A Weighted Average Likelihood Ratio Scan Test for Spatial Clustering of Disease

Ronald E. Gangnon

and

Murray K. Clayton

Department of Biostatistics and Medical Informatics
Department of Statistics
University of Wisconsin – Madison
March 28, 2001

http://www.biostat.wisc.edu/~ronald
Questions to Answer in Clustering Studies

- Is there evidence of clustering?
  
  Small areas with high (or low) leukemia rates.

- Where is the cluster located?

- What is the risk associated with the cluster?

- Why?
  
  - Screening.
  
  - Surveillance.
  
  - Hypothesis generation.
New York Leukemia Data
Tests for Spatial Clustering

- Distance-based statistics,
  e.g., Whittemore et al. (1987), Bonetti (2000).

- Join count statistics.

- Multiple cluster locations/sizes.
  - Openshaw et al. (1988)
  - Turnbull et al. (1990)
  - Besag and Newell (1991)
  - Gangnon and Clayton (2000)
Basic Statistical Model

- \( i = 1, 2, ..., N \) cells.
- \( O_i \) = number of cases in cell \( i \).
- \( n_i \) = population at risk in cell \( i \).
- \( r_i = \mathbb{E}(\frac{O_i}{n_i}) \) = true rate of disease in cell \( i \).
- \( O_i \sim \text{Poisson}(r_i \cdot n_i) \).
- \( H_0 : r_i \equiv r \) (no clustering)
Potential Clusters for New York Data

- Circular clusters centered at cell centroids.
- Radii ranging from 0 km up to a maximum geographic radius.
- Cell belongs to cluster if centroid falls inside circle.
- Many other choices available.
  
  Maximum population radius (K&N, 1995)
  
  Fixed population radius (Turnbull, et al, 1990)
  
  Fixed case radius (Besag and Newell, 1991)
Models/Likelihoods for Clusters

For cluster $i, j$,

1. Consider a two-parameter Poisson model

   $$r_i = \lambda_1 \text{ inside cluster;}$$

   $$r_i = \lambda_0 \text{ outside cluster.}$$

2. Calculate the LR test statistic for $H_0 : \lambda_1 = \lambda_0$:

   $$LR_{i,j} = \left( \frac{O_{i,j}/n_{i,j}}{O_t/n_t} \right)^{O_{i,j}} \left( \frac{(O_t - O_{i,j})/(n_t - n_{i,j})}{O_t/n_t} \right)^{(O_t - O_{i,j})}$$
Weights for Clusters

- Select a cell by throwing a dart at study region.
- Center the cluster at centroid of that cell.
- Select cluster radius uniformly from available radii.
- Weight for cluster \(i, j\):

\[
 w_{i,j} = \frac{a_i}{A} \cdot \frac{r_{i,j} + 1 - r_{i,j}}{r_{max}} ,
\]
Likelihood-Based Test Statistics

- Scan statistic (Kulldorff, 1997)

  \[ LR_{\text{max}} = \max_{i,j} LR_{i,j} \]

  - Associated cluster is MLE.

  - Specific, but biased.

- WALR statistic (Gangnon & Clayton, 2000)

  \[ \text{WALR} = \sum_{i,j} w_{i,j} LR_{i,j} \]

  - Posterior probability of cluster \( i, j \) = \( w_{i,j} LR_{i,j} / \text{WALR} \)

  - Unbiased, but less specific.
Proposed Likelihood-Based Test

Statistics

- Penalized scan statistic
  \[ LR_{\text{max}} = \max_{i,j} w_{i,j} LR_{i,j} \]
  - Associated cluster maximizes posterior probability.
  - Weight penalizes clusters in dense areas.

- WALR scan (WALRS) statistic
  \[ \text{WALRS} = \max_{k} \frac{\sum_{i=1}^{N} \sum_{j=1}^{m_i} w_{i,j} LR_{i,j} I_{\{k \in c_{i,j}\}}}{\sum_{i=1}^{N} \sum_{j=1}^{m_i} w_{i,j} I_{\{k \in c_{i,j}\}}} \]
  - Posterior probability of cluster \( i, j \) = \( w_{i,j} LR_{i,j} I_{\{k \in c_{i,j}\}} / \text{WALRS} \)
  - Localizes test via common cell.
Unbiased Cluster Selection

- Tests associated w/ “estimated clusters.”

- Under null hypothesis, each cell should have an equal chance of belonging to the “estimated cluster.”

- Tests satisfying the above condition are unbiased.

- Evaluate bias thru simulations under null.
Estimated Bias

- Scan
- Penalized Scan
- WALR
- WALRS

March 2001
Power Comparisons

- Simulations for 15 different clusters.
- Variety of different scenarios.
  - Geographically large and small clusters.
  - Large and small populations.
- Designed to be detectable, but not obvious.
- Rejection rates based on 1,000 simulations.
- Rate of false rejections low.
Simulated Clusters
Power Comparisons –
High Rate Clusters

March 2001
Power Comparisons –
Low Rate Clusters

March 2001
New York Leukemia Data –

All Tests

- Scan statistic 13.05 (5% CV = 9.87)
- Penalized scan statistic 2.00 (5% CV = -0.02)
- WALR statistic 4.96 (5% CV = 1.83)
- WALRS statistic 8.11 (5% CV = 4.87)
- In all cases, reject null hypothesis.
New York Leukemia Data –
Cluster Memberships

Scan

Penalized Scan

WALR

WALRS

March 2001
New York Leukemia Data –
Disease Rates

Scan

Penalized Scan

WALR

WALRS
Concluding Remarks

- Proposed penalized scan and WALRS test for cluster detection.

- WALR and WALRS tests unbiased and powerful.

- Penalized scan statistic overcorrects for bias in scan statistic.

- Estimates from WALR quantify uncertainty about cluster location.

- Estimates from WALRS quantify uncertainty about cluster composition.