A Bayesian Approach for Spatial Analysis of Lung Cancer Rates in Ohio

Hu Wei
Department of Geography

June 2005

Disease mapping

- Definition

  -Mapping the spatial dispersion of a certain disease across the study area

- Objective

  -Infer the geographic distribution of the rates and then identify areas of higher or lower incidence.

Mapping Relative Risk

- Relative risk measures how much a particular risk factor influences the risk of a specified outcome (e.g., cancer mortality)
- Classical approach is mapping SMRs (standardized mortality/morbidity rates) for subregions based on Poisson model
- Compute P-values for SMRs to identify areas with significantly high (or low) relative risk

Poisson Model

- For rare events a Poisson model is commonly adopted.

$$O_i \mid E_i, y_i \sim Poisson(E_i y_i)$$

- SMR = O_i/E_i is the MLE estimator of Relative Risk from the Poisson model, with estimated standard error $s_i = \sqrt{O_i} / E_i$ asymptotically.
- P-value can then be computed for each area with a certain SMR

Problems of SMR

- More extreme values of the estimates may be based on a few cases only in areas with small population.
- Rare events in small areas can lead to extra-Poisson variation.
- Spatial correlation in the Relative risks is not taken into account.

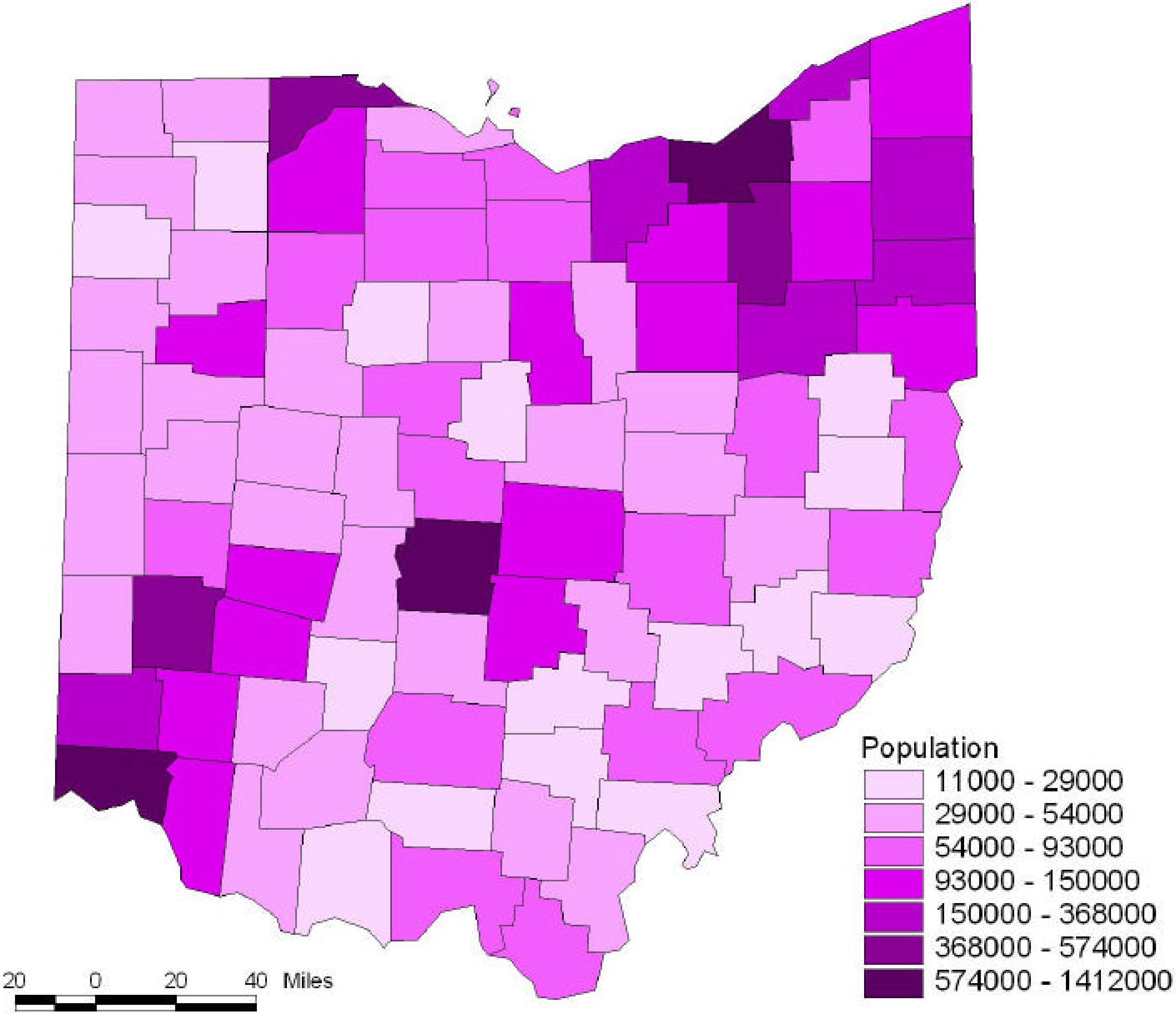
Bayesian approach

- Hierarchical model
  - Enable us to incorporate multiple sources of data and knowledge (e.g., spatial autocorrelation)
- Prior specification
  - Nonspatial random effect to describe unstructured heterogeneity.
  - Spatial random effect can be expressed via Markov random fields models (CAR, Exp)

