Cancer Clusters: The Myth and the Method

Tim E. Aldrich, Ph.D., M.P.H.
NAACCR Symposium: Cancer Informatics
Toronto, Canada
June 9, 2002
CCR JUSTIFICATION

- Prevention
- Treatment
- Research
- Environment
- Ethics
The Myth

What exactly IS a cluster?

“A cluster is an unusual aggregation, real or perceived, of health events that are grouped together in time and space, and are reported to a health agency.”

MMWR, 39/No.RR-11; 7/27/90.
“While I don’t play poker often, I do know that four-of-a-kind can be beaten by a straight flush. But, when I draw four-of-a-kind, I am going to bet on it.”

Alexander Langmuir, M.D.
Director, Epidemiology Section
Centers for Disease Control, 1965
SPATIAL CLUSTER

DISTANCE (miles)

DISTANCE (miles)
SPACE–TIME CLUSTER

TIME (years)
The Method

These guidelines represent a range of options and a recommended systematic strategic approach to the practice of investigating clusters of health events.

MMWR, 39/No.RR-11; 7/27/90.
Potential Cancer Cluster

Female Breast Cancer By Age

- 60 to 80
- 50 to 60
- 40 to 50
- 30 to 40
- 0 to 30

500 meters
“The constructive approach to this situation, in my opinion is not to develop highly refined statistical techniques to determine whether or not a certain cluster might have resulted by chance alone. But, rather to investigate each cluster as it is reported and see if additional association of possible interest can be found. If none turn up, there is obviously a cold trail, and any good hunting dog will abandon it and go look for a better one. If the scent strengthens, then hot pursuit is in order.”

Alexander Langmuid, M.D.
Director, Epidemiology Section
Centers for Disease Control, 1965
Search for Global Clustering

Adjacency

Occupancy

Progressive Curtailment of Randomness
Age-Adjusted Cancer Incidence Rates by A.D.D. in Kentucky
MALE LUNG CASES, 1998-1999

Copyright (C) 2000 Kentucky Cancer Registry
Age-Adjusted Cancer Incidence Rates by County in Kentucky
MALE LUNG CASES, 1998-1999

Rate / 100,000
62.7 - 92.9
93.2 - 110.4
111.0 - 134.7
135.4 - 235.1

Total Male Population 1998
Total Male Population 1999
Age-adjusted to the 1970 U.S. Standard Population

Copyright © 2000 Kentucky Cancer Registry
Male Lung Cancer 1998-99
Statistical Scoring

White < 10 Cases  Yellow - Same Rate as State  Red - Statistically Signif.
Green - Significantly lower  Yuck - Higher than the state, but not signif.
Occupancy within a progressive small geographic area

This progressive logic for occupancy over a probability surface may be connected to disease cluster analyses for cell occupancy using a negative binomial solution

\[
Pr = \frac{n}{m} \left[\frac{m-1}{m}\right]^{n-k}
\]

The sequential search for a recurring rare [or aberrant] event in a small area, over time.
So. Carolina County-Level and Race/Sex Level Analyses for Brain Cancer

- **Very High**
- **Higher than SC**
- **SC Rate or less**
**Illustration of Progressive Stochastic Curtailment: Brain Cancer**

<table>
<thead>
<tr>
<th>South Carolin</th>
<th>State Rate: 5.4 (Poisson 95% UCL = 9.936)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 Health Districts</td>
<td>Palmetto: 10.43 WM: 26.32 WF: 21.64 Lexington (No F/U)</td>
</tr>
<tr>
<td></td>
<td>App II: 11.71 WF 45-54 yo 13.87 AF 25-29 yo 11.82</td>
</tr>
<tr>
<td></td>
<td>App III: 9.75 White 50+ AA Both, AfrAm F 5-9 yo 35-49 yo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Counties</th>
<th>Newbey 15-9: 7.2 55+ 9.4</th>
<th>Richld 60-64 yo 5.2</th>
<th>Greenville 25-29 45-54</th>
<th>Pickn &lt; 10 45-54</th>
<th>Spartanbg 25-35 50-54</th>
<th>Unin 3 age strat</th>
<th>Greenwood All Older F END</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Age-Groups</th>
<th>WM 15-9 12.58</th>
<th>Groups All Go Back to State Pattern END</th>
<th>25-9 AF 45-54 WF END</th>
<th>White WF END</th>
<th>Only AF move at all END</th>
<th>All WM End</th>
<th>Colleton 15-19 AM Still END</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Race-/Sex</th>
<th>WM 55+ 12.5</th>
<th>Barnwell: 5-9 WF: 5.5 60-64 WF: 9.75 12.5</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Follow-up</th>
<th>CumRat WM 27.85</th>
<th>Marlboro: 20-24: 8.1 AM: 12.0 Cumulative Rate 40-44: 5.0 AM: 12.7 Afr-Am. Males 45-49: 5.9 AM: 13.0</th>
</tr>
</thead>
</table>

