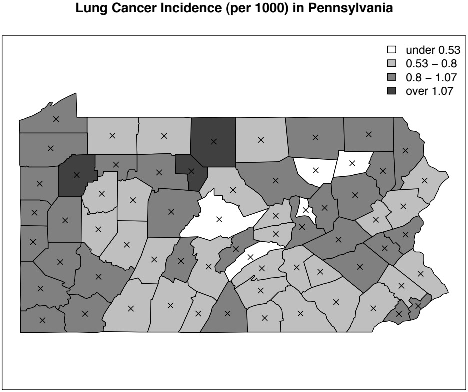
# A Bayesian Method for Disease Cluster Detection


Albert Y. Kim

Middlebury College
Middlebury VT


August 2015

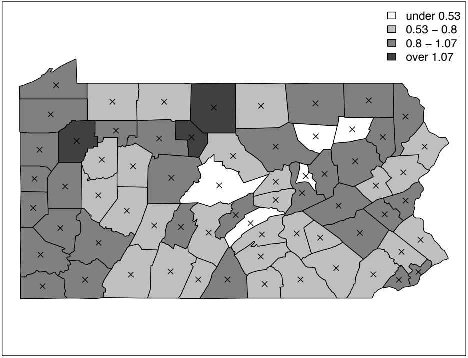
# Scenario



Lung Cancer Incidence (per 1000) in Pennsylvania
under 0.53
0.53 – 0.8
0.8 – 1.07
over 1.07

# Scenario

Lung Cancer Incidence (per 1000) in Pennsylvania

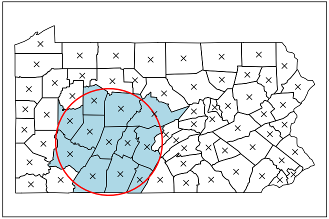



Disease counts $Y_i$ are modeled via Poisson$(E_i \times \theta_i)$ where

- $E_i =$ *expected numbers of disease*
- $\theta_i =$ *relative risk of disease*. $\theta_i > 1$ suggests high risk.
- Point estimate $\widehat{\theta}_i$ is the *standardized mortality ratio* $\frac{y_i}{E_i}$

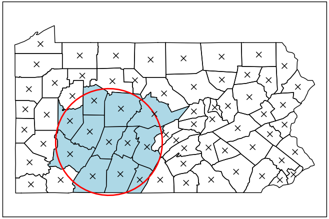
# Geographic Units

We define a set of single zones $z$
by placing circles on the map.
Each $z$ is a candidate cluster.

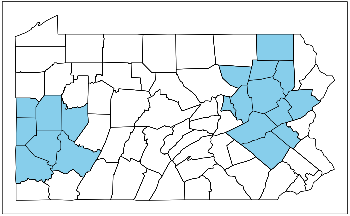


# Geographic Units

We define a set of single zones $z$ by placing circles on the map. Each $z$ is a candidate cluster.

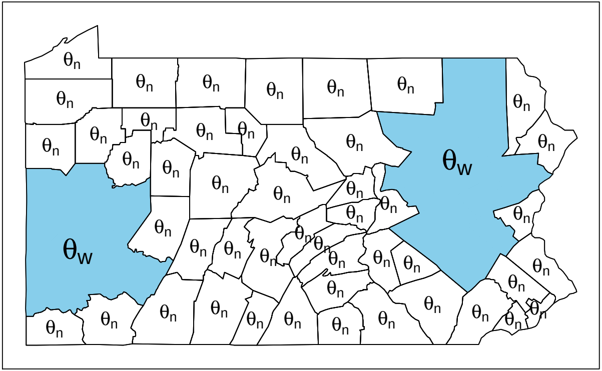


We define a set of configurations $c$: combinations of single zones. Each $c$ is a candidate explanatory model for $\vec{y}$



# Explanatory model for $\vec{y}$

Based on a particular $c$, we derive an explanatory model for $\vec{y}$:



# Posterior Probability of a Configuration

The (discrete) parameter of interest is the configuration $c$ with posterior probability:

$$\pi(c|\vec{y}) = \frac{p(\vec{y}|c)\pi(c)}{\sum_c p(\vec{y}|c)\pi(c)}$$

# Posterior Probability of a Configuration

The (discrete) parameter of interest is the configuration $c$ with posterior probability:

$$\pi(c|\vec{y}) = \frac{p(\vec{y}|c)\pi(c)}{\sum_c p(\vec{y}|c)\pi(c)}$$

Problem: Intractable # of configurations.

# Posterior Probability of a Configuration

The (discrete) parameter of interest is the configuration $c$ with posterior probability:

$$\pi(c|\vec{y}) = \frac{p(\vec{y}|c)\pi(c)}{\sum_c p(\vec{y}|c)\pi(c)}$$

Problem: Intractable # of configurations.

Solution: Use MCMC to estimate $\pi(c|\vec{y})$. Specifically a Metropolis-Hastings algorithm with proposal function $Q$.

# Five Types of Moves in Proposal Function $Q$