MODULE 12: Spatial Statistics in Epidemiology and Public Health
Lecture 5: Spatial regression

Jon Wakefield and **Lance Waller**

References

- Waller and Gotway (2004, Chapter 9) *Applied Spatial Statistics for Public Health Data*. New York: Wiley.
- Elliott, P., et al. (2000) *Spatial Epidemiology: Methods and Applications*, Oxford: Oxford University Press.
- Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.
- Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2014) *Hierarchical Modeling and Analysis for Spatial Data, 2nd Ed.* Boca Raton, FL: CRC/Chapman & Hall.
- Blangiardo, M. and Cameletti, M. (2015) *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Chichester: Wiley.

What do we have so far?

- Point process ideas (intensities, K-functions).
  - Data: (x, y) event locations.
  - Where are the clusters? Use intensities.
  - How are events clusters? Use K-functions.
- Disease clustering with point data.
- Disease clustering with regional counts.

What's left?

- So we know how to describe and evaluate spatial patterns in health outcome data.
- What about linking patterns in health outcomes to patterns in exposures?
- With *independent* observations we know how to use *linear* and *generalized linear* models such as linear, Poisson, logistic regression.
- What happens with *dependent* observations?

Caveat

"...all models are wrong. The practical question is how wrong do they have to be to not be useful."
Box and Draper (1987, p. 74)

What changes with dependence?

- In statistical modeling, we are often trying to describe the mean of the outcome as a function of covariates, assuming error terms are mutually independent.
- That means we usually model any trend in the data as a trend in *expectations*.
- Allows estimation of covariate effects.
- With *dependent* error terms, observed trends may be due to covariates, correlation, or both.
- May impact the identifiability of covariate effects.
- Could have different effects equally likely under different correlation models.

Residual correlation

- Where do correlated errors come from?
- Perhaps outcomes truly correlated (infectious disease).
- Perhaps we omitted an important variable that has spatial structure itself.
- If temperature is important and we left it out of a model applied to the continental U.S., what would the residuals look like?

Residual maps important

- If high temperatures associated with high outcomes, we would *underfit* in southern states (observations $>$ model predictions $\Rightarrow$ positive residuals), and *overfit* in northern states (observations $<$ model prediction $\Rightarrow$ negative residuals).
- The "missing covariate" idea suggests that *maps* of residuals are important spatial diagnostics.
- Also, we may want to apply tests of clustering or to detect clusters to residuals.
- Moran's I, LISAs.

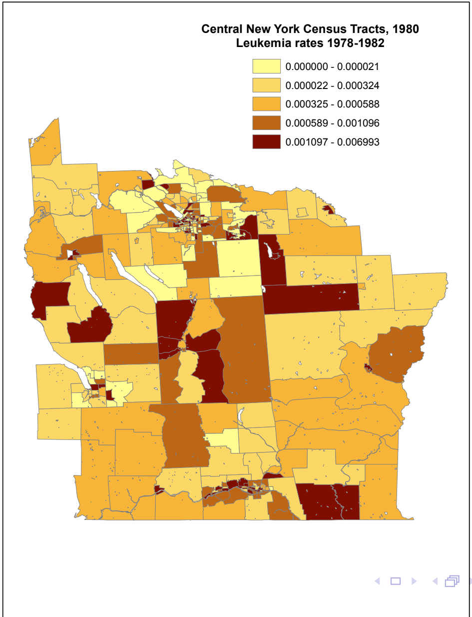
Our plan

- NY leukemia data and add some covariates.
- We will fit linear and Poisson regression models with various spatial correlation structures and compare inferences.
- Remember, all of these models are wrong, but some may be useful.

Illustrating regression models

- New York leukemia data from Waller et al. (1994)
- 281 census tracts (1980 Census).
- 8 counties in central New York.
- 592 cases for 1978-1982.
- 1,057,673 people at risk.

Crude Rates (per 100,000)



Building the model

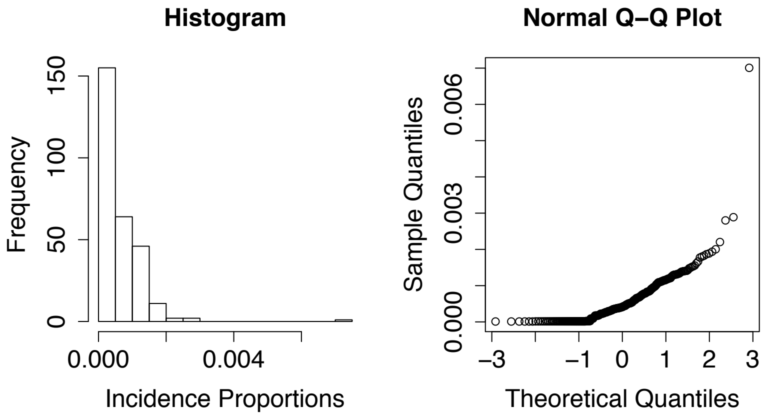
- Let Y_i = count for region i.
- Let E_i = *expected* count for region i.
- $x_{i,TCE}$ = inverse distance to TCE site.
- $x_{i,65}$ = percent over age 65 (census).
- $x_{i,home}$ = percent who own own home (census).
- The model:

$$Y_i = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home} + \epsilon_i.$$