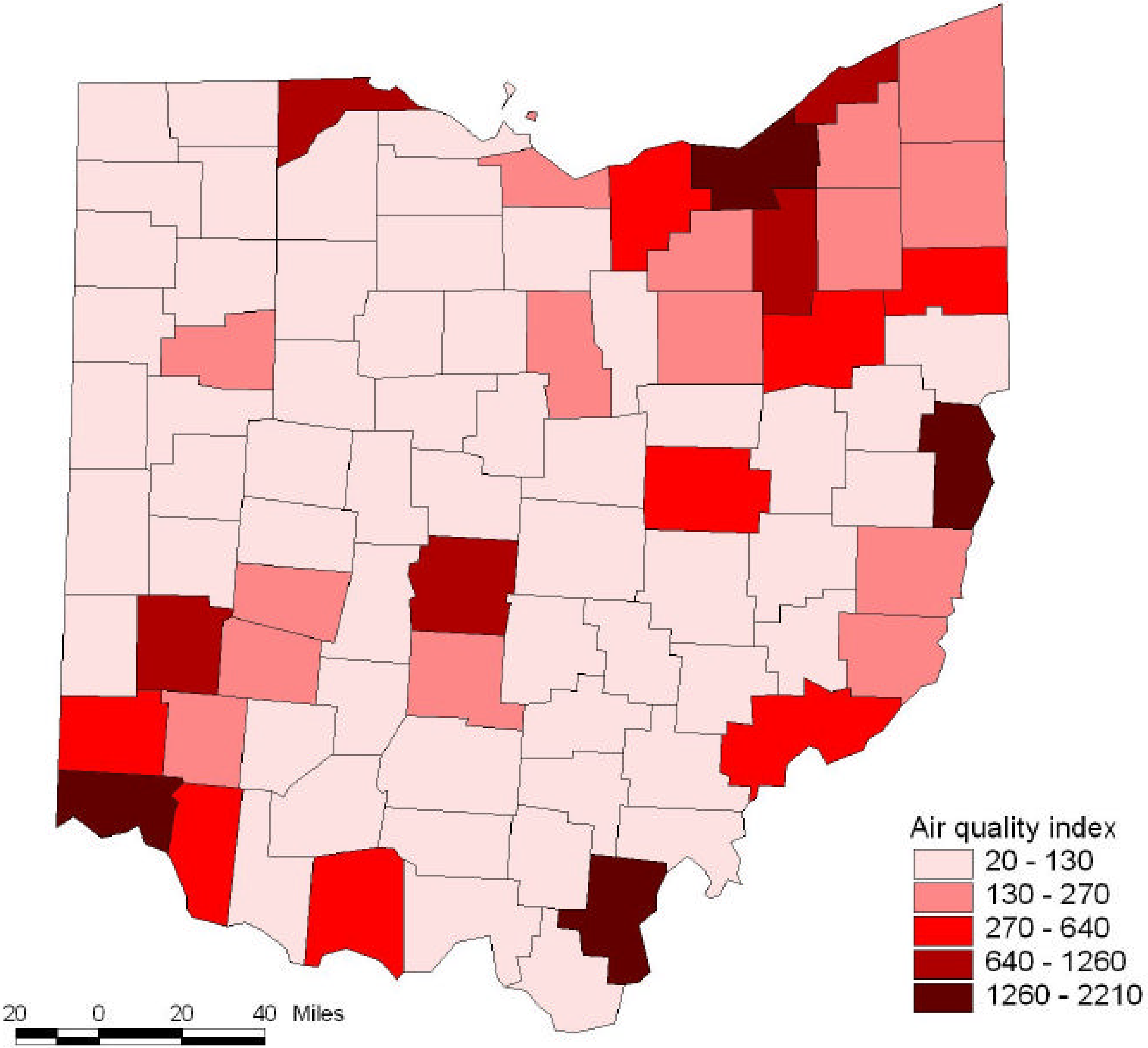
Data

- County map and census population for Ohio
- Observed lung cancer mortality at county level
  - National Cancer Institute
- Expected lung cancer mortality
  - Population in a county multiplied by crude rate
- Covariate variables
  - Air quality data from EPA
  - Poverty level: Census
- Software:
  - ArcView GIS, WinBUGS (GeoBUGS), and R.

