.0884
<b>TLT</b>	<b>2.5%</b>	0.0958	0.0964	0.1754	0.0515	0	0.1089	0.104
	<b>97.5%</b>	0.1682	0.169	0.2543	0.1191	0	0.1835	0.1779
<b>LN2</b>	<b>2.5%</b>	-0.0476	-0.0468	0.0349	-0.0937	-0.1835	0	-0.0387
	<b>97.5%</b>	0.0191	0.0199	0.1024	-0.0281	-0.1089	0	0.0282
<b>TRW</b>	<b>2.5%</b>	-0.0423	-0.0415	0.0401	-0.0884	-0.1779	-0.0282	0
	<b>97.5%</b>	0.0243	0.0251	0.1076	-0.0229	-0.104	0.0387	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998)  
model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with  
Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model  
with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk =  
1.5, sf=0.5 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.19	-0.0352	-0.0797	-0.2896	-0.0106	-0.1948
	<b>97.5%</b>	0	-0.1353	0.0281	-0.0227	-0.2325	0.0493	-0.1395
<b>KHB</b>	<b>2.5%</b>	0.1353	0	0.1319	0.0856	-0.1188	0.1533	-0.0253
	<b>97.5%</b>	0.19	0	0.1864	0.1373	-0.0779	0.2108	0.0163
<b>BC</b>	<b>2.5%</b>	-0.0281	-0.1864	0	-0.0789	-0.2853	-0.01	-0.1915
	<b>97.5%</b>	0.0352	-0.1319	0	-0.0165	-0.2296	0.0558	-0.1357
<b>LN1</b>	<b>2.5%</b>	0.0227	-0.1373	0.0165	0	-0.2357	0.0417	-0.1413
	<b>97.5%</b>	0.0797	-0.0856	0.0789	0	-0.1839	0.0995	-0.0905
<b>TLT</b>	<b>2.5%</b>	0.2325	0.0779	0.2296	0.1839	0	0.2509	0.0738
	<b>97.5%</b>	0.2896	0.1188	0.2853	0.2357	0	0.3099	0.114
<b>LN2</b>	<b>2.5%</b>	-0.0493	-0.2108	-0.0558	-0.0995	-0.3099	0	-0.2148
	<b>97.5%</b>	0.0106	-0.1533	0.01	-0.0417	-0.2509	0	-0.1583
<b>TRW</b>	<b>2.5%</b>	0.1395	-0.0163	0.1357	0.0905	-0.114	0.1583	0
	<b>97.5%</b>	0.1948	0.0253	0.1915	0.1413	-0.0738	0.2148	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998)  
model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with  
Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model  
with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk =  
1.5, sf=0.3 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.1004	-0.1822	-0.0687	-0.2547	-0.0162	-0.0848
	<b>97.5%</b>	0	-0.0388	-0.1228	-0.0075	-0.1914	0.0467	-0.0218
<b>KHB</b>	<b>2.5%</b>	0.0388	0	-0.1098	0.0007	-0.1823	0.053	-0.0127
	<b>97.5%</b>	0.1004	0	-0.0561	0.0621	-0.1246	0.1167	0.0453
<b>BC</b>	<b>2.5%</b>	0.1228	0.0561	0	0.085	-0.097	0.1372	0.0713
	<b>97.5%</b>	0.1822	0.1098	0	0.1438	-0.044	0.1984	0.1271
<b>LN1</b>	<b>2.5%</b>	0.0075	-0.0621	-0.1438	0	-0.2147	0.0226	-0.0456
	<b>97.5%</b>	0.0687	-0.0007	-0.085	0	-0.1551	0.0842	0.0153
<b>TLT</b>	<b>2.5%</b>	0.1914	0.1246	0.044	0.1551	0	0.2058	0.1402
	<b>97.5%</b>	0.2547	0.1823	0.097	0.2147	0	0.2708	0.1993
<b>LN2</b>	<b>2.5%</b>	-0.0467	-0.1167	-0.1984	-0.0842	-0.2708	0	-0.0998
	<b>97.5%</b>	0.0162	-0.053	-0.1372	-0.0226	-0.2058	0	-0.0373
<b>TRW</b>	<b>2.5%</b>	0.0218	-0.0453	-0.1271	-0.0153	-0.1993	0.0373	0
	<b>97.5%</b>	0.0848	0.0127	-0.0713	0.0456	-0.1402	0.0998	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, RR = 1.5, sf=0.1  
in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0525	-0.0482	-0.0469	-0.15	-0.0241	-0.0438
	<b>97.5%</b>	0	0.0129	0.0218	0.0178	-0.0781	0.0412	0.0215
<b>KHB</b>	<b>2.5%</b>	-0.0129	0	-0.0264	-0.0274	-0.1287	-0.0046	-0.0235
	<b>97.5%</b>	0.0525	0	0.0396	0.038	-0.0598	0.0613	0.0408
<b>BC</b>	<b>2.5%</b>	-0.0218	-0.0396	0	-0.0365	-0.1347	-0.0139	-0.0318
	<b>97.5%</b>	0.0482	0.0264	0	0.0339	-0.067	0.0574	0.036
<b>LN1</b>	<b>2.5%</b>	-0.0178	-0.038	-0.0339	0	-0.1346	-0.0094	-0.0293
	<b>97.5%</b>	0.0469	0.0274	0.0365	0	-0.0644	0.0555	0.036
<b>TLT</b>	<b>2.5%</b>	0.0781	0.0598	0.067	0.0644	0	0.0861	0.068
	<b>97.5%</b>	0.15	0.1287	0.1347	0.1346	0	0.1591	0.1377
<b>LN2</b>	<b>2.5%</b>	-0.0412	-0.0613	-0.0574	-0.0555	-0.1591	0	-0.0525
	<b>97.5%</b>	0.0241	0.0046	0.0139	0.0094	-0.0861	0	0.0131
<b>TRW</b>	<b>2.5%</b>	-0.0215	-0.0408	-0.036	-0.036	-0.1377	-0.0131	0
	<b>97.5%</b>	0.0438	0.0235	0.0318	0.0293	-0.068	0.0525	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998)  
model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with  
Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model  
with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk = 2,  
sf=1 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0823	-0.0622	-0.0213	-0.0823	-0.0141	-0.0824
	<b>97.5%</b>	0	-0.0649	-0.0456	-0.0027	-0.0648	0.0051	-0.065
<b>KHB</b>	<b>2.5%</b>	0.0649	0	0.016	0.0539	-0.0018	0.0607	-0.0019
	<b>97.5%</b>	0.0823	0	0.0235	0.0692	0.0018	0.0774	0.0017
<b>BC</b>	<b>2.5%</b>	0.0456	-0.0235	0	0.034	-0.0235	0.0408	-0.0236
	<b>97.5%</b>	0.0622	-0.016	0	0.0497	-0.016	0.0579	-0.0161
<b>LN1</b>	<b>2.5%</b>	0.0027	-0.0692	-0.0497	0	-0.0692	-0.0015	-0.0693
	<b>97.5%</b>	0.0213	-0.0539	-0.034	0	-0.0539	0.0166	-0.054
<b>TLT</b>	<b>2.5%</b>	0.0648	-0.0018	0.016	0.0539	0	0.0607	-0.0019
	<b>97.5%</b>	0.0823	0.0018	0.0235	0.0692	0	0.0774	0.0017
<b>LN2</b>	<b>2.5%</b>	-0.0051	-0.0774	-0.0579	-0.0166	-0.0774	0	-0.0775
	<b>97.5%</b>	0.0141	-0.0607	-0.0408	0.0015	-0.0607	0	-0.0608
<b>TRW</b>	<b>2.5%</b>	0.065	-0.0017	0.0161	0.054	-0.0017	0.0608	0
	<b>97.5%</b>	0.0824	0.0019	0.0236	0.0693	0.0019	0.0775	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollié (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk = 2,  
sf=0.5 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.1966	-0.1024	-0.0687	-0.1983	-0.0286	-0.1997
	<b>97.5%</b>	0	-0.1581	-0.0575	-0.0315	-0.1599	0.0241	-0.1613
<b>KHB</b>	<b>2.5%</b>	0.1581	0	0.0847	0.1115	-0.0065	0.1571	-0.0078
	<b>97.5%</b>	0.1966	0	0.1101	0.1431	0.003	0.1931	0.0015
<b>BC</b>	<b>2.5%</b>	0.0575	-0.1101	0	0.0101	-0.112	0.058	-0.1134
	<b>97.5%</b>	0.1024	-0.0847	0	0.0497	-0.0862	0.0973	-0.0877
<b>LN1</b>	<b>2.5%</b>	0.0315	-0.1431	-0.0497	0	-0.1448	0.024	-0.1462
	<b>97.5%</b>	0.0687	-0.1115	-0.0101	0	-0.1132	0.0716	-0.1147
<b>TLT</b>	<b>2.5%</b>	0.1599	-0.003	0.0862	0.1132	0	0.1586	-0.006
	<b>97.5%</b>	0.1983	0.0065	0.112	0.1448	0	0.195	0.0031
<b>LN2</b>	<b>2.5%</b>	-0.0241	-0.1931	-0.0973	-0.0716	-0.195	0	-0.1962
	<b>97.5%</b>	0.0286	-0.1571	-0.058	-0.024	-0.1586	0	-0.1602
<b>TRW</b>	<b>2.5%</b>	0.1613	-0.0015	0.0877	0.1147	-0.0031	0.1602	0
	<b>97.5%</b>	0.1997	0.0078	0.1134	0.1462	0.006	0.1962	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk = 2,  
sf=0.3 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.2498	-0.1375	-0.0985	-0.3126	-0.0165	-0.2522
	<b>97.5%</b>	0	-0.1973	-0.0819	-0.0451	-0.2592	0.0408	-0.1989
<b>KHB</b>	<b>2.5%</b>	0.1973	0	0.0927	0.1289	-0.0757	0.2087	-0.0162
	<b>97.5%</b>	0.2498	0	0.135	0.1746	-0.049	0.2628	0.0123
<b>BC</b>	<b>2.5%</b>	0.0819	-0.135	0	0.012	-0.1974	0.0931	-0.1377
	<b>97.5%</b>	0.1375	-0.0927	0	0.0638	-0.1549	0.1506	-0.094
<b>LN1</b>	<b>2.5%</b>	0.0451	-0.1746	-0.0638	0	-0.2369	0.0569	-0.1767
	<b>97.5%</b>	0.0985	-0.1289	-0.012	0	-0.1912	0.111	-0.1308
<b>TLT</b>	<b>2.5%</b>	0.2592	0.049	0.1549	0.1912	0	0.2708	0.0471
	<b>97.5%</b>	0.3126	0.0757	0.1974	0.2369	0	0.3253	0.0736
<b>LN2</b>	<b>2.5%</b>	-0.0408	-0.2628	-0.1506	-0.111	-0.3253	0	-0.2648
	<b>97.5%</b>	0.0165	-0.2087	-0.0931	-0.0569	-0.2708	0	-0.2106
<b>TRW</b>	<b>2.5%</b>	0.1989	-0.0123	0.094	0.1308	-0.0736	0.2106	0
	<b>97.5%</b>	0.2522	0.0162	0.1377	0.1767	-0.0471	0.2648	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollié (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk = 2,  
sf=0.1 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0804	-0.0754	-0.0747	-0.2879	-0.0186	-0.0588
	<b>97.5%</b>	0	-0.0176	-0.0095	-0.0128	-0.2251	0.0451	0.0053
<b>KHB</b>	<b>2.5%</b>	0.0176	0	-0.0237	-0.0256	-0.237	0.0302	-0.0081
	<b>97.5%</b>	0.0804	0	0.0368	0.0362	-0.178	0.0943	0.0527
<b>BC</b>	<b>2.5%</b>	0.0095	-0.0368	0	-0.0337	-0.2436	0.0218	-0.0163
	<b>97.5%</b>	0.0754	0.0237	0	0.0312	-0.1845	0.0897	0.0477
<b>LN1</b>	<b>2.5%</b>	0.0128	-0.0362	-0.0312	0	-0.2423	0.0258	-0.0147
	<b>97.5%</b>	0.0747	0.0256	0.0337	0	-0.1833	0.0881	0.0487
<b>TLT</b>	<b>2.5%</b>	0.2251	0.178	0.1845	0.1833	0	0.2377	0.199
	<b>97.5%</b>	0.2879	0.237	0.2436	0.2423	0	0.3018	0.2606
<b>LN2</b>	<b>2.5%</b>	-0.0451	-0.0943	-0.0897	-0.0881	-0.3018	0	-0.0722
	<b>97.5%</b>	0.0186	-0.0302	-0.0218	-0.0258	-0.2377	0	-0.0077
<b>TRW</b>	<b>2.5%</b>	-0.0053	-0.0527	-0.0477	-0.0487	-0.2606	0.0077	0
	<b>97.5%</b>	0.0588	0.0081	0.0163	0.0147	-0.199	0.0722	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk = 3,  
sf=1 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0191	-0.019	-0.0033	-0.0191	-0.003	-0.0191
	<b>97.5%</b>	0	-0.0134	-0.0134	0.003	-0.0134	0.0032	-0.0134
<b>KHB</b>	<b>2.5%</b>	0.0134	0	-0.0001	0.0133	-0.0001	0.0135	-0.0001
	<b>97.5%</b>	0.0191	0	0.0003	0.0189	0.0002	0.0192	0.0001
<b>BC</b>	<b>2.5%</b>	0.0134	-0.0003	0	0.0132	-0.0003	0.0135	-0.0003
	<b>97.5%</b>	0.019	0.0001	0	0.0188	0.0001	0.0191	0.0001
<b>LN1</b>	<b>2.5%</b>	-0.003	-0.0189	-0.0188	0	-0.0189	-0.0028	-0.019
	<b>97.5%</b>	0.0033	-0.0133	-0.0132	0	-0.0133	0.0034	-0.0133
<b>TLT</b>	<b>2.5%</b>	0.0134	-0.0002	-0.0001	0.0133	0	0.0135	-0.0002
	<b>97.5%</b>	0.0191	0.0001	0.0003	0.0189	0	0.0192	0.0001
<b>LN2</b>	<b>2.5%</b>	-0.0032	-0.0192	-0.0191	-0.0034	-0.0192	0	-0.0193
	<b>97.5%</b>	0.003	-0.0135	-0.0135	0.0028	-0.0135	0	-0.0135
<b>TRW</b>	<b>2.5%</b>	0.0134	-0.0001	-0.0001	0.0133	-0.0001	0.0135	0
	<b>97.5%</b>	0.0191	0.0001	0.0003	0.019	0.0002	0.0193	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk = 3,  
sf=0.5 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0298	-0.0281	-0.0056	-0.0295	-0.0057	-0.0299
	<b>97.5%</b>	0	-0.022	-0.0204	0.0029	-0.0218	0.0029	-0.0221
<b>KHB</b>	<b>2.5%</b>	0.022	0	0.0008	0.0208	-0.0003	0.0208	-0.0006
	<b>97.5%</b>	0.0298	0	0.0025	0.0283	0.0008	0.0282	0.0005
<b>BC</b>	<b>2.5%</b>	0.0204	-0.0025	0	0.0192	-0.0023	0.0192	-0.0026
	<b>97.5%</b>	0.0281	-0.0008	0	0.0266	-0.0005	0.0265	-0.0009
<b>LN1</b>	<b>2.5%</b>	-0.0029	-0.0283	-0.0266	0	-0.028	-0.0043	-0.0284
	<b>97.5%</b>	0.0056	-0.0208	-0.0192	0	-0.0206	0.0041	-0.0209
<b>TLT</b>	<b>2.5%</b>	0.0218	-0.0008	0.0005	0.0206	0	0.0205	-0.0009
	<b>97.5%</b>	0.0295	0.0003	0.0023	0.028	0	0.028	0.0003
<b>LN2</b>	<b>2.5%</b>	-0.0029	-0.0282	-0.0265	-0.0041	-0.028	0	-0.0283
	<b>97.5%</b>	0.0057	-0.0208	-0.0192	0.0043	-0.0205	0	-0.0208
<b>TRW</b>	<b>2.5%</b>	0.0221	-0.0005	0.0009	0.0209	-0.0003	0.0208	0
	<b>97.5%</b>	0.0299	0.0006	0.0026	0.0284	0.0009	0.0283	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk = 3,  
sf=0.3 in Hot Spots