Assumptions for regression

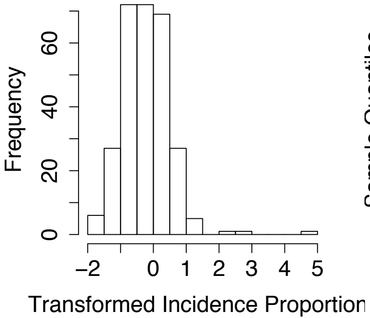
- The error terms, $\epsilon_i \overset{ind}{\sim} N(0, \sigma^2)$;
- The data have a constant variance, σ^2;
- The data are uncorrelated (OLS) or have a specified parametric covariance structure (GLS);

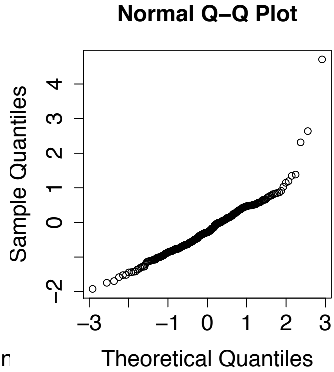
Y normally distributed?



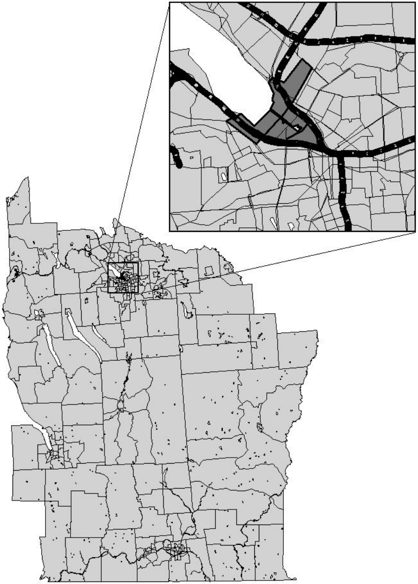
Transformation?

$$Z_i = \log\left(\frac{1000(Y_i + 1)}{n_i}\right).$$

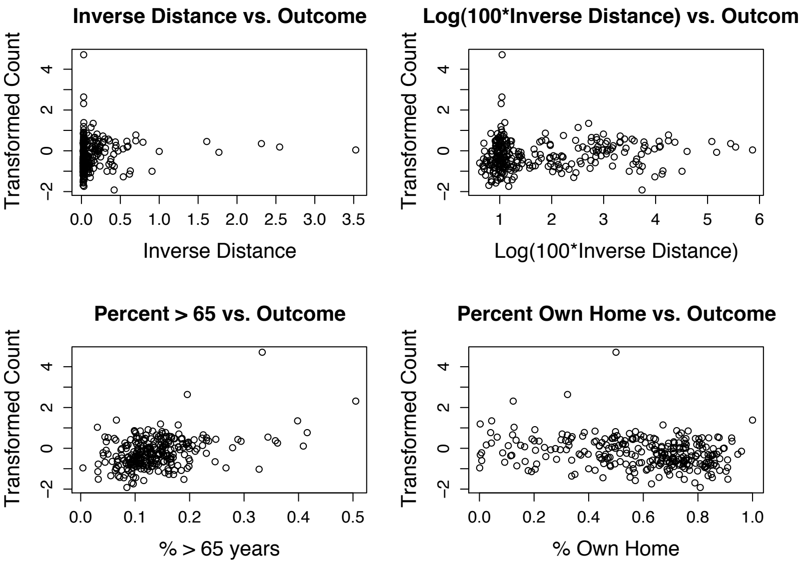




Outliers, where are the top 3?



Scatterplots



Linear Regression (OLS)

Parameter	Estimate	Std. Error	p-value
$\hat{\beta}_0$ (Intercept)	-0.5173	0.1586	0.0012
$\hat{\beta}_1$ (TCE)	0.0488	0.0351	0.1648
$\hat{\beta}_2$ (% Age $>$ 65)	3.9509	0.6055	$<$0.0001
$\hat{\beta}_3$ (% Own home)	-0.5600	0.1703	0.0011
$\hat{\sigma}^2$	0.4318	277 df	
R^2=0.1932	AIC=567.5		

Is OLS appropriate?

- Zs roughly Gaussian (symmetric).
- Do Zs have constant variance?
- No, since population sizes vary.
- $\text{Var}(Z_i) = \text{Var}\left(\log\left(\frac{1000(Y_i+1)}{n_i}\right)\right)$
- Try *weighted least squares* with weights $1/n_i$.