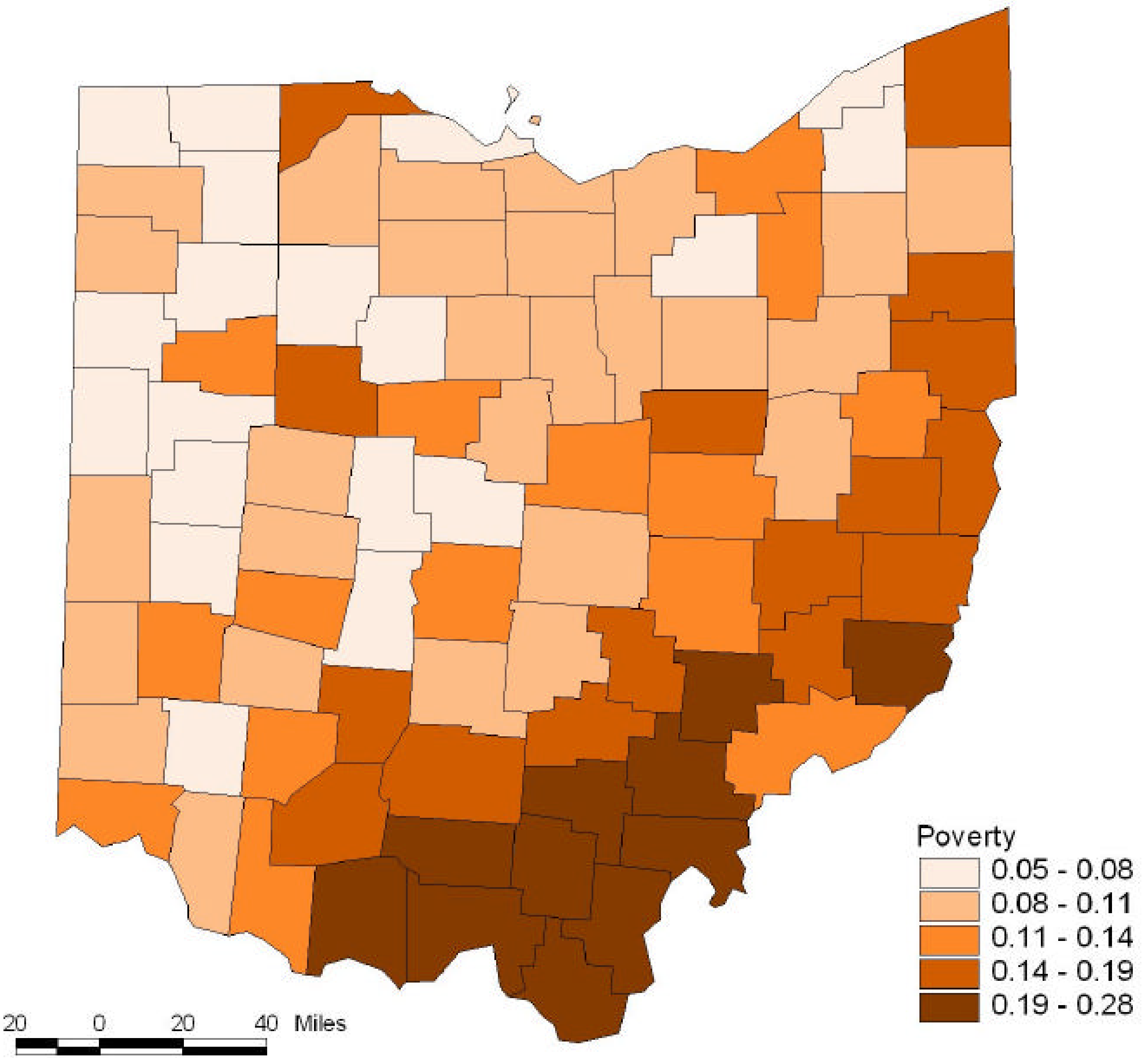
Population Distribution



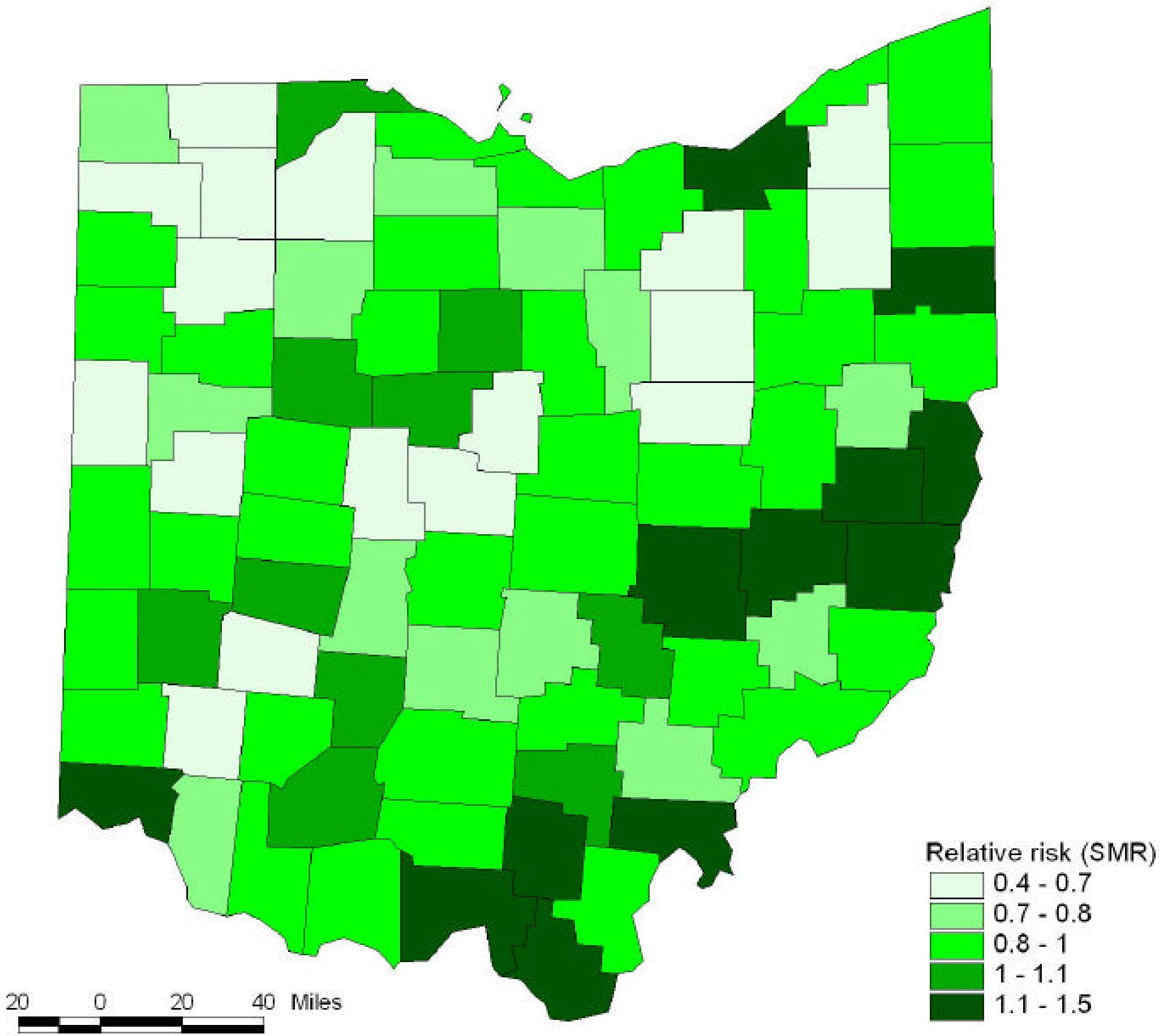