		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.0782	-0.0672	-0.0163	-0.0778	-0.0142	-0.0783
	<b>97.5%</b>	0	-0.0616	-0.0508	0.0015	-0.0614	0.0037	-0.0618
<b>KHB</b>	<b>2.5%</b>	0.0616	0	0.0084	0.0549	-0.001	0.0569	-0.0014
	<b>97.5%</b>	0.0782	0	0.0134	0.0701	0.0016	0.0724	0.0012
<b>BC</b>	<b>2.5%</b>	0.0508	-0.0134	0	0.044	-0.0131	0.046	-0.0135
	<b>97.5%</b>	0.0672	-0.0084	0	0.0592	-0.0081	0.0616	-0.0085
<b>LN1</b>	<b>2.5%</b>	-0.0015	-0.0701	-0.0592	0	-0.0698	-0.0064	-0.0702
	<b>97.5%</b>	0.0163	-0.0549	-0.044	0	-0.0546	0.0108	-0.0551
<b>TLT</b>	<b>2.5%</b>	0.0614	-0.0016	0.0081	0.0546	0	0.0566	-0.0018
	<b>97.5%</b>	0.0778	0.001	0.0131	0.0698	0	0.0721	0.0009
<b>LN2</b>	<b>2.5%</b>	-0.0037	-0.0724	-0.0616	-0.0108	-0.0721	0	-0.0726
	<b>97.5%</b>	0.0142	-0.0569	-0.046	0.0064	-0.0566	0	-0.057
<b>TRW</b>	<b>2.5%</b>	0.0618	-0.0012	0.0085	0.0551	-0.0009	0.057	0
	<b>97.5%</b>	0.0783	0.0014	0.0135	0.0702	0.0018	0.0726	0

Entries refer to row heading – column heading.

BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

Confidence Limits for difference in AUCs for Scenario 4, Relative Risk = 3,  
sf=0.1 in Hot Spots