**Cumulative Rate** **WF**: 26.75 **MM**: 55.5
“The closer the [geographic] approach, the more rates should increase, if the data are sensitive to the underlying causal process.”

Dr. Tom Mason, USC SPH
Tampa, FL. February 14, 1995
Sequential Spatial Stochastic Curtailment

Established Risk Factor:  
Smoking for Lung Ca  
$\text{SMR} = 2.1 \ p = 0.12$

Regional Risk for AA  
$\text{SMR} = 4.6 \ p < 0.10$

County AA Males 45-64 yo  
$\text{SMR} = 6.5 \ p < 0.01$

Cluster in Specific Community
South Carolina

13 Health Districts
South Carolina DHEC District Analyses for Colorectal Cancer: Five Districts High

DHEC Districts High Colorectal Incidence
S. Carolina Analysis for DHEC Districts Colorectal Ca. Mortality: Four High

DHEC Districts High for Colorectal Mortality
So. Carolina County-Level And Race/Sex Analysis for Colorectal Cancer

- **High Mortality**
- **High Incidence**
- **High for Both**
<table>
<thead>
<tr>
<th>Reference SC Colon Cancer Analysis</th>
<th>Incidence: 43.5 WM: 52.8  WF: 35.2  AfAmM: 54.3  AfAmF: 40.4</th>
<th>Mortality: 17.7 WM: 20.7  WF: 13.0  AfAmM: 28.6  AfAmF: 18.7</th>
<th>Late Stage State: 56.70  White: 56.28  AfAm: 58.38</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR Area</td>
<td>Region 2: Hi For Incidence 46.58</td>
<td>None High For Mortality</td>
<td>Region 1 High for Late Stage: 62.74%</td>
</tr>
<tr>
<td>District Incidence</td>
<td>Catawba 45.21</td>
<td>Edisto 45.75</td>
<td>Watere 49.04</td>
</tr>
<tr>
<td></td>
<td>Palmett 45.99</td>
<td>PeeDe 44.24</td>
<td>Triden 46.42</td>
</tr>
<tr>
<td>District Mortality</td>
<td>Catawba 17.00</td>
<td>Edisto 16.09</td>
<td>Watere 21.31</td>
</tr>
<tr>
<td></td>
<td>Palmett 16.58</td>
<td>PeeDe 18.51</td>
<td>Triden 17.13</td>
</tr>
<tr>
<td>Late Stage</td>
<td>Chester Hi Unk</td>
<td>All Groups High Late Stage</td>
<td>High All Groups</td>
</tr>
<tr>
<td></td>
<td>Dillon High Unk High Both Races</td>
<td>Union High AfAm Cherokee High All</td>
<td></td>
</tr>
<tr>
<td>Screenin</td>
<td>Poor</td>
<td>High Sigmoidoscopy</td>
<td></td>
</tr>
<tr>
<td>Follow-up</td>
<td>Focus AfAmr, esp Female</td>
<td>Site Two</td>
<td>Focus All Groups; AA esp Female</td>
</tr>
<tr>
<td></td>
<td>AAM</td>
<td></td>
<td>Product of Good Work</td>
</tr>
</tbody>
</table>
Community-based Indicators

- Established Risk Relationship
- Indicator
- Indicator
- Indicator
- At Risk Community
Tier I vs