Air quality index



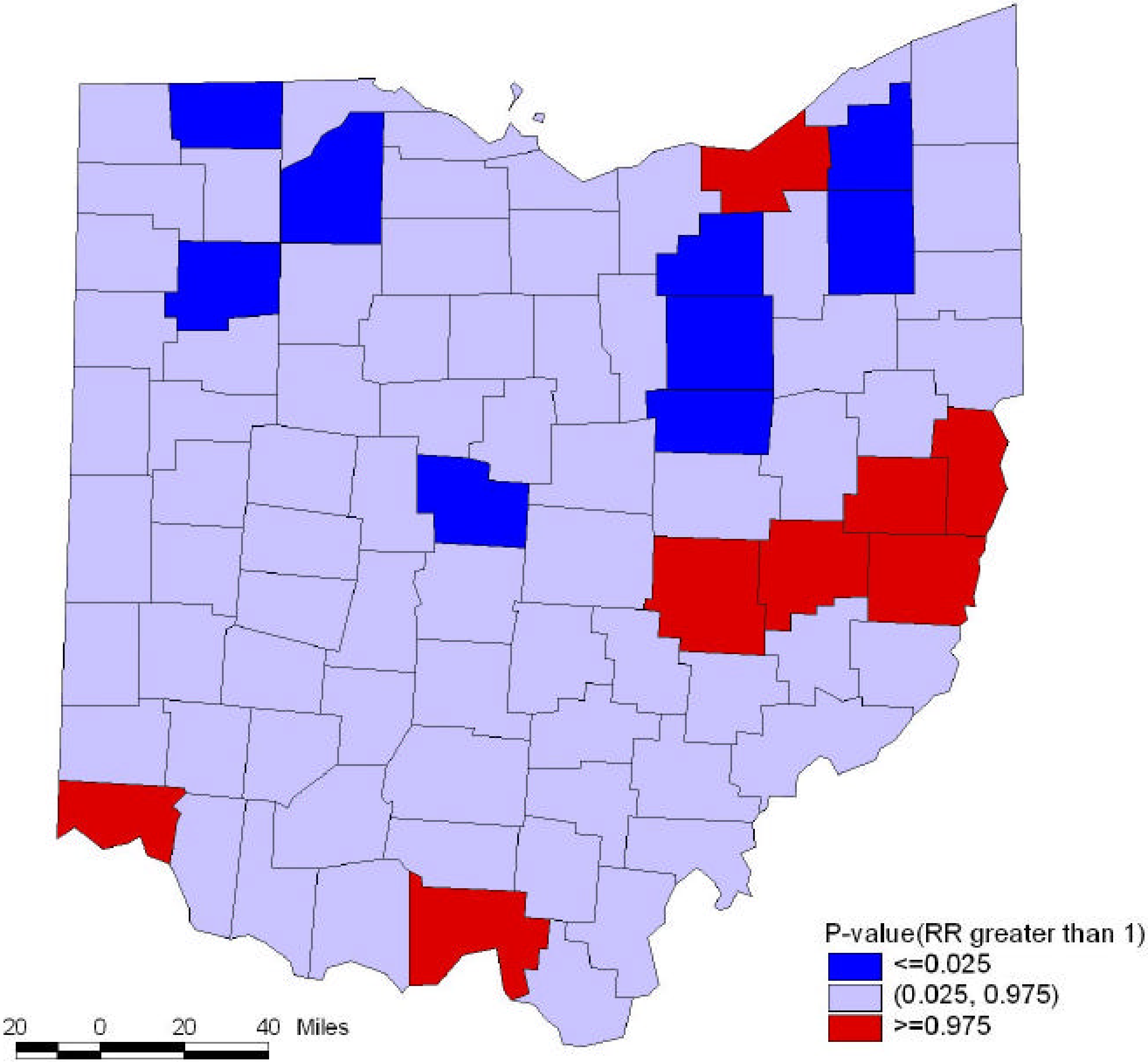
Poverty percentage



Standard mortality rate (SMR)