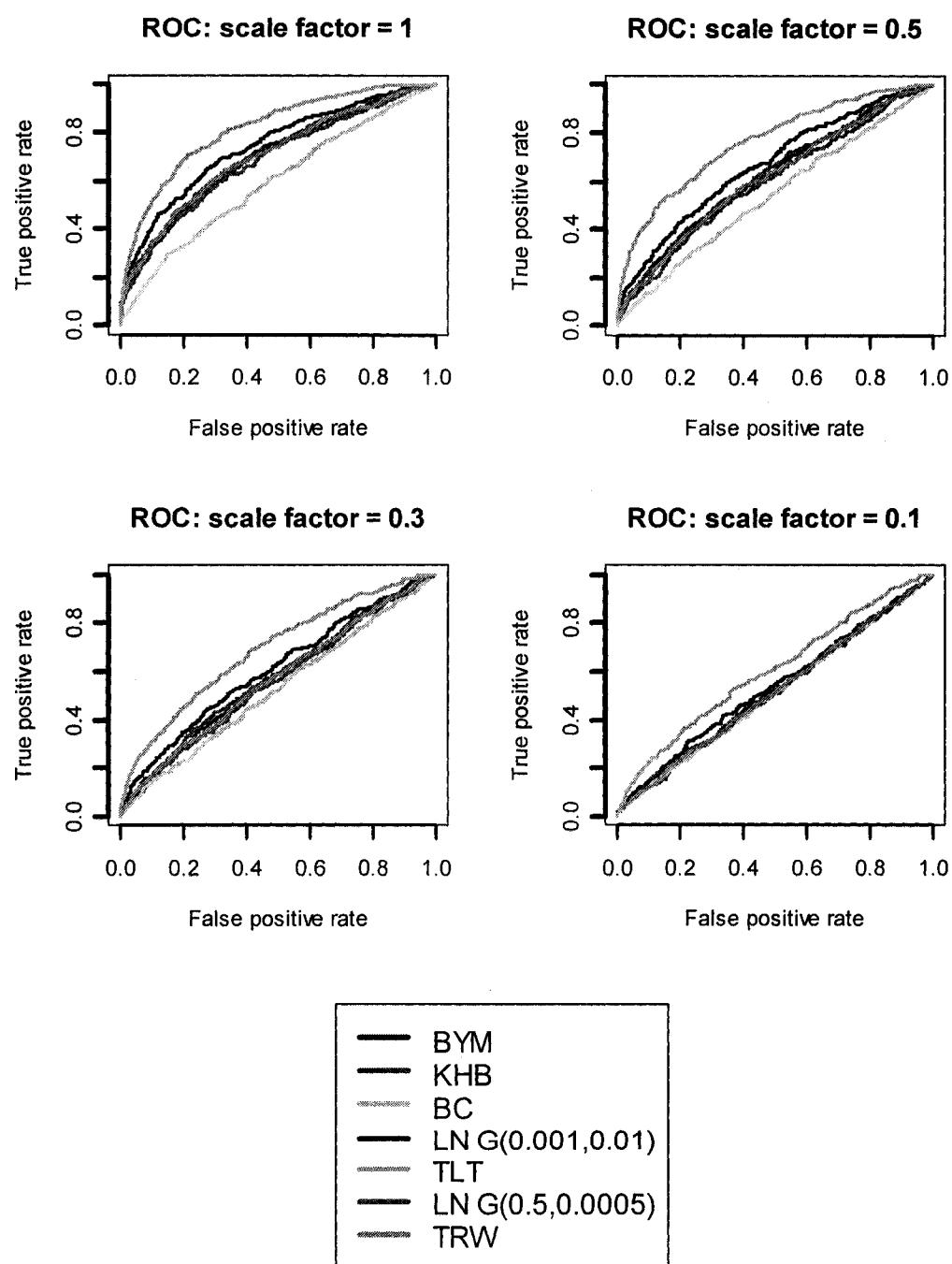
		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>BYM</b>	<b>2.5%</b>	0	-0.2957	-0.1088	-0.0977	-0.3021	-0.0169	-0.2993
	<b>97.5%</b>	0	-0.2467	-0.0548	-0.0478	-0.2525	0.0372	-0.2502
<b>KHB</b>	<b>2.5%</b>	0.2467	0	0.1695	0.1781	-0.0125	0.2563	-0.0101
	<b>97.5%</b>	0.2957	0	0.2094	0.2189	0.0004	0.3065	0.0031
<b>BC</b>	<b>2.5%</b>	0.0548	-0.2094	0	-0.0161	-0.2159	0.0634	-0.2132
	<b>97.5%</b>	0.1088	-0.1695	0	0.0341	-0.1751	0.1204	-0.1728
<b>LN1</b>	<b>2.5%</b>	0.0478	-0.2189	-0.0341	0	-0.2252	0.0576	-0.2225
	<b>97.5%</b>	0.0977	-0.1781	0.0161	0	-0.1839	0.1081	-0.1816
<b>TLT</b>	<b>2.5%</b>	0.2525	-0.0004	0.1751	0.1839	0	0.262	-0.0038
	<b>97.5%</b>	0.3021	0.0125	0.2159	0.2252	0	0.3128	0.0088
<b>LN2</b>	<b>2.5%</b>	-0.0372	-0.3065	-0.1204	-0.1081	-0.3128	0	-0.3101
	<b>97.5%</b>	0.0169	-0.2563	-0.0634	-0.0576	-0.262	0	-0.2598
<b>TRW</b>	<b>2.5%</b>	0.2502	-0.0031	0.1728	0.1816	-0.0088	0.2598	0
	<b>97.5%</b>	0.2993	0.0101	0.2132	0.2225	0.0038	0.3101	0

Entries refer to row heading – column heading.

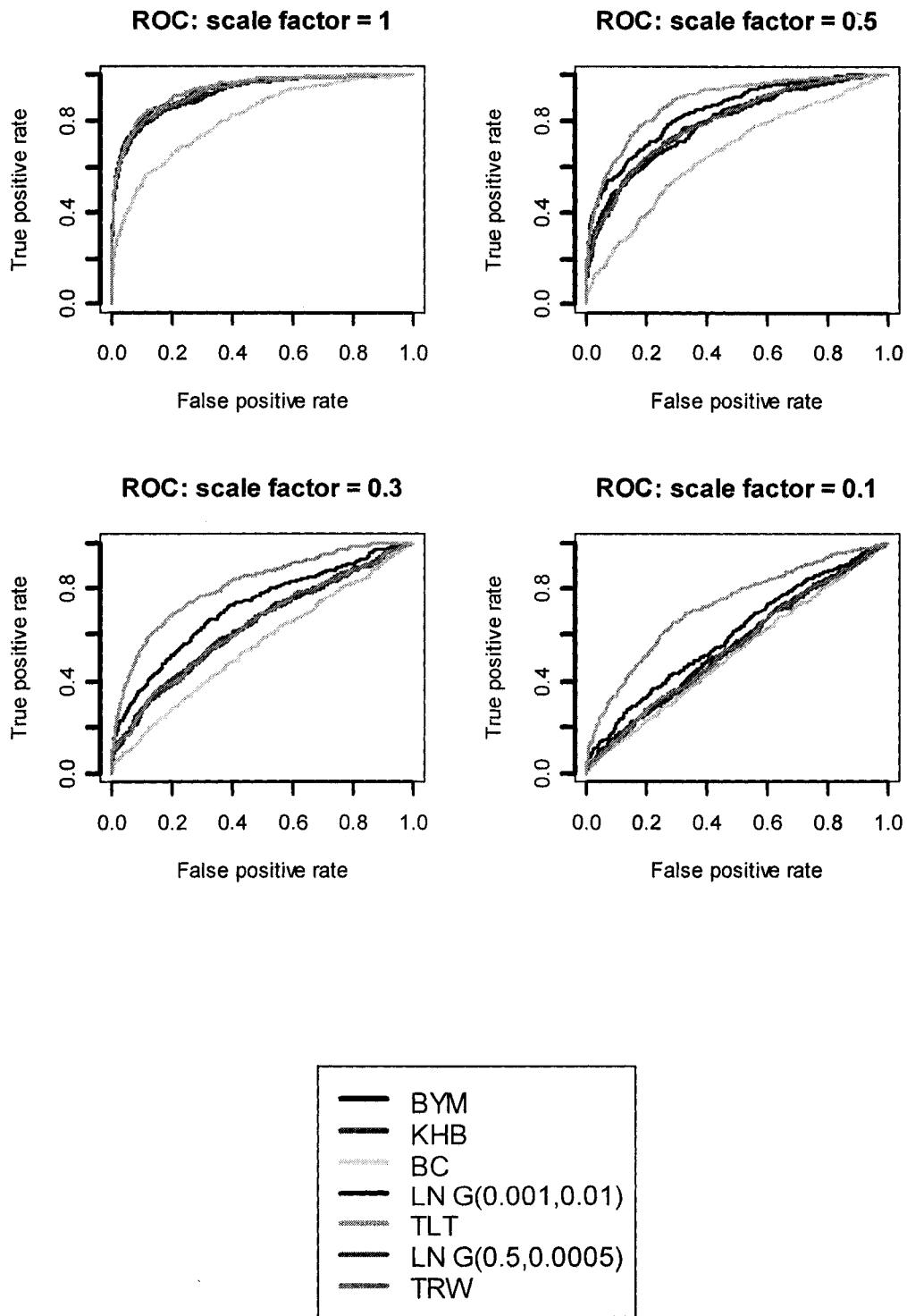
BYM = Besag, York and Mollie (1991) model, KHB = Knorr-Held and Besag (1998) model, BC = Bernardinelli (and Clayton) et al (1995), LN1 = log-normal model with Gamma(0.01, 0.01) hyperprior, TLT = Temporal Linear Trend, LN2 = log-normal model with Gamma(0.5, 0.0005) hyperprior, TRW = Temporal Random Walk model