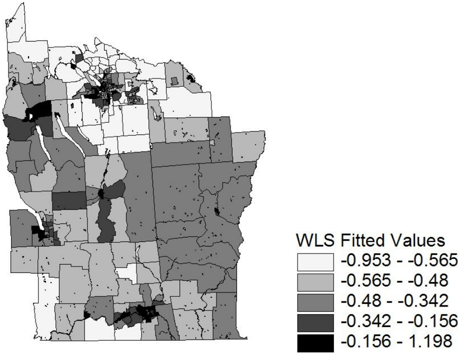
Linear Regression (WLS)

Parameter	Estimate	Std. Error	p-value
$\hat{\beta}_0$ (Intercept)	-0.7784	0.1412	$<$0.0001
$\hat{\beta}_1$ (TCE)	0.0763	0.0273	0.0056
$\hat{\beta}_2$ (% Age $>$ 65)	3.8566	0.5713	$<$0.0001
$\hat{\beta}_3$ (% Own home)	-0.3987	0.1531	0.0097
$\hat{\sigma}^2$	1121.94	277 df	
R^2=0.1977	AIC=513.5		

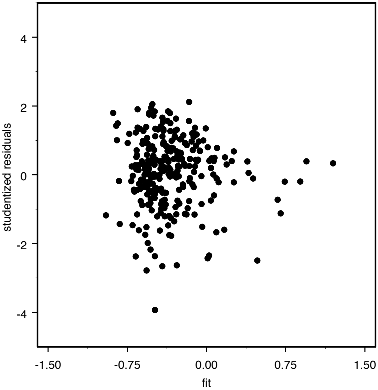
What changed?

- The three outliers are all in regions with small n_i.
- Weighting reduced their impact on estimates.
- Most profound effect is with respect to TCE.

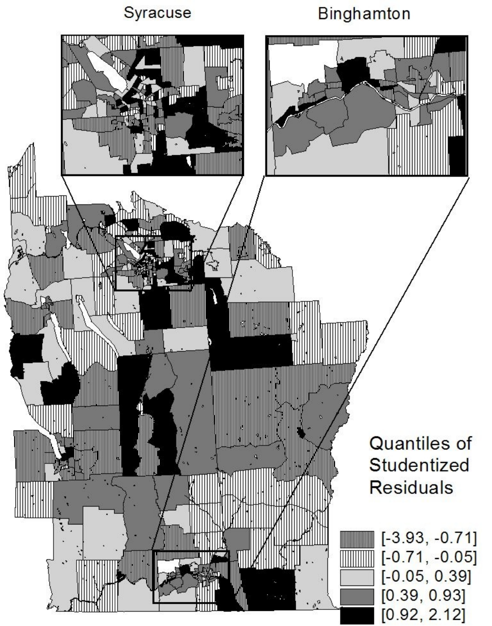
WLS fitted values



Residual plot



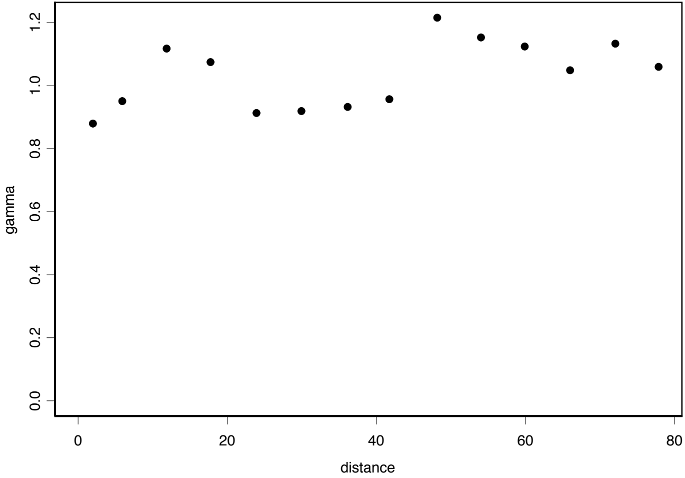
Residual map



What are we looking for?

- Patterns in locations of residuals.
- Model underfit (predictions too low) near cities?
- Correlations in residuals?
- Let's try semivariograms for the residuals.
- Let's try local Moran's I for residuals.

Residual correlation?



Comments

- Residual semivariogram not too impressive.
- We can try *maximum likelihood* fit incorporating residual correlation via the semivariogram (which defines covariance matrix).

Linear Regression, Correlated Errors (ML)

Parameter	Estimate	Std. Error	p-value
$\hat{\beta}_0$ (Intercept)	-0.7222	0.1972	$<$0.0001
$\hat{\beta}_1$ (TCE)	0.0826	0.0434	0.0576
$\hat{\beta}_2$ (% Age $>$ 65)	3.7093	0.6188	$<$0.0001
$\hat{\beta}_3$ (% Own home)	-0.3245	0.2044	0.1136
$\hat{c}_0$=0.3740	$\hat{c}_s$=0.0558	$\hat{a}$=6.93	
AIC=565.6	277 df		

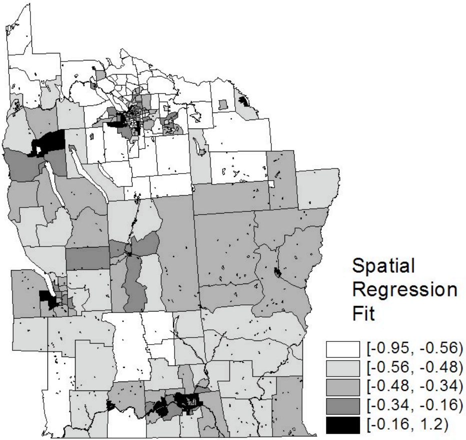
Weighting?