P-value



Hierarchical Bayesian Model using CAR prior

Likelihood:

O[i] ~ Poisson(mu[i])

First stage:

Log(mu[i]) = Log(E[i]) + phi[i] + theta[i]

RR[i] = exp(phi[i])

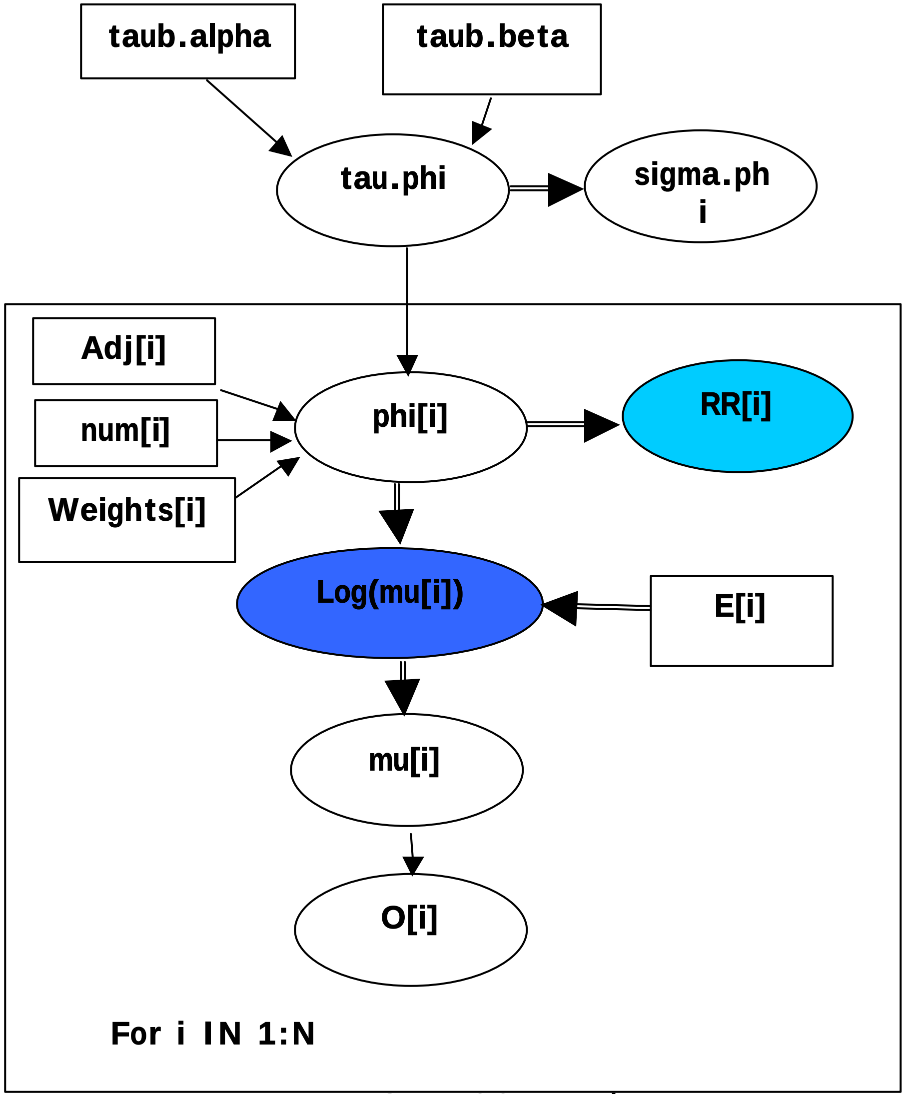
Second stage:

phi[1:N] ~ car.normal(adj[], weights[], num[], tau.phi)

Priors:

tau.phi ~ Gamma(taub.alpha, taub.beta)