## A5. ROC curves for spatial and spatio-temporal models

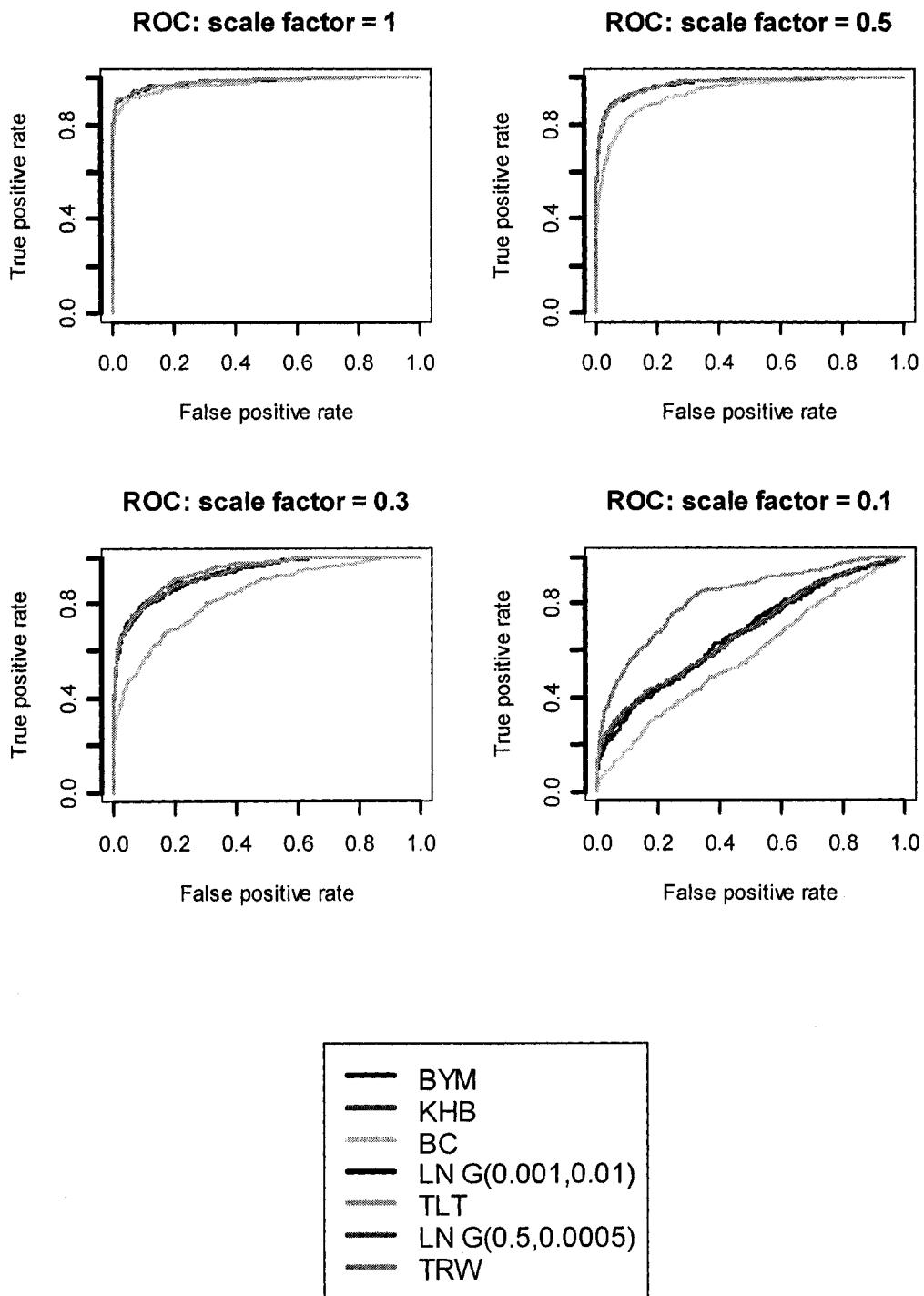
ROC Scenario 1, RR=1.5



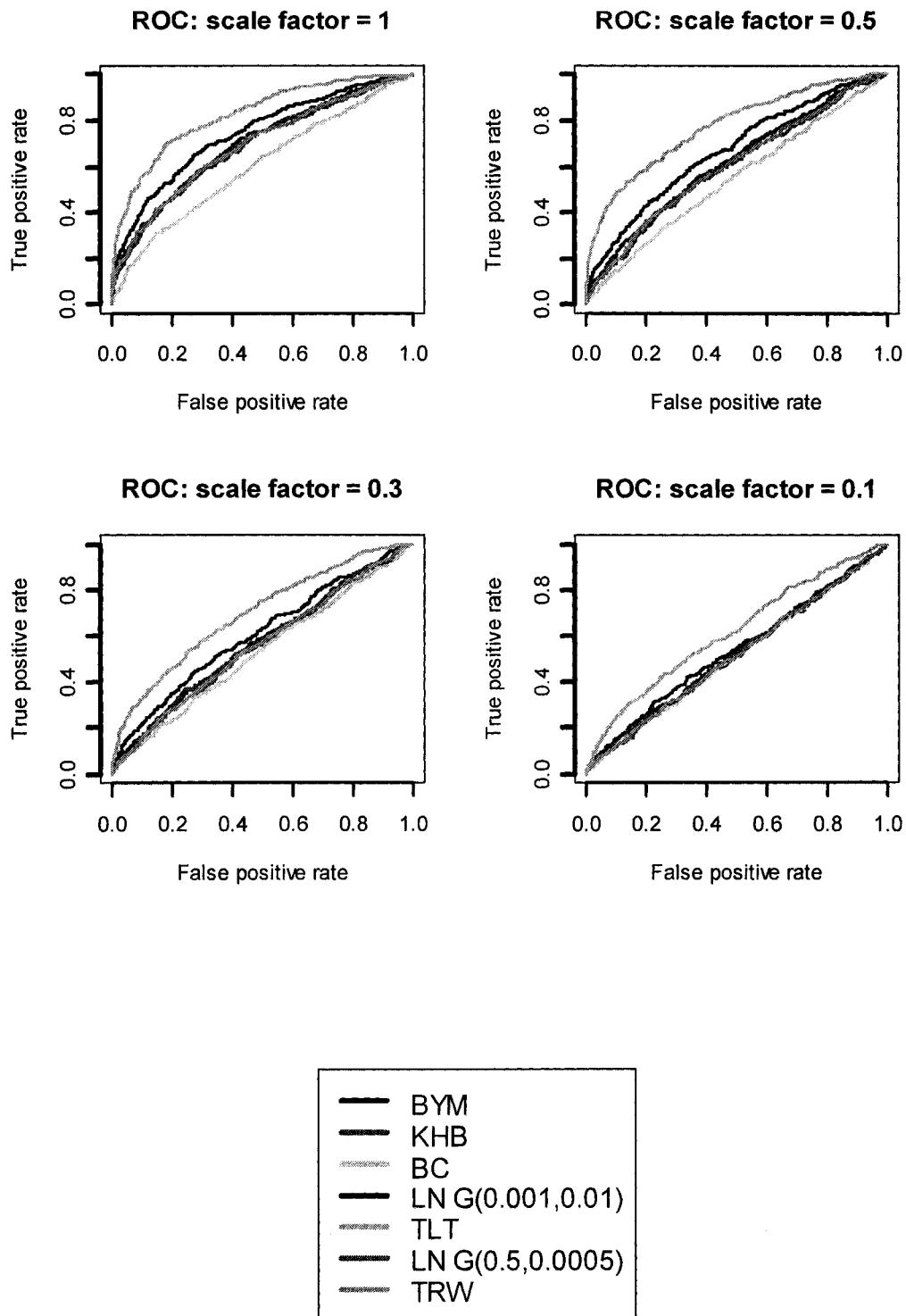
ROC Scenario 1, RR=2



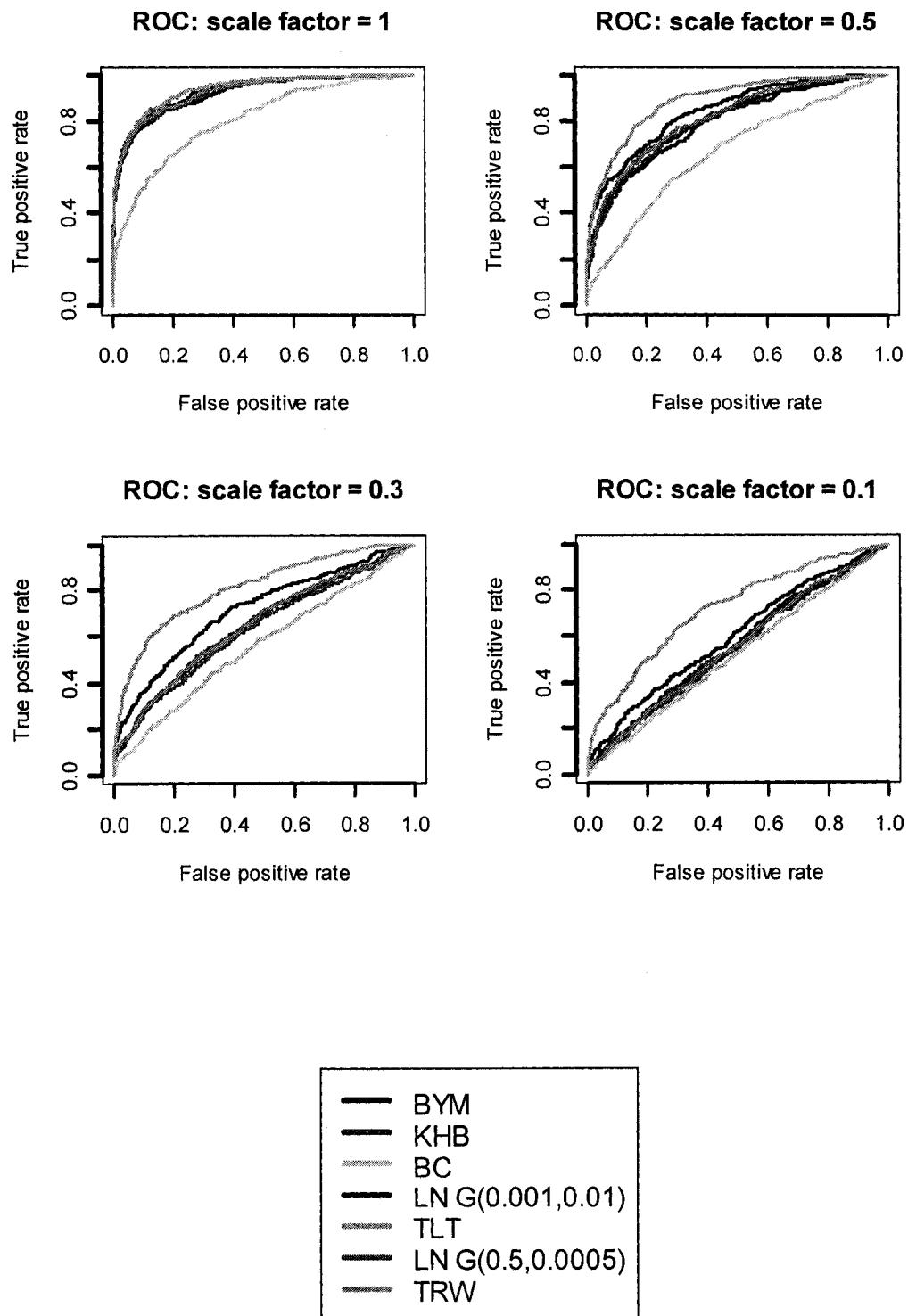
ROC Scenario 1, RR=3



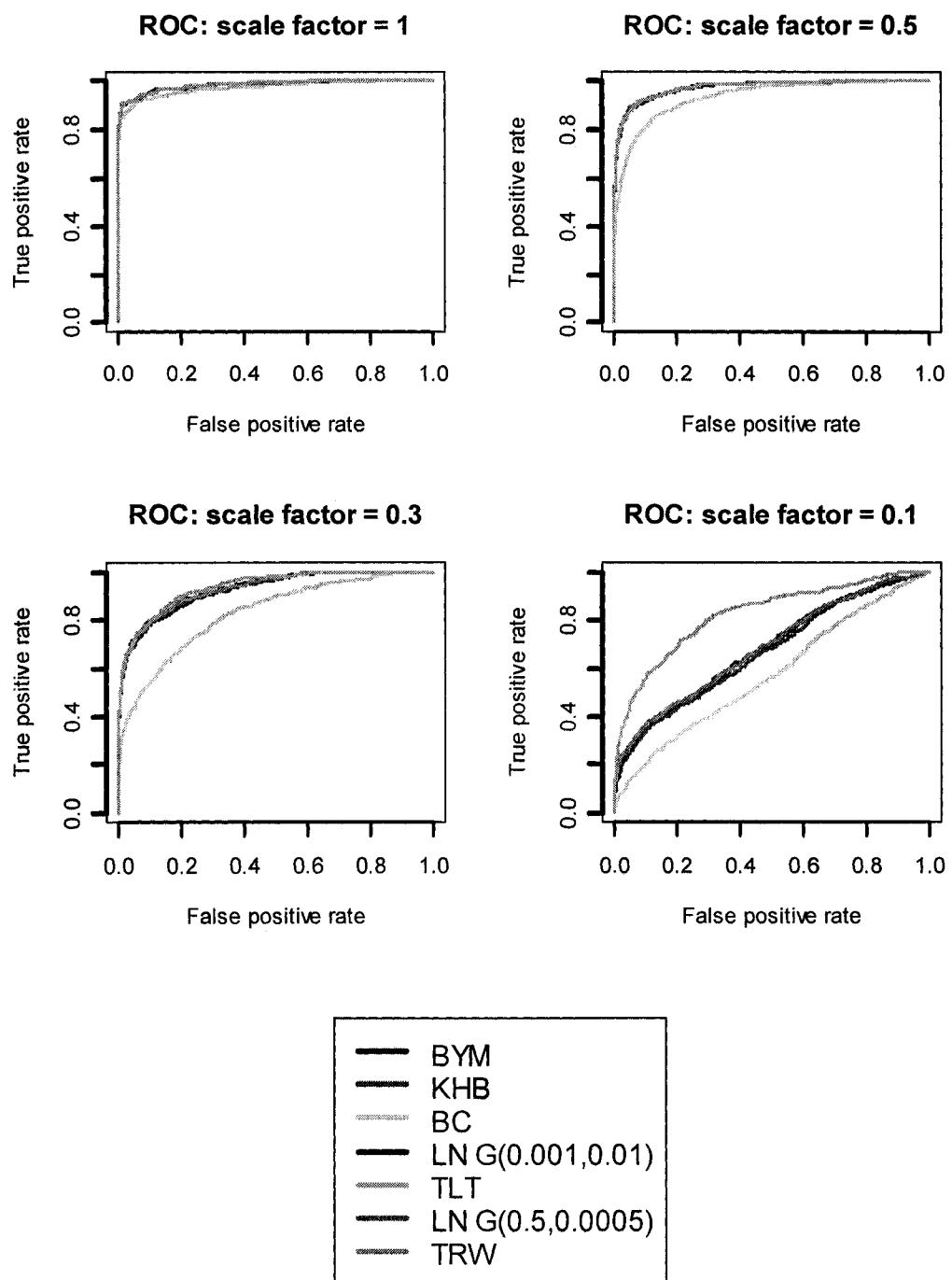
ROC Scenario 2, RR=1.5



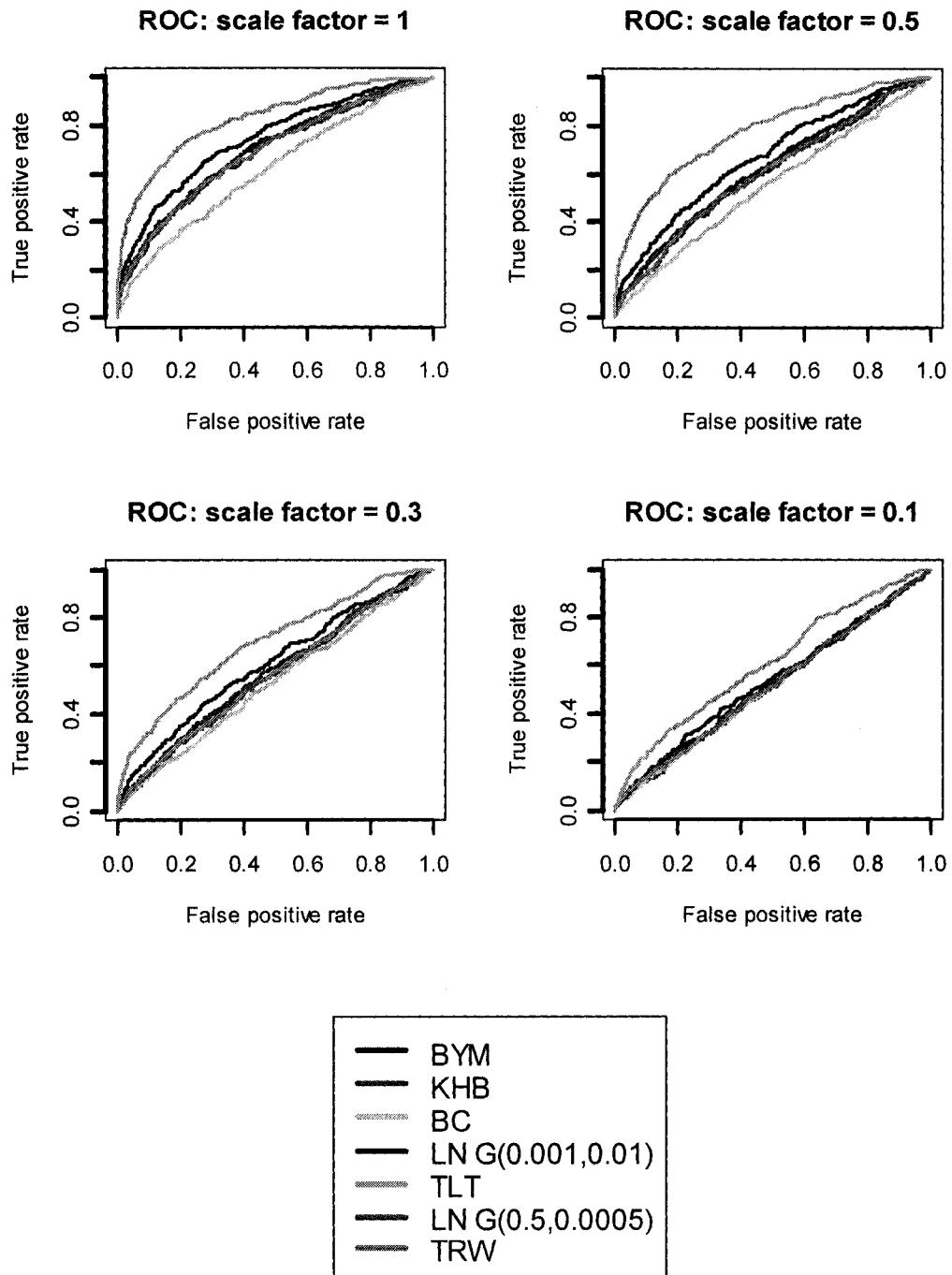
ROC Scenario 2, RR=2



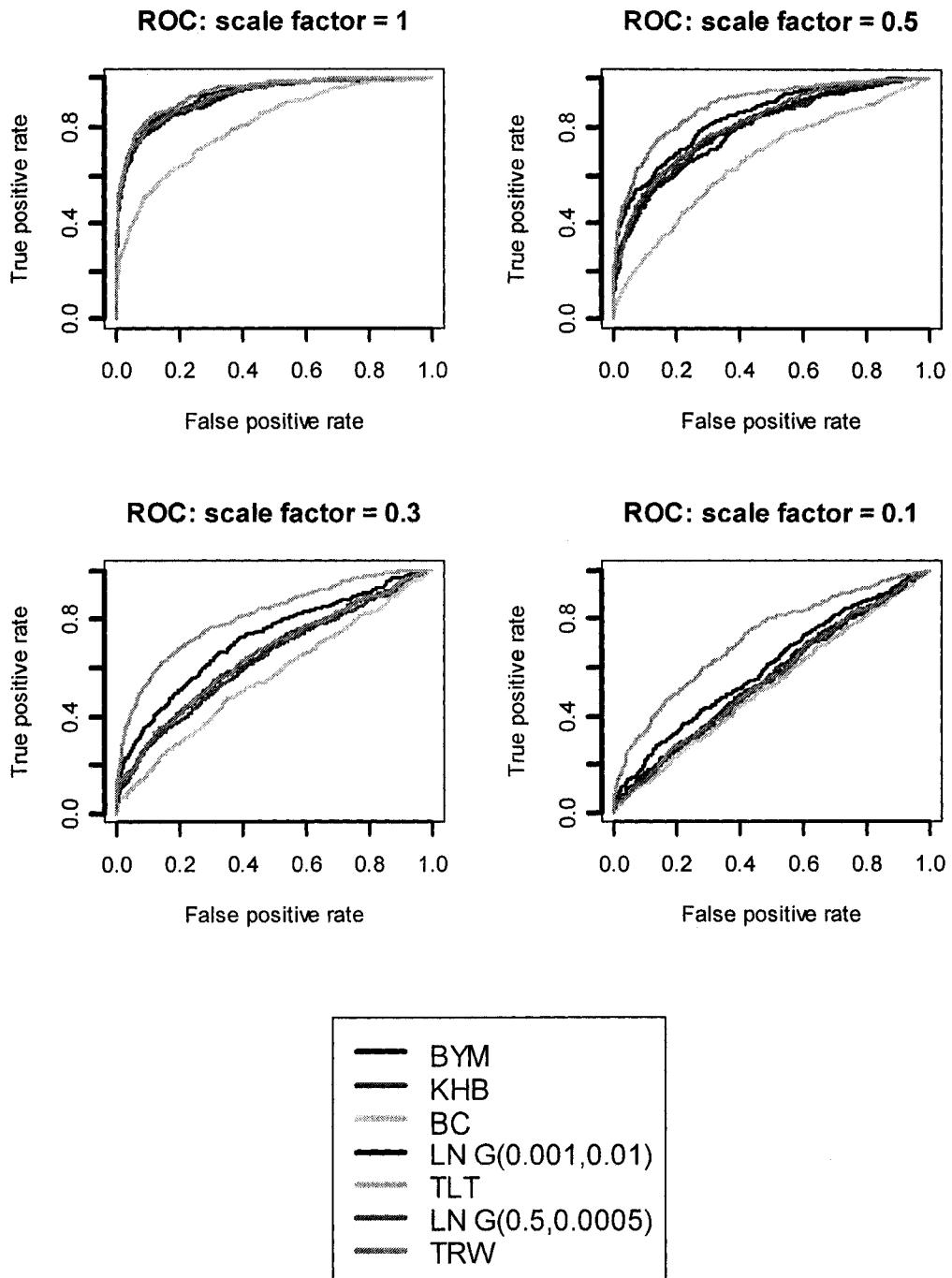
ROC Scenario 2, RR=3



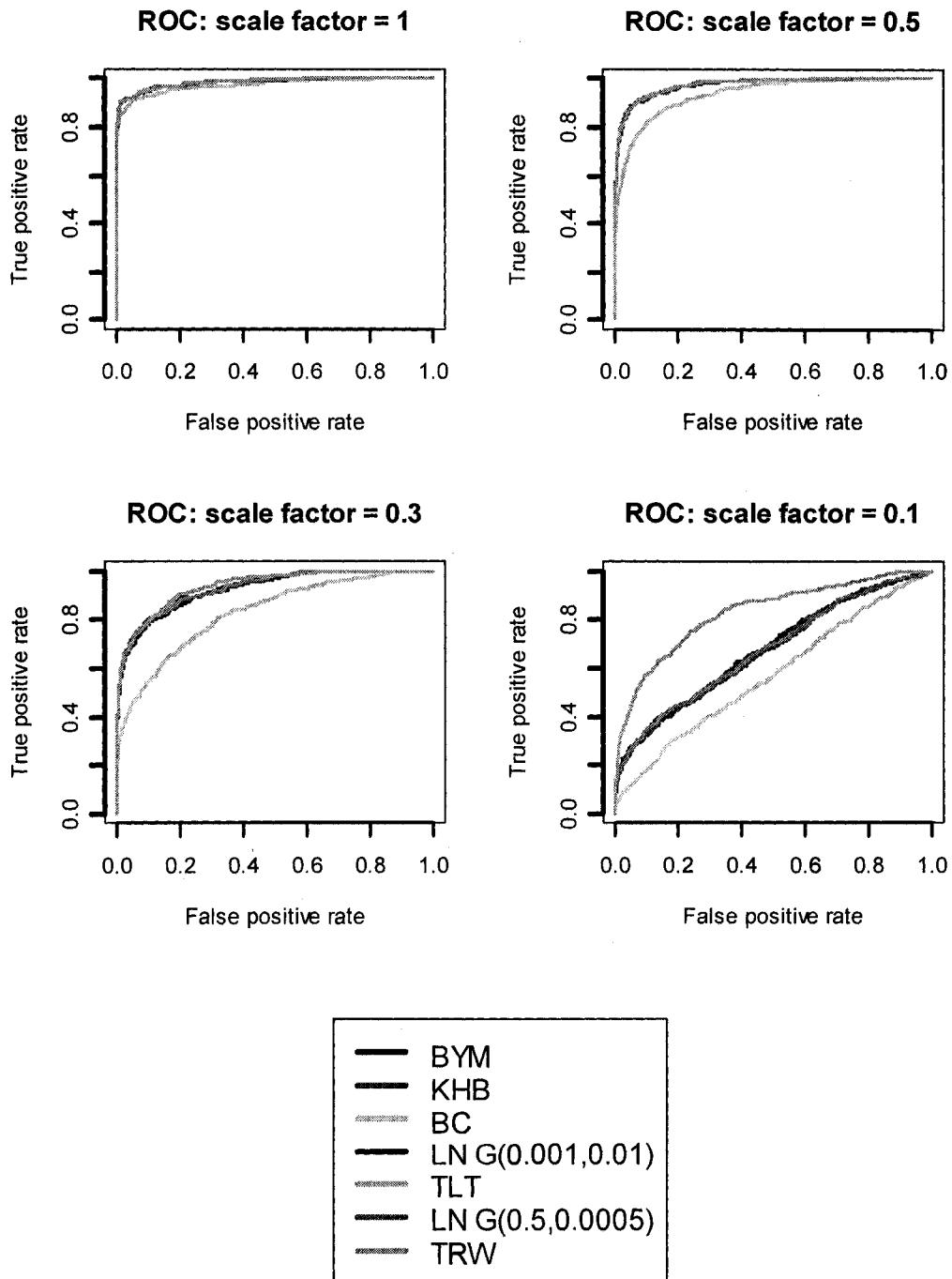
ROC Scenario 3, RR=1.5



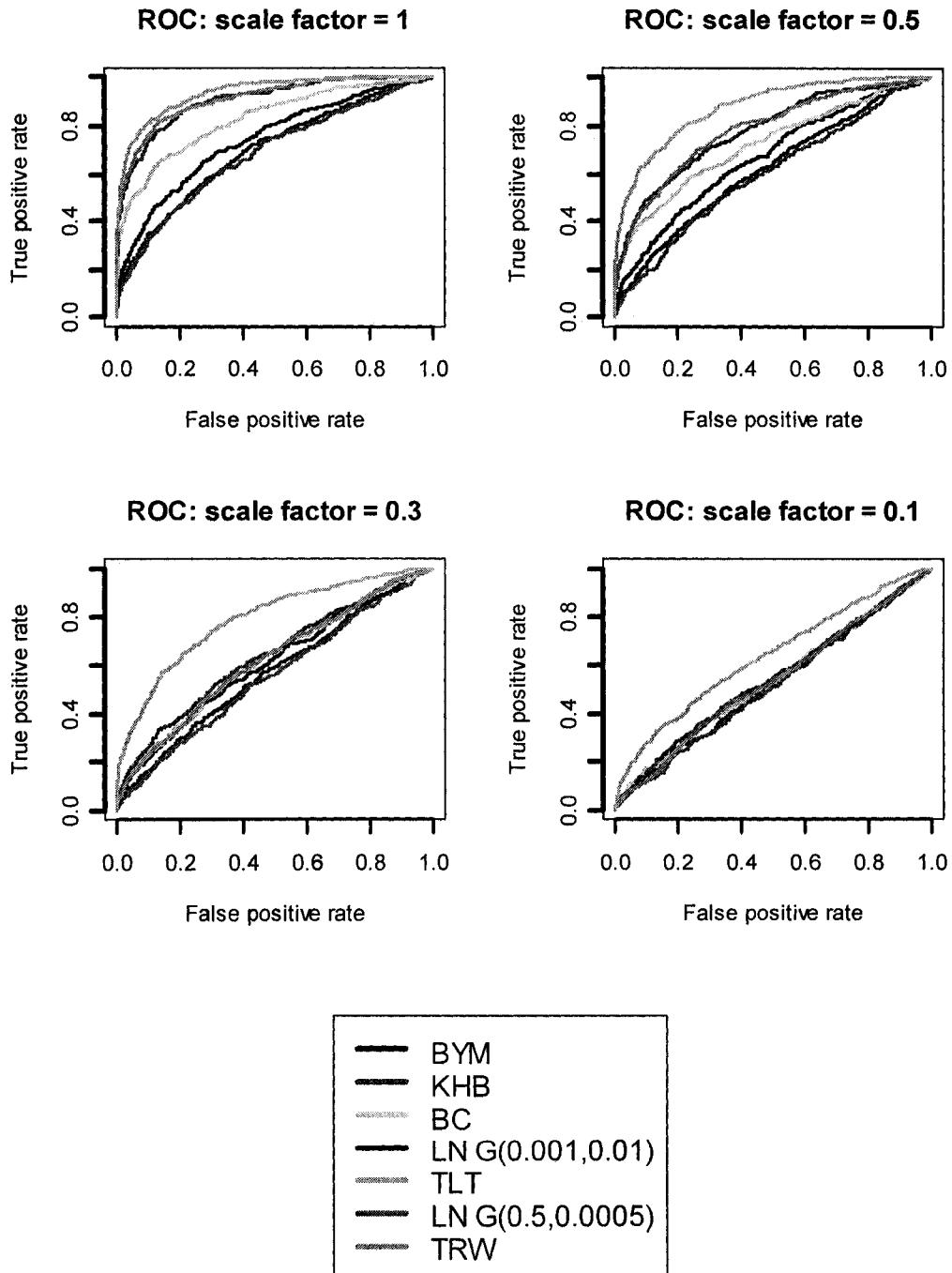
ROC Scenario 3, RR=2



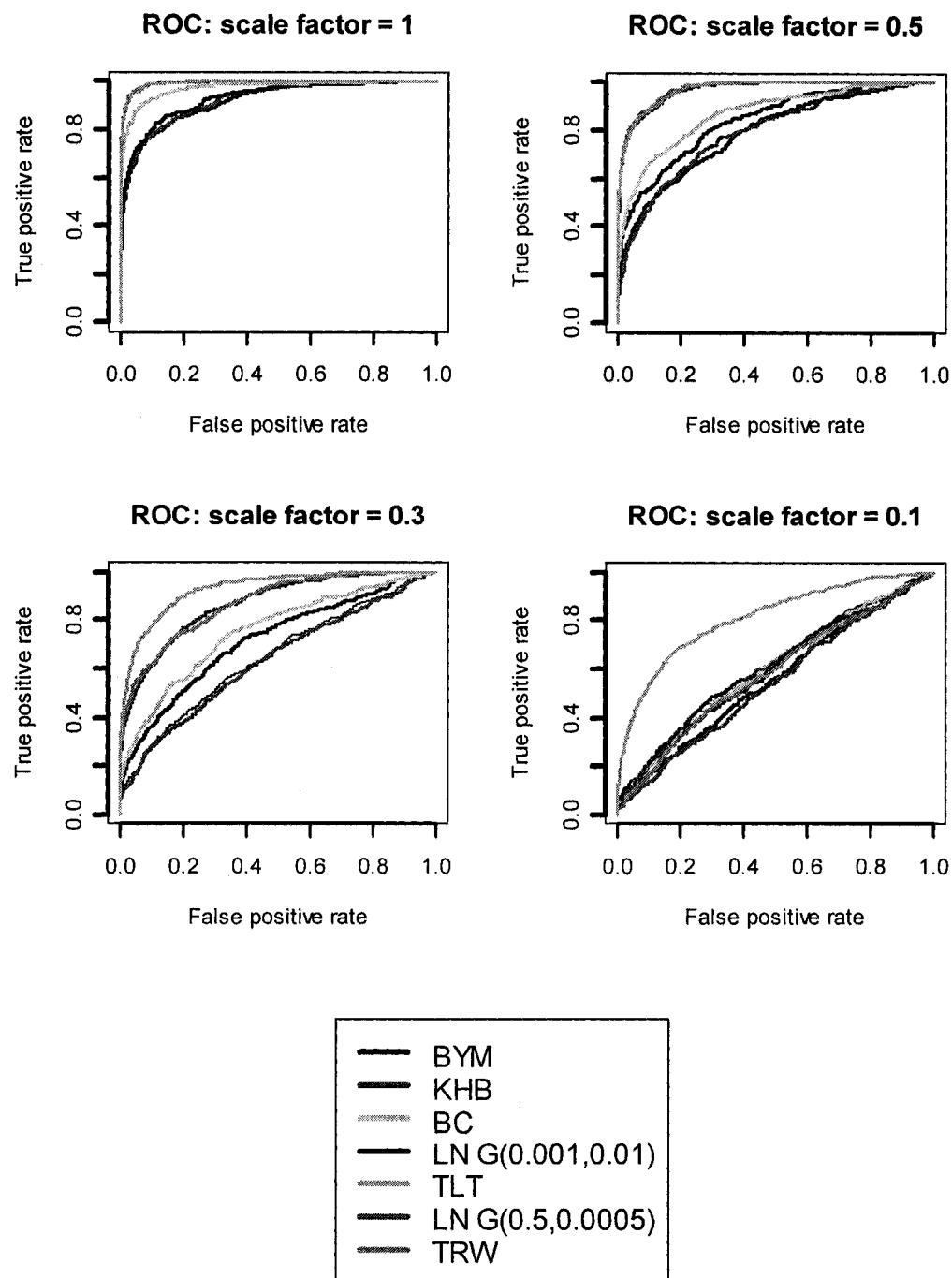
ROC Scenario 3, RR=3



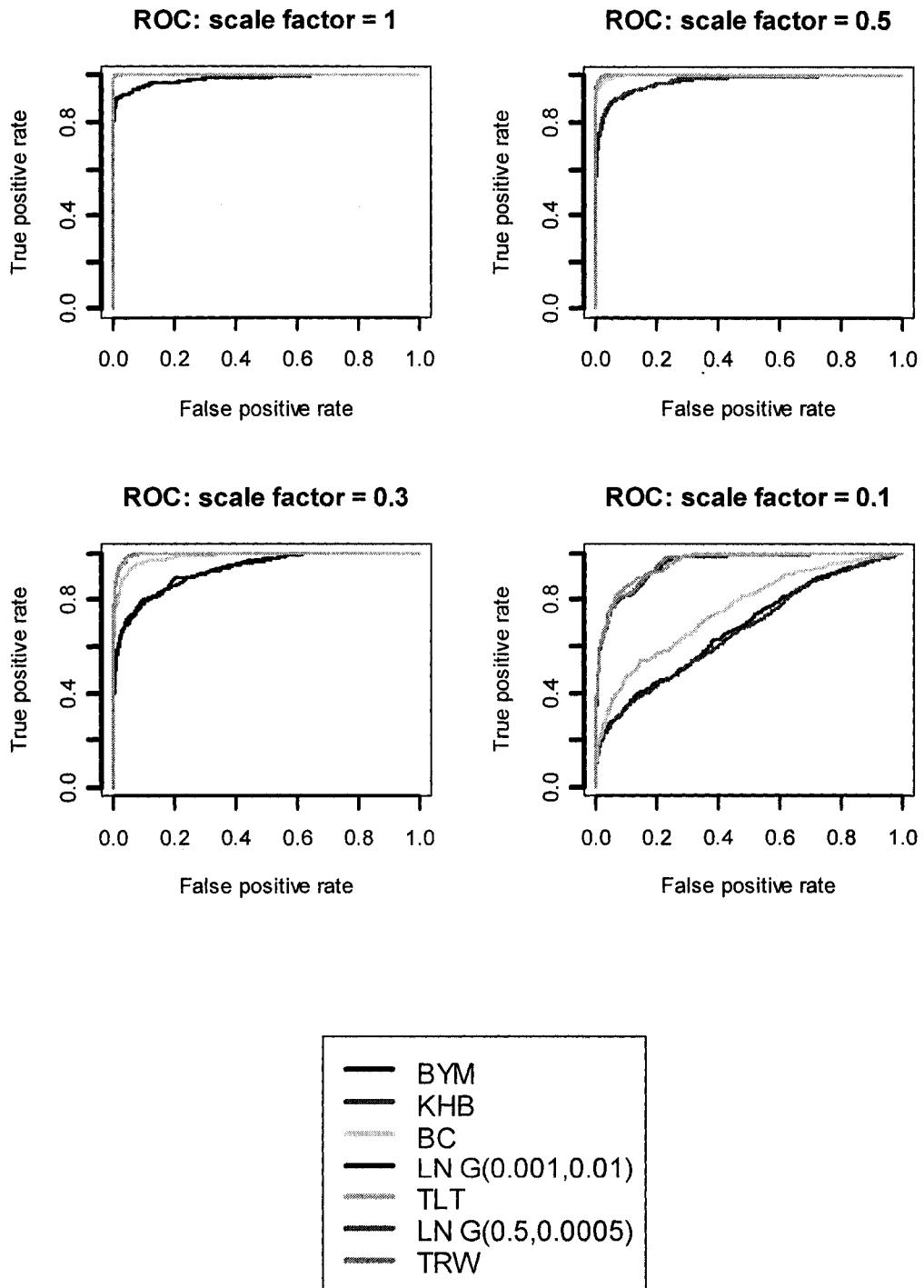
ROC Scenario 4, RR=1.5



ROC Scenario 4, RR=2



ROC Scenario 4, RR=3



## A6. Maximum $\hat{R}$ values

Maximum  $\hat{R}$  Scenario 2

<b>Hot Spot RR</b>		<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>1.5</b>	<b>1</b>	1.02	1.01	1.08	1.01	1.01	1.03	1.02
	<b>0.5</b>	1.05	1.02	1.01	1.02	1.01	1.02	1.00