- We also need to include weights to account for heteroskedasticity.
- Again we use weights equal to $1/n_i$.
- What changes?

Linear regression, Correlated, Weighted

Parameter	Estimate	Std. Error	p-value
$\hat{\beta}_0$ (Intercept)	-0.9161	0.1648	$<$0.0001
$\hat{\beta}_1$ (TCE)	0.0956	0.0322	0.0032
$\hat{\beta}_2$ (% Age $>$ 65)	3.5763	0.5920	$<$0.0001
$\hat{\beta}_3$ (% Own home)	-0.2285	0.1761	0.1956
$\hat{c}_0$=997.65	$\hat{c}_s$=127.12	$\hat{a}$=6.86	
AIC=514.7	277 df		

Fitted values (correlated, weighted)



Modelling counts directly

- Using linear regression required a fair amount of data transformation, just to meet modelling assumptions.
- Can we model the counts directly?
- In epidemiology, common to use logistic or Poisson regression.
- For rare disease, little difference between logistic and Poisson.
- Both are examples of *generalized linear models* (McCullagh and Nelder, 1989).

Building the model

- Let Y_i = count for region i.
- Let E_i = *expected* count for region i.
- Let $(x_{i,TCE}, x_{i,65}, x_{i,home})$ be the associated covariate values.
- Poisson regression:

$$Y_i \sim Poisson(E_i \zeta_i)$$

where

$$\log(\zeta_i) = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home}.$$

What's different?

- Poisson distribution for counts, rather than transforming proportions for normality.
- *Link function*: Natural log of mean of Y_i is a linear function of covariates.
- So βs represent multiplicative increases in expected counts, e^{β} a measure of relative risk associated with one unit increase in covariate.
- E_i an *offset*, what we expect if the covariates have no impact.
- Age, race, sex adjustments in either E_i (standardization) or covariates.

How do we add spatial correlation?

- Trickier than in regression, since mean and variance are related for Poisson observations.
- Two general approaches:
  - *Marginal specification* defining correlation among means.
  - *Conditional specification* defining correlation through the use of *random effects*.

Marginal and conditional models

- We often think of a model representing the *marginal mean*, $E(\mathbf{Y})$ as a function of fixed, unknown parameters.
- That is, the parameters define the *population average* effect of the covariates ("On average, how does a given level of air pollution impact a person?")
- Another approach is to consider a model of the *conditional mean* for each subject.
- In this setting we think of fixed effects of parameters and *random* effects specific to the subjects.

Marginal versus conditional interpretation

- For us: *fixed effects* apply equally to all subjects, *random effects* apply to a particular subject.
- Interpret fixed effects *conditional on* levels of the random effects.
- "What is the effect of aspirin on a headache averaged over all individuals in the study?" (Marginal effect).
- "What is the effect of aspirin on a headache in this individual?" (Conditional effect).
- Random effects allow different parameter values for individuals, following some distribution.

Random intercepts

- A model with fixed and random effects is a *mixed* model.
- A very common formulation is to have fixed parameter values and a *random intercept*. This says everyone has the same response to the treatment, but that individuals have different starting points.
- In Poisson regression setting, if we add random effects we generate a *generalized linear mixed model* (GLMM).

Random effects and the conditional specification

- We add a *random effect* (intercept).
- Represents an impact of region i, not accounted for in E_i or the covariates.
- We define this random effect to have a *spatial* distribution.

Building the model

- Let Y_i denote the *observed* number of cases in region i.
- Let E_i denote the *expected* number of cases, *ignoring covariate effects*.
- Assume E_i known, perhaps age-standardized, or based on global (external or internal) rates.
- First stage:

$$Y_i|\zeta_i \stackrel{ind}{\sim} \text{Poisson}(E_i\zeta_i)$$

- ζ_i represent a relative risk associated with region i *not accounted for by the* E_i.

Building the model

- Note $Y_i/E_i = SMR_i$, the MLE of ζ_i.
- Also note, $E[Y_i|\zeta_i] \neq E_i$, since E_i does not include the impact of the random effect.
- Create a GLMM with log link by

$$\log(E[Y_i|\zeta_i]) = \log(E_i) + \log(\zeta_i)$$

- If we add covariates and rename $\log(\zeta_i) = \psi_i$, then

$$\log(\zeta_i) = \mathbf{x}_i'\boldsymbol{\beta} + \psi_i$$

New York data

- So our model is

$$Y_i|\boldsymbol{\beta}, \psi_i \stackrel{ind}{\sim} \text{Poisson}(E_i \exp(\mathbf{x}_i'\boldsymbol{\beta} + \psi_i)),$$

$$\log(\zeta_i) = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home} + \psi_i.$$

- The ψ_i represent the *random interecpts*.
- Add *overdispersion* via $\boldsymbol{\psi}_i \stackrel{ind}{\sim} N(0, v_\psi)$.
- Add spatial correlation via

$$\boldsymbol{\psi} \sim MVN(\mathbf{0}, \Sigma).$$

Priors and "shrinkage"

- Overdispersion model (i.i.d. ψ_i) results in each estimate being a compromise between the *local* SMR and the *global average* SMR.
- "Borrows information (strength)" from other observations to improve precision of local estimate.
- "Shrinks" estimate toward global mean. (Note: "shrink" does not mean "reduce", rather means "moves toward").