Model diagram



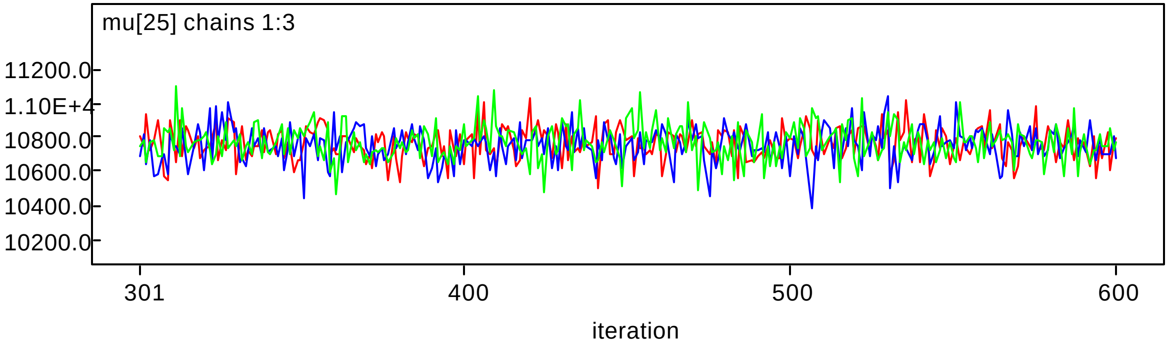
Example of Source Code

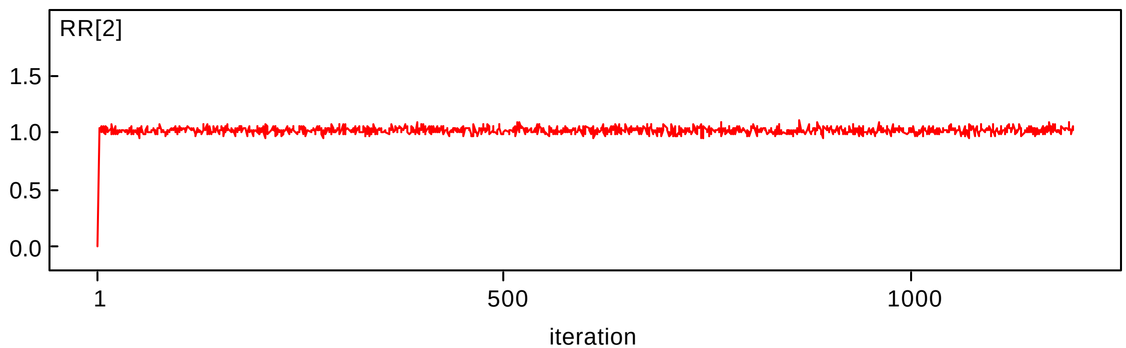
```
model {
    # Likelihood
    for (i in 1 : N) {
        obs.m[i]  ~ dpois(mu[i])
        theta[i]  ~ dnorm(0, tau.theta)
        log(mu[i]) <- log(e.m[i]) + beta0 + beta1*log.emis[i] + phi[i] + theta[i]
        RR[i] <- exp(beta0 + beta1*pov[i] + phi[i] + theta[i])
    }
    # CAR prior distribution for relative risk:
    phi[1:N] ~ car.normal(adj[], weights[], num[], tau.phi)
    for(k in 1:sumNumNeigh) { weights[k] <- 1}
    # Other priors:
    beta0 ~ dflat()
    beta1 ~ dnorm(0.0, 1.0E-5)
    #beta2 ~ dnorm(0.0, 1.0E-5)
    tau.phi  ~ dgamma(0.5, 0.0005)
    tau.theta  ~ dgamma(0.5, 0.0005)
    sigma.phi <- sqrt(1 / tau.phi)
}
```

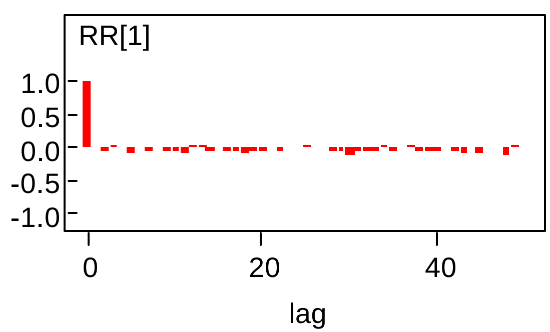
Model selection

Model		DIC	pD
No spatial structured variance		922.3	87.7
Structured & unstructured	CAR	926.5	93.5
	EXP	916.3	83.4
1 covariate log.emis	CAR	923.6	89.7
	EXP	916.7	83.3
1 covariate pov	CAR	924.3	90.6
	EXP	916.2	82.9
2 covariates log.emis and pov	CAR	921.7	88.5
	EXP	917.4	84.5