	<b>0.3</b>	1.03	1.05	1.11	1.02	1.01	1.04	1.05
	<b>0.1</b>	1.01	1.07	1.14	1.04	1.02	1.02	1.01
<b>2</b>	<b>1</b>	1.02	1.00	1.02	1.00	1.00	1.00	1.01
	<b>0.5</b>	1.03	1.04	1.01	1.00	1.01	1.00	1.01
	<b>0.3</b>	1.08	1.09	1.16	1.01	1.03	1.12	1.04
	<b>0.1</b>	1.10	1.05	1.05	1.03	1.02	1.03	1.06
<b>3</b>	<b>1</b>	1.00	1.01	1.00	1.00	1.00	1.00	1.00
	<b>0.5</b>	1.00	1.01	1.01	1.00	1.00	1.00	1.00
	<b>0.3</b>	1.02	1.01	1.02	1.00	1.01	1.00	1.01
	<b>0.1</b>	1.04	1.05	1.05	1.02	1.01	1.14	1.03

Maximum  $\hat{R}$  Scenario 3

<b>Hot Spot RR</b>	<b>SF</b>	<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>1.5</b>	<b>1</b>	1.02	1.00	1.00	1.01	1.01	1.03	1.00
	<b>0.5</b>	1.05	1.01	1.01	1.02	1.01	1.02	1.01
	<b>0.3</b>	1.03	1.01	1.08	1.02	1.01	1.04	1.00
	<b>0.1</b>	1.01	1.11	1.02	1.04	1.02	1.02	1.01
<b>2</b>	<b>1</b>	1.02	1.01	1.03	1.00	1.01	1.00	1.01
	<b>0.5</b>	1.03	1.01	1.03	1.00	1.00	1.00	1.01
	<b>0.3</b>	1.08	1.01	1.09	1.01	1.02	1.12	1.09
	<b>0.1</b>	1.1	1.04	1.11	1.03	1.03	1.03	1.02