Local shrinkage

- Spatial model (correlated ψ_i) results in each estimate begin a compromise between the *lcoal* SMR and the *local average* SMR.
- Shrinks each ψ_i toward the average of its *neighbors*.
- Can also include *both* global and local shrinkage (Besag, York, and Mollié 1991).
- How do we fit these models?

Bayesian inference regarding model parameters based on *posterior distribution*

$$Pr[\boldsymbol{\beta}, \boldsymbol{\psi} | \mathbf{Y}]$$

proportional to the product of the likelihood times the prior

$$Pr[\mathbf{Y} | \boldsymbol{\beta}, \psi] Pr[\psi] Pr[\beta].$$

Defers spatial correlation to the prior rather than the likelihood.

Spatial priors

- Could model *joint* distribution

$$\boldsymbol{\psi} \sim MVN(\mathbf{0}, \Sigma).$$

- Could also model *conditional* distribution

$$\psi_i | \psi_{j \neq i} \sim N\left(\frac{\sum_{j \neq i} c_{ij} \psi_j}{\sum_{j \neq i} c_{ij}}, \frac{1}{v_{CAR} \sum_{j \neq i} c_{ij}}\right), i = 1, \ldots, N.$$

  where c_{ij} are *weights* defining the neighbors of region i.
- Adjacency weights: $c_{ij} = 1$ if j is a neighbor of i.

CAR priors

- The conditional specification defines the *conditional autoregressive* (CAR) prior (Besag 1974, Besag et al. 1991).
- Under certain conditions on the c_{ij}, the CAR prior defines a valid multivariate joint Gaussian distribution.
- Variance covariance matrix a function of the *inverse* of the matrix of neighbor weights.

Perspective: Generalized linear mixed model

- Given the values of the random effects (ψ_is), observations (Y_is) are independent.
- Taking into account correlation in the ψ_is, the Y_is are correlated.
- Conditionally independent $Y_i|\psi_i$ give *likelihood* function.
- (Spatially correlated) distribution of the ψ_is a *prior distribution*.

Fitting Bayesian models: Markov chain Monte Carlo

- Posterior often difficult to calculate mathematically.
- Iterative simulation approach to model fitting.
- Given *full conditional* distributions, simulate a new value for each parameter, holding the other parameter values fixed.
- The set of simulated values converges to a sample from the posterior distribution.
- `WinBUGS` software.
  `www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`

Conceptual MCMC example

- Suppose we have a model with data $\mathbf{Y}$ and three parameters θ_1, θ_2, and θ_3.
- "Gibbs sampler" simulates values from the *full conditional* distributions

$$
\begin{gathered}
f(\theta_1|\theta_2, \theta_3, \mathbf{Y}), \\
f(\theta_2|\theta_1, \theta_3, \mathbf{Y}), \\
f(\theta_3|\theta_1, \theta_2, \mathbf{Y}).
\end{gathered}
$$

Conceptual MCMC

- Start with values $\theta_1^{(1)}$, $\theta_2^{(1)}$, and $\theta_3^{(1)}$.

$$\text{sample } \theta_1^{(2)} \text{ from } f(\theta_1|\theta_2^{(1)}, \theta_3^{(1)}, \mathbf{Y}),$$
$$\text{sample } \theta_2^{(2)} \text{ from } f(\theta_2|\theta_1^{(2)}, \theta_3^{(1)}, \mathbf{Y}),$$
$$\text{sample } \theta_3^{(2)} \text{ from } f(\theta_3|\theta_1^{(2)}, \theta_2^{(2)}, \mathbf{Y}).$$

- As we continue to update $\boldsymbol{\theta}$, sampled values become indistinguishable from a sample from the joint posterior distribution $f(\theta_1, \theta_2, \theta_3|\mathbf{Y})$.

MCMC example

- Gelman et al. (2004). Theoretical and MCMC results.

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim MVN\left(\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

- Uniform priors on θ_1, θ_2, yield posterior

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \sim MVN\left(\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

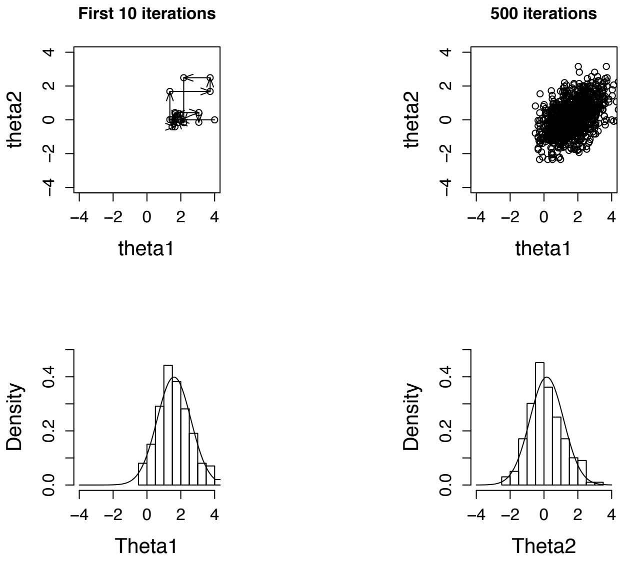
Full conditionals

- Multivariate results give *full conditionals*

$$\begin{aligned} \theta_1|\theta_2, \mathbf{Y} &\sim N(Y_1 + \rho(\theta_2 - Y_2), 1 - \rho^2), \\ \theta_2|\theta_1, \mathbf{Y} &\sim N(Y_2 + \rho(\theta_1 - Y_1), 1 - \rho^2). \end{aligned}$$

- Let's try a Gibbs sampler and compare to the theoretical results.

MCMC example



Back to CAR prior

- Almost custom-made for MCMC.
- Defined for ψ_i, given ψ_j for $j \neq i$.
- We define neighborhood weights c_{ij}.

Complete model specification

$$Y_i|\boldsymbol{\beta}, \psi_i \overset{ind}{\sim} \text{Poisson}(E_i \exp(\mathbf{x}_i'\boldsymbol{\beta} + \psi_i)),$$

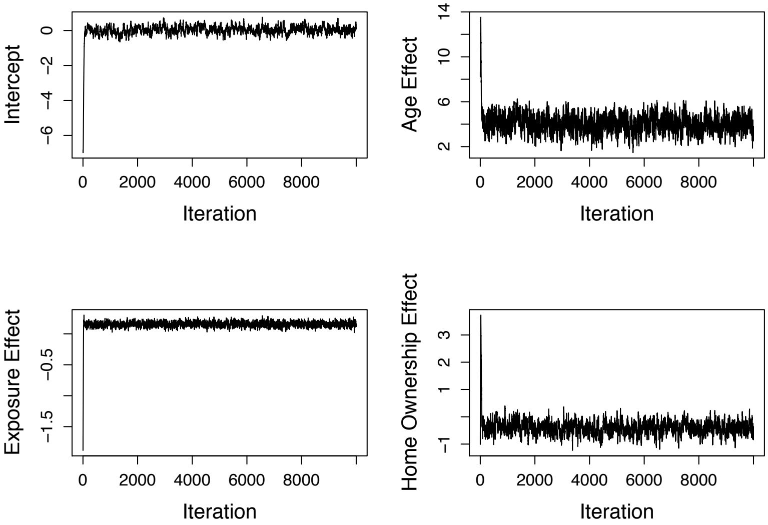
$$\log(\zeta_i) = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home} + \psi_i.$$

$$\beta_k \sim \text{Uniform}.$$

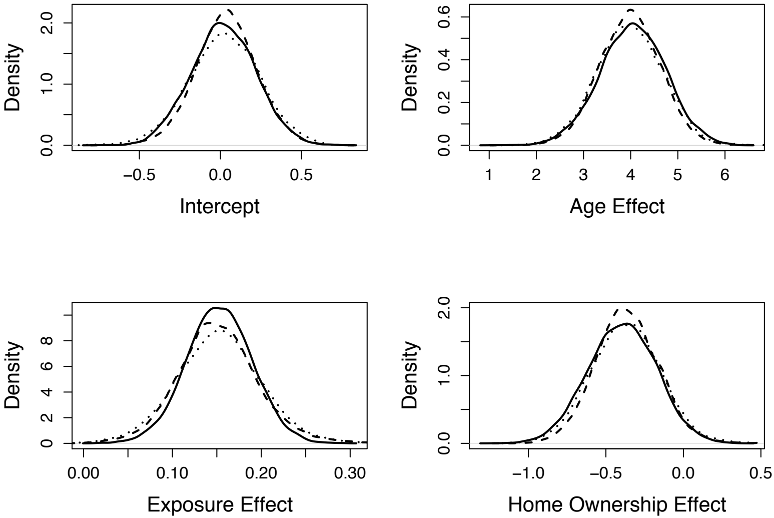
$$\psi_i|\psi_{j\neq i} \sim N\left(\frac{\sum_{j\neq i} c_{ij}\psi_j}{\sum_{j\neq i} c_{ij}}, \frac{1}{v_{CAR}\sum_{j\neq i} c_{ij}}\right), i = 1, \ldots, N.$$