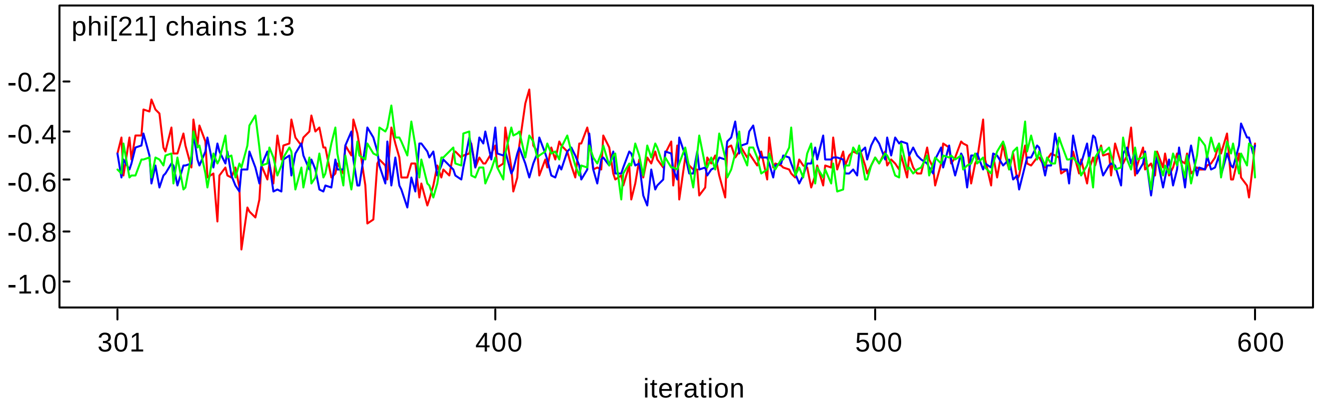
History graphs

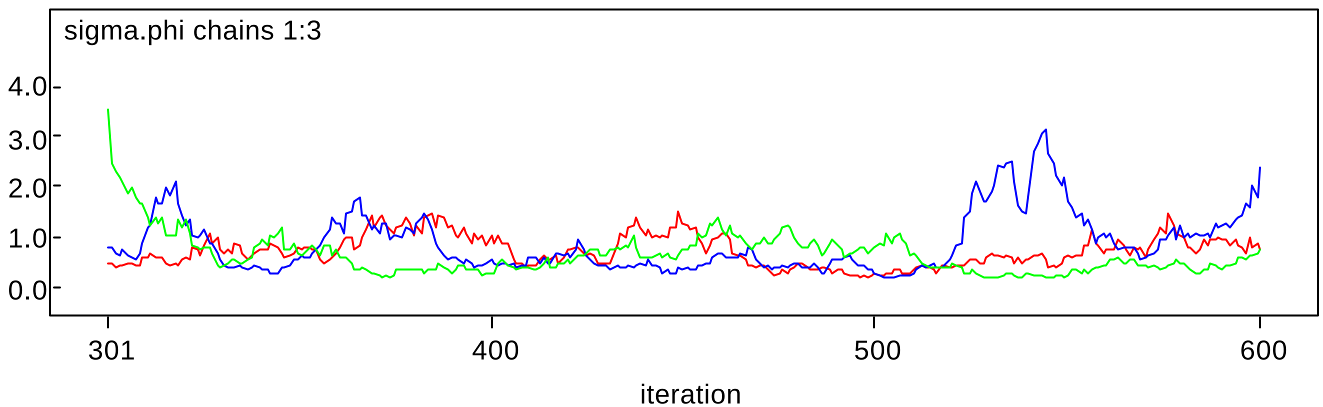






History graphs





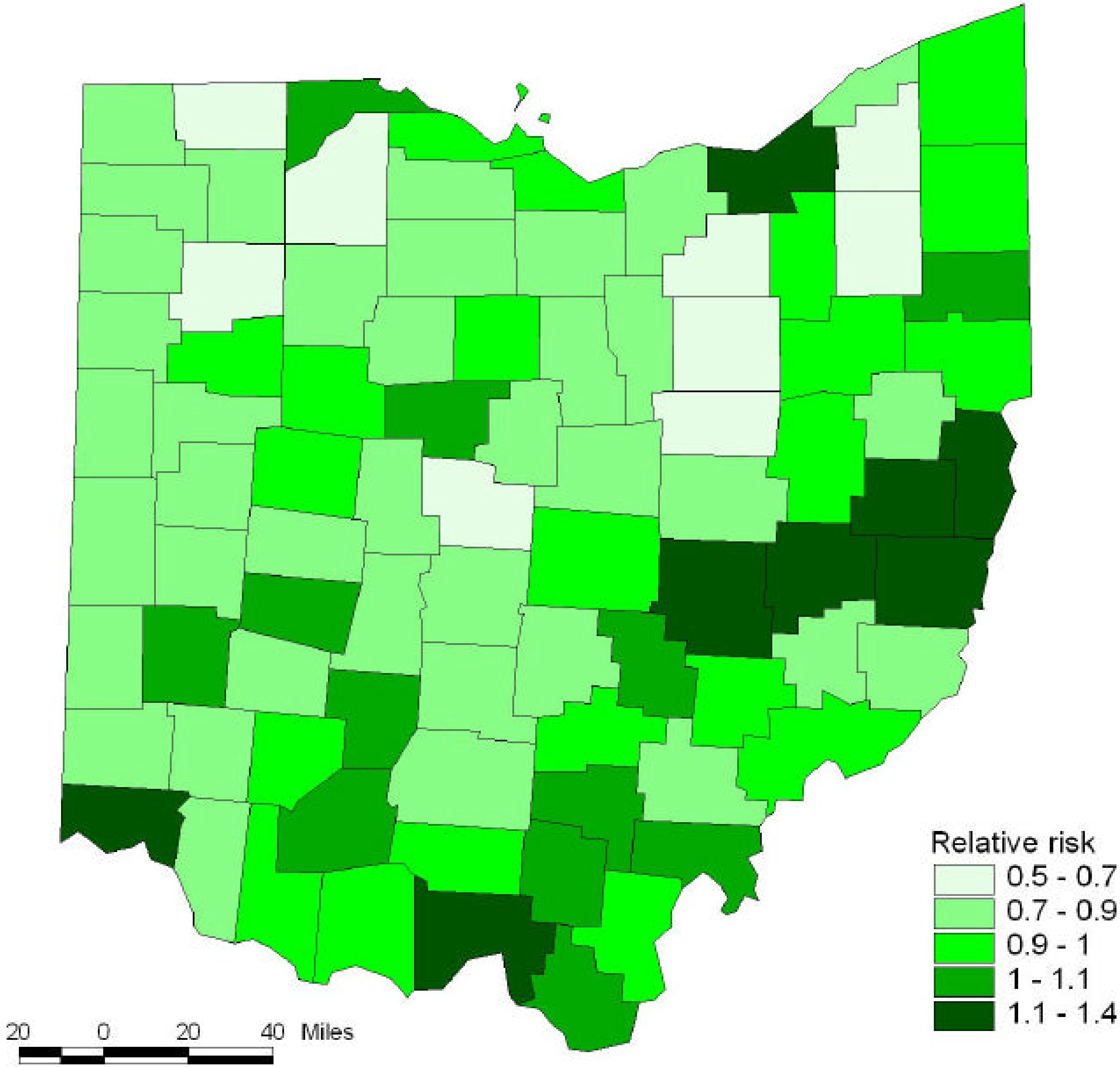
Statistical results

One covariate (pov) with spatial structured variance

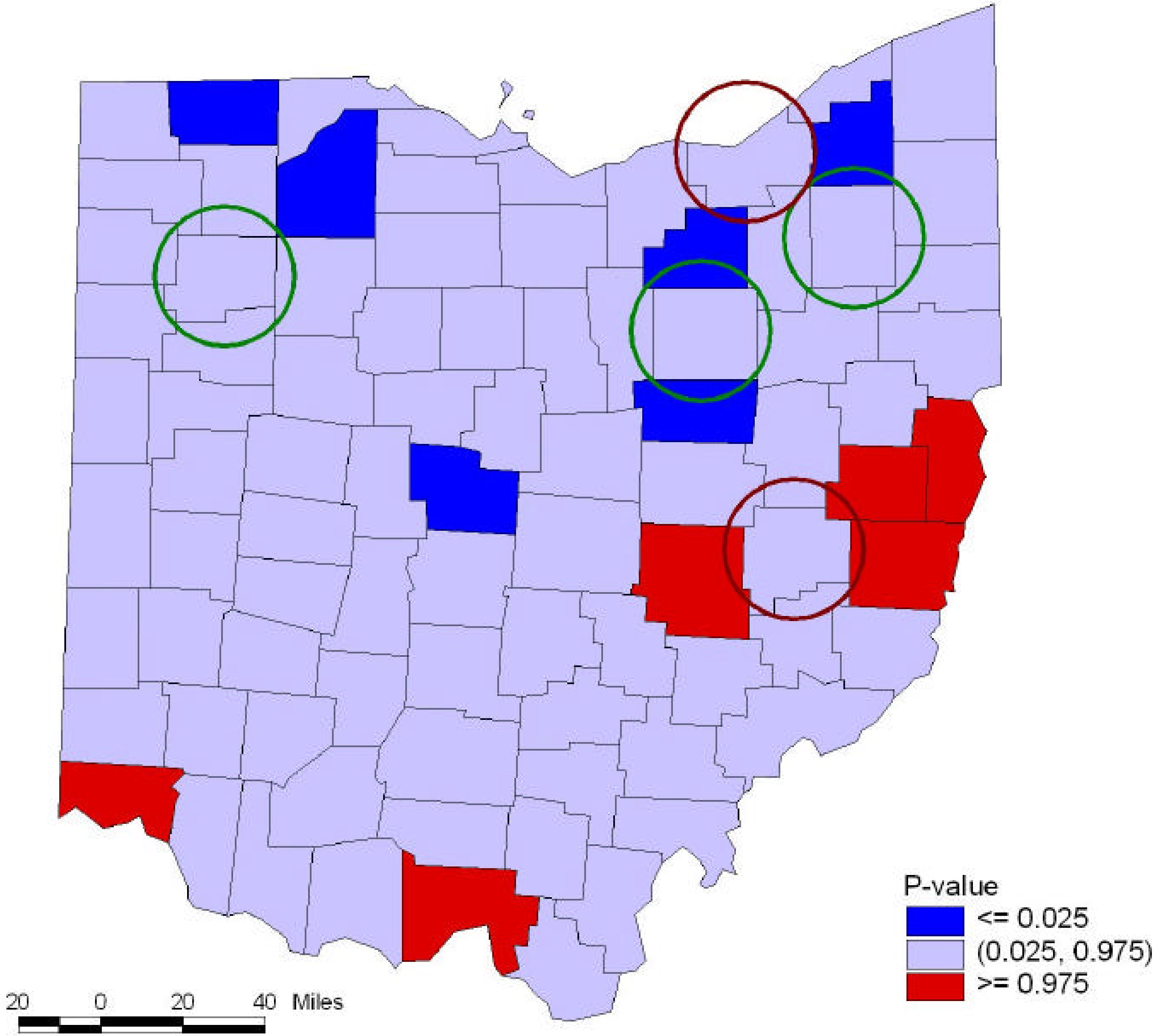
sigma.phi	210.8	56.5	120.3	141.5	234.4	254.6	286.8	7.0	3
deviance	833.3	12.9	809.8	824.7	832.6	841.9	859.4	1.0	900

pD = 82.9 and DIC = 916.2 (using the rule, pD = var(deviance)/2)

Relative risks



P-value of RR>1



Discussion and Conclusions

- Bayesian approach helps create more interpretable map by:
  - Applying priors
  - Incorporate covariates
- p-value map identify several potential hotspots.
- Age and race adjusted rates may be used to compute expected number of cases.
- Space and time