<b>3</b>	<b>1</b>	1.00	1.01	1.01	1.00	1.00	1.00	1.00
	<b>0.5</b>	1.00	1.01	1.01	1.00	1.01	1.00	1.01
	<b>0.3</b>	1.02	1.01	1.09	1.00	1.01	1.00	1.00
	<b>0.1</b>	1.04	1.11	1.02	1.02	1.02	1.14	1.01

Maximum  $\hat{R}$  Scenario 4

<b>Hot Spot RR</b>	<b>SF</b>	<b>BYM</b>	<b>KHB</b>	<b>BC</b>	<b>LN1</b>	<b>TLT</b>	<b>LN2</b>	<b>TRW</b>
<b>1.5</b>	<b>1</b>	1.02	1.01	1.01	1.01	1.00	1.03	1.02
	<b>0.5</b>	1.05	1.03	1.01	1.02	1.02	1.02	1.03
	<b>0.3</b>	1.03	1.01	1.02	1.02	1.01	1.04	1.01
	<b>0.1</b>	1.01	1.09	1.12	1.04	1.02	1.02	1.01
<b>2</b>	<b>1</b>	1.02	1.00	1.04	1.00	1.01	1.00	1.01
	<b>0.5</b>	1.03	1.05	1.02	1.00	1.01	1.00	1.01
	<b>0.3</b>	1.08	1.01	1.19	1.01	1.01	1.12	1.01
	<b>0.1</b>	1.1	1.04	1.04	1.03	1.02	1.03	1.06
<b>3</b>	<b>1</b>	1.00	1.01	1.01	1.00	1.00	1.00	1.00
	<b>0.5</b>	1.00	1.01	1.01	1.00	1.00	1.00	1.01
	<b>0.3</b>	1.02	1.01	1.08	1.00	1.01	1.00	1.00
	<b>0.1</b>	1.04	1.05	1.02	1.02	1.02	1.14	1.02

## REFERENCES

1. Aamodt G, Samuelsen S, Skrondel A. A simulation study of three methods for detecting disease clusters. *International Journal of Health Geographics* 2006; 5:15.
2. Alperovitch A, Bouvier MH. Geographical pattern of death rates from Multiple Sclerosis in France. An analysis of 4012 deaths. *Acta Neurol Scand* 1982; 66: 454-461.
3. Angers JF, Biswas A. A Bayesian analysis of zero-inflated generalized Poisson model. *Computational Statistics and Data Analysis* 2003; 42: 37-46.
4. Arnold RA, Diamond ID, Wakefield JC. The use of population data in spatial epidemiology. In: Elliot P, Wakefield J, Best N, Briggs D, editors. *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press, 2000b.
5. Bernardinelli L, Clayton D, Montomoli C. Bayesian estimates of disease maps : How important are priors? *Statistics in Medicine* 1995b; 14: 2411-2431.
6. Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M. Bayesian Analysis of Space-Time Variation in Disease Risk. *Statistics in Medicine* 1995a;14: 2433-2443.

7. Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; 43: 1-59.
8. Best NG, Ickstadt K, Wolpert RL. Spatial Poisson regression for health and exposure data measured at disparate resolutions. *Journal of the American Statistical Association* 2000; 95: 1076-1088.
9. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research* 2005; 14: 35-39.
10. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; 43: 671-668.
11. Congdon P. *Bayesian Statistical Modelling*. Chichester: John Wiley and Sons Ltd; 2001.
12. Cressie N and Chan NH. Spatial modelling of regional variables. *Journal of the American Statistical Association* 1989; 84: 393-401.
13. Denison DGT, Holmes CC. Bayesian partitioning for estimating disease risk. *Biometrics* 2001; 56: 922-935.
14. Elliot P, Wakefield JC. Bias and confounding in spatial epidemiology. In: Elliot P, Wakefield J, Best N, Briggs D, ed. *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press, 2000: 68-84.

15. Gangnon RE, Clayton MK. Bayesian partitioning for estimating disease risk. *Biometrics* 2001; 57: 143-149.
16. Gangnon RE, Clayton MK. A hierarchical model for spatially clustered disease rates. *Statistics in Medicine* 2003; **22**(20):3213-3228.
17. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Boca Raton: Chapman&Hall; 2003.
18. Granieri E. Introduction. In: Riise T, Wolfson C, editors. The Epidemiologic Study of Exogenous Factors in the Etiology of Multiple Sclerosis. *Neurology* 1997; 49:(Suppl2) 2-3.
19. Green PJ, Richardson S. Hidden Markov Models and Disease Mapping. *Journal of the American Statistical Association*. 2002; 97: 1055-1070.
20. Hanley JA, McNeil BJ. The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve. *Radiology* 1982; 143: 29-36.
21. Hanley JA, McNeil BJ. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. *Radiology* 1983;148: 839-843.
22. Hogg RV, Tanis EA. *Probability and Statistical Inference*. New York: Macmillan Publishing Company; 1988.

22. Jarup L, Best N, Toledano MB, Wakefield J, Elliott P. Geographical Epidemiology of Prostate Cancer in Great Britain. *Int. J. Cancer* 2002; 97:695-699.
23. Kahana E. Epidemiologic studies of multiple sclerosis: a review. *Biomed&Pharmacother* 2000; 54: 100-2.
24. Kelsall JE, Wakefield JC. Discussion of 'Bayesian models for spatially correlated disease and exposure data', by Best et al. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, ed. Bayesian Statistics 6, 151, Oxford University Press, 1999.
25. Kelsall J, Wakefield JC. Modeling spatial variation in disease risk: A geostatistical approach. *Journal of the American Statistical Association* 2002; 97:692-701.
26. Kentucky Cancer Registry. (2006). Cancer Incidence/Mortality Rates in Kentucky. Retrieved (7/21/2005) from <http://www.kcr.uky.edu/>.
27. Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* 2000;19 (17-18):2555-67.
28. Knorr-Held L, Besag J. Modelling risk from a disease in time and space. *Statistics in Medicine* 1998;17: 2045-2060.
29. Knorr-Held L, Raßer G. Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* 2000; 56: 13-21.
30. Knorr-Held L, Richardson S. A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Applied Statistics* 2003; 52: 169-183.

31. Kurtzke F, Heltberg A. Multiple sclerosis in the Faroe Islands: an epitome. *Journal of Clinical Epidemiology* 2001; 54:1-22.
32. Lauer K. Environmental associations with the risk of multiple sclerosis: the contribution of ecological studies. *Acta Neurol Scand* 1995; 91: (Suppl 161): 77-88.
33. Lawson AB, Browne WJ, Rodeiro CLV. *Disease Mapping with WinBUGS and MLWin*. Chichester: John Wiley & Sons, Ltd; 2003.
34. Lawson A, Clark A. Spatial mixture relative risk models applied to disease mapping. *Statistics in Medicine* 2002; 21:359-370.
35. Maplandia.net. Available at: [www.maplandia.net](http://www.maplandia.net). Accessed December, 2006. (map of Sardinia relative to Italy)
36. MacNab YC. Hierarchical Bayesian Modeling of Spatially Correlted Health Service Outcome and Utilization Rates. *Biometrics* 2003; 59: 305-315.
37. MacNab YC, Dean CB. Spatio-temporal modeling of rates for the construction of disease maps. *Statistics in Medicine* 2002;21: 347-358.
38. Montomoli C, Allemani C, Solinas G, Motta G, Bernardinelli L, Clemente S, Murgia BS, Ticca AF, Musu L, Piras ML, Ferrai R, Caria A, Sanna S, Porcu O. An Ecologic Study of Geographical Variation in Multiple Sclerosis Risk in Central Sardinia, Italy. *Neuroepidemiology* 2002; 21:187-193.

39. MS Society of Canada website. 2006. Available at:  
<http://www.mssociety.ca>. Accessed November 2006.
40. Nobre AA, Schmidt AM, Lopes HF. Spatio-temporal models for mapping the incidence of malaria in Pará. *Environmetrics* 2005; 16:291-304.
41. Olson K, Grannis SJ, Mandl KD. Privacy Protection Versus Cluster Detection in Spatial Epidemiology. *American Journal of Public Health* 2006; Vol 96, No.11;2002-2008.
42. Pascutto C, Wakefield JC, Best NG, Richardson S, Bernardinelli L, Staines A, Elliot P. Statistical issues in the analysis of disease mapping data. *Statistics in Medicine* 2000;19: 2493-2519.
43. Pickle LW. Exploring spatio-temporal patterns of mortality using mixed effects models. *Statistics in Medicine* 2000;19 (17-18): 2251-2264.
44. Pugliatti M, Riise T, Sotgiu MA, Sotgiu S, Satta W, Mannu L, Sanna G, Rosati G. Increasing Incidence of Multiple Sclerosis in the Province of Sassari, Northern Sardinia. *Neuroepidemiology* 2005; 25: 129-134.
45. Pugliatti M, Solinas G, Sotgiu S, Castiglia P, Rosati G. Multiple sclerosis distribution in northern Sardinia: spatial cluster analysis of prevalence. *Neurology* 2002; 58:277-282.
46. R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available at:  
<http://www.R-project.org>.

47. Richardson S, Thomson A, Best N, Elliott P. Interpreting Posterior Relative Risk Estimates in Disease-Mapping Studies. *Environmental Health Perspectives* 2004;112:1016-1025.
48. Salway R, Wakefield J. Sources of bias in ecological studies of non-rare events. *Environmental and Ecological Statistics* 2005; 12:321-347.
49. Sardinian Dreams Available at: [www.sardiniandreams.com](http://www.sardiniandreams.com). Accessed December, 2006. (map of Sardinia)
50. Sardinia. Wikipedia Encyclopedia. Available at: [www.wikipedia.org](http://www.wikipedia.org). Accessed December, 2006.
51. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* 2002; 64 (Part 4): 583-639.
52. Spiegelhalter DJ, Thomas A Best NG, Lunn DJ. 2002. WinBUGS: Bayesian Inference Using Gibbs Sampling Manual, Version 1.4. London: Imperial College; Cambridge, UK:MRC Biostatistics Unit. Available at: <http://www.mrc-bsu.cam.ac.uk/winbugs>.
53. Sun D, Tsutakawa RK, Kim H, He Z. Spatio-temporal interaction with disease mapping. *Statistics in Medicine* 2000;19: 2015-2035.
54. Thomas A, Best N, Lunn D, Arnold R, Spiegelhalter D. GeoBUGS User Manual, Version 1.2. 2004. URL <http://www.mrc-bsu.cam.ac.uk/bugs>.

55. Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS Open. *R News* 2006; 6: 12-17.
56. Tobias Sing, Oliver Sander, Niko Beerenwinkel and Thomas Lengauer (2005). ROCR: Visualizing the performance of scoring classifiers. R package version 1.0-1. Available at: <http://rocr.bioinf.mpi-sb.mpg.de/>.
57. Vince Carey (2004). ROC: utilities for ROC, with uarray focus. R package version 1.4.0. Available at: <http://www.bioconductor.org>
58. Wakefield JC, Best NG, Waller L. Bayesian approaches to disease mapping. In: Elliot P, Wakefield J, Best N, Briggs D, ed. *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press, 2000a:104-127.
59. Wakefield JC, Kelsall JE, Morris SE. Clustering, cluster detection and spatial variation in risk. In: Elliot P, Wakefield J, Best N, Briggs D, ed. *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press, 2000b 128-152.
60. Waller LA, Carlin BP, Xia H. Structuring correlation within hierarchical spatio-temporal models for disease rates. In: Gregore TG, Brillinger DR, Diggle PJ, Russek-Cohen E, Warren WG, Woldinger RD, ed. *Modeling Longitudinal and Spatially Correlated Data: Lecture Notes in Statistics* 122, New York: Springer-Verlag, 1997.
61. Waller LA, Carlin BP, Xia H, Gelfand A. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* 1997; 92: 607-617.

62. Walter SD. Disease mapping: a historical perspective. In: Elliot P, Wakefield J, Best N, Briggs D, ed. *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press, 2000: 223-239.
63. Wolpert RL, Ickstadt K. Poisson/Gamma random field models for spatial statistics. *Biometrika* 1998; 85: 251-267.