$$\frac{1}{v_{CAR}} \sim \text{Gamma}(0.5, 0.0005).$$

MCMC trace plots



Posterior densities



MCMC posterior estimates

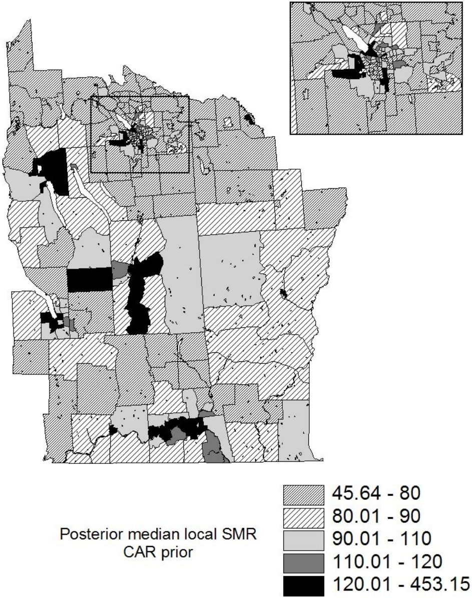
Covariate	Posterior Median	95% Credible Set
β_0	0.048	(-0.355, 0.408)
β_{65}	3.984	(2.736, 5.330)
β_{TCE}	0.152	(0.066, 0.226)
β_{home}	-0.367	(-0.758, 0.049)

But there's more!

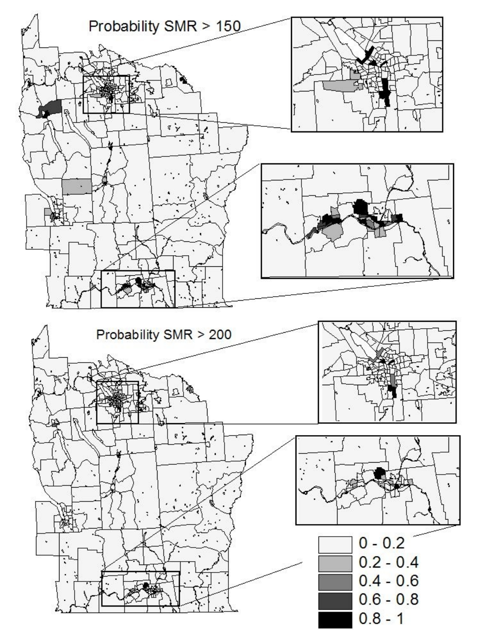
- A nifty thing about MCMC estimates:

  We get posterior samples from any function of model parameters by taking that function of the sampled posterior parameter values.
- Gives us posterior inference for $SMR_i = Y_{i,fit}/E_i$.
- Also can get $Pr[SMR_i > 200|\mathbf{Y}]$ and map these *exceedence probabilities*.

Posterior median SMRs



Posterior exceedence probabilities



Example 2

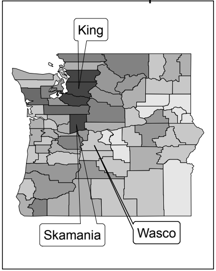
- Cryptozoology Example: Waller and Carlin (2010) Disease Mapping. In *Handbook of Spatial Statistics*, Gelfand et al. (eds.). Boca Raton: CRC/Chapman and Hall.

Cryptozoology example

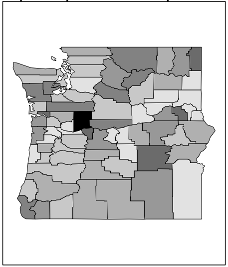
- County-specific reports of encounters with *Sasquatch* (Bigfoot).
- "...which brings us to the appropriateness of the Bigfoot example."
- Data downloaded from `www.bfro.net`
- Sightings from counties in Oregon and Washington (Pacific Northwest).
- Probability of report related to population density?
- (Hopefully) rare events in small areas.
- Perhaps spatial smoothing will stabilize local rate estimates.
- Fit models with no random effects, exchangeable random effects, CAR random effects, convolution random effects.

Sasquatch Data

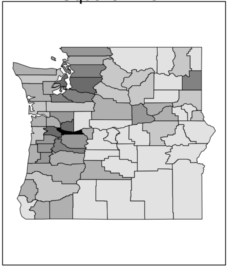
Number of Reports



Reports per 2000 Population



2000 Population per
Square Mile



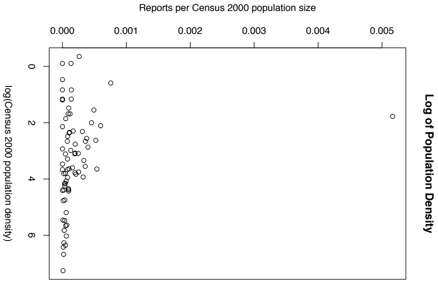
Legend

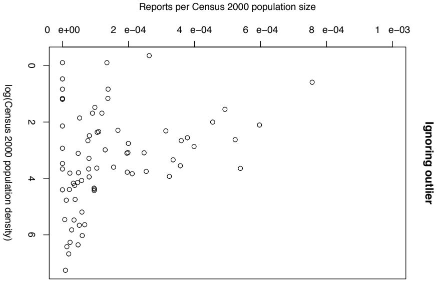
	Reports	Reports/ Person	Population/ Sq. Mi.
	0	0.00000 - 0.00003	0.7 - 12.9
	1-5	0.00003 - 0.00008	13.0 - 32.1
	6-10	0.00008 - 0.00016	32.2 - 69.9
	11-15	0.00016 - 0.00026	70.0 - 180.1
	16-20	0.00026 - 0.00046	180.2 - 414.0
	21-25	0.00046 - 0.00076	414.1 - 793.3
	25-51	0.00076 - 0.00517	793.4 - 1419.3

0 100200 400 600 800

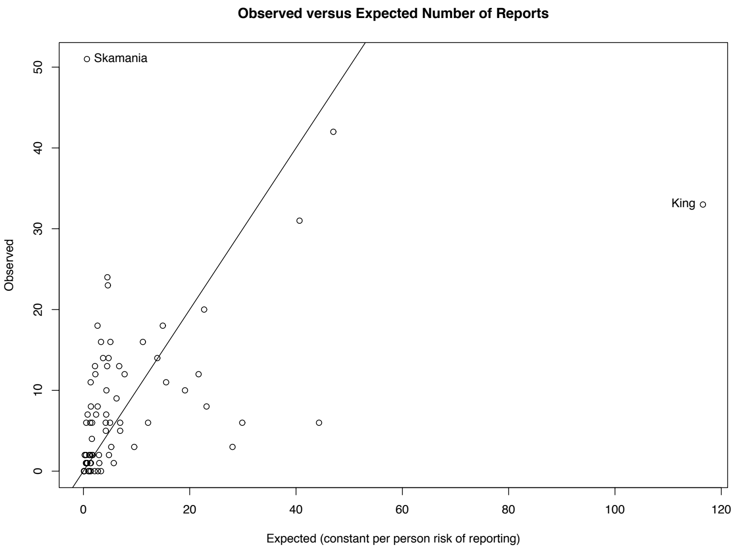
Kilometers

Reports vs. Population Density



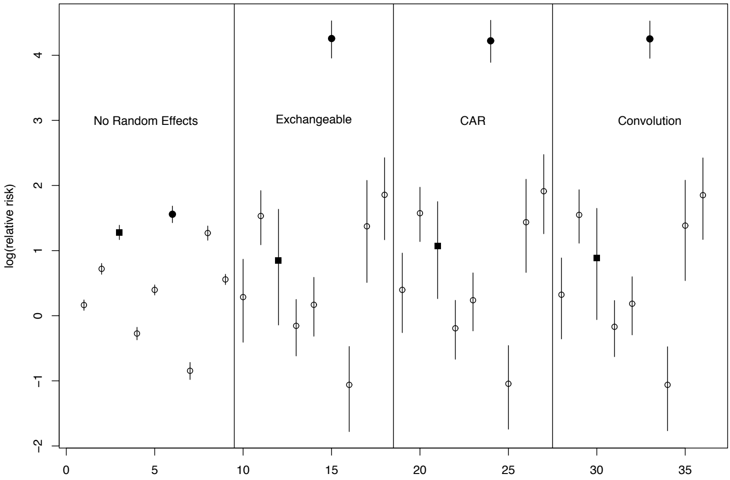


Observed vs. Expected

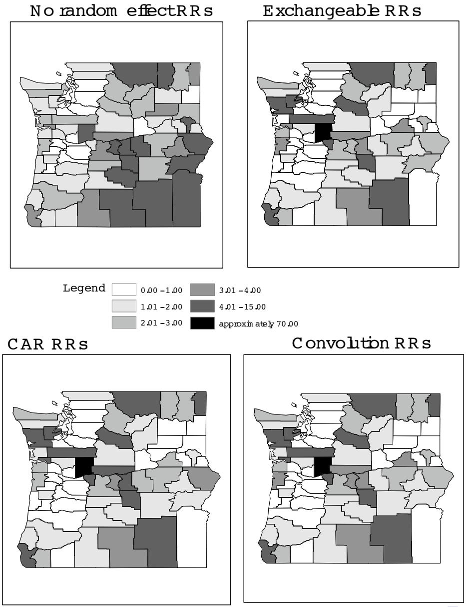


Predicted relative risks and credible sets

Filled circle = Skamania, Filled square = Wasco



Mapped relative risks



Skamania Sasquatch Ordinances

- http://www.skamaniacounty.org/commissioners/homepage/ordinances-2/
- Big Foot Ordinance 69-1: "THEREFORE BE IT RESOLVED that any premeditated, willful and wonton slaying of any such creature shall be deemed a felony punishable by a fine not to exceed Ten Thousand Dollars ($10,000.00) and/or imprisonment in the county jail for a period not to exceed Five (5) years. ADOPTED this 1st day of April, 1969."
- Big Foot Ordinance 1984-2:
  - Repealed felony and jail sentence.
  - Established a Sasquatch Refuge (Skamania County).
  - Clarified penalty (gross misdemeanor vs. misdemeanor) and penalty (fine and jail time), disallowed insanity defense, and clarified distinction between coroner designation of victim as humanoid (murder) or anthropoid (this ordinance).

Conclusions

- What method to use depends on what you want data you have and what question you want to answer.
- All methods try to balance trend (fixed effects) with correlation (here, with random effects).
- All models wrong, some models useful.
- Trying more than one approach often sensible.
- Few methods (including Monte Carlo simulation) in current GIS packages.