August 3, 2003
University of Wisconsin – Madison
Department of Biostatistics and Medical Informatics

Ronald E. Canzian and Murray K. Clayton

Covariates
Case-Control Data in the Presence of Detecting Spatial Clustering with Covariates

Detecting Spatial Clustering with Covariates

ftp://www.biotstat.wisc.edu/~ronald
Fitted values from logistic regression available.

History, alcohol use, BMI.

Age, parity, age at menarche, education level, family.

Information on known breast cancer risks available.

- Cases & controls geocoded to residence.

- 471 case & 471 controls in Dane County, Wisconsin.

- Subset of larger case-control study of breast cancer in Dane County, Wisconsin.

Dane County Breast Cancer Data
Map of Dane County
Goals of Analysis

- Is there evidence of clustering of cases?
- If so, where are the clusters located?
- If so, what are the risks associated with the clusters?
- Interest is clustering NOT explained by known risk factors.
standardization

Covariate adjustment easily accomplished using indirect

Denoted by \( LR_e \).

Find likelihood-ratio for the two-parameter Poisson model.

One rate inside cluster, another rate outside cluster.

Each cluster defines a two-parameter Poisson model.

Define a large set of potential clusters.

(census tracts, zip codes, etc.).

Observe number of cases and population at risk for regions

Cluster Detection for Case-Count Data

Detecting Spatial Clustering with Covariates

Gangnon & Clayton
Posterior probability of cluster $c = \frac{w^c \cdot \text{TR}}{\text{WALT}}$

\[ \text{WALT} = \sum_{c} w^c \cdot \text{TR} \]

- Wight (prior probability) for cluster $c$, $w^c$

- WALT statistic (Cangnan & Clayton, 2000)

• Associated cluster is MLE.

\[ \text{LR}_{\text{max}} = \max \text{LR}^c \]

- Scan statistic (Kulldorff, 1997)

Likelihood-Based Test Statistics

Cangnan & Clayton

Detecting Spatial Clustering with Covariates
Use unconditional sampling instead.

• Conditional sampling given number of cases very difficult.

Use approximations to minimize computations.

Iterative algorithms needed to find likelihood ratio.

• Covariate adjustment is more difficult

Poisson likelihood replaced with binomial likelihood.

Adaptations to Case-Control Data
Potential Clusters for Dane County Data

- Fixed number of cases (Besag and Newell, 1991)
- Fixed number of subjects (Turnbull et al., 1990)
- Maximum number of subjects (KNN, 1995)

- Many other choices available.
- Subject belongs to cluster if centroid falls inside circle.
- Radius ranging from 0 km up to 5 km.
- Circular clusters centered at subjects (cases/controls).
(no clustering).

\[ H \quad \bullet \]

\( \beta \text{ Binomial}(\theta) \quad \bullet \)

\[ \begin{align*}
&\text{Outside cluster: } \logit(p) = \logit(0) + (1) \\
&\text{Inside cluster: } \logit(p) = \logit(0) + (1)
\end{align*} \quad \bullet \]

Consider a two-parameter logistic regression model.

\[ \hat{p} = \text{predicted probability based on covariates.} \quad \bullet \]

\[ \hat{y}_i = \text{case/control indicator.} \quad \bullet \]

\[ i = 1, 2, \ldots, N \text{ subjects.} \quad \bullet \]

Model for Clusters

Detecting Spatial Clustering with Covariates

Gagnon & Clifton
Use similar approach to evaluate outside cluster likelihood.

Evaluate exact likelihood using 2-step estimate.

Update estimate of cluster risk and evaluate estimate of cluster likelihood.

1. If $|\hat{\theta}|$ is small, find one-step estimate of cluster risk.
2. If $|\hat{\theta}|$ is large, use quadratic approximation to likelihood.

For the inside cluster likelihood:

**Approximate Likelihoods for Clusters**
Why weight clusters?

- Weight cluster using probability of selection.
- Select cluster radius uniformly from available radii.
- Center the cluster at subject location.

Density

- Probability inversely proportional to estimated population
- Select a subject at random.

Weights for Clusters

Detecting Spatial Clustering with Coordinates

Gangnan G. Chajin
Probability Subject Belongs to Cluster (unweighted)
(Weighted)

Probability Subject Belongs to Cluster

Detecting Spatial Clustering with Covariates

Gangnan & Chajison
Impact of covariate adjustment is minimal.

\[
\text{VATR statistic } 7.15 (d = 0.019),
\]

\[
\text{Scan statistic } 7.26 (d = 0.293),
\]

\[
\text{VATR statistic } 7.80 (d = 0.033),
\]

\[
\text{Scan statistic } 7.21 (d = 0.462),
\]

Without covariate adjustment.

Cluster Detection Test Results
Likelihood

Adequacy of Approximation to

Detecting Spatial Clustering with Covariates

Gangan E. Chajib

Scan Statistic (Exact LR)

WALR Statistic (Exact LR)

Difference (Approx LR - Exact LR)

Difference (Approx LR - Exact LR)

4 3 2 1

-0.05 -0.04 -0.03 -0.02 -0.01 0.0

4 3 2 1

-0.12 -0.08 -0.04 0.0
Quantile of Scan Statistic (Conditional Sampling)

Quantile of WALR Statistic (Conditional Sampling)
Formal estimation of cluster location based on posterior.

Power comparisons.

Future work:

Impact of unconditional vs. conditional sampling is minimal.

Approximation to binomial likelihood fast and accurate.

Conclusion: Remarks

Detecting Spatial Clustering with Coarse Data

Gangnan & Chiu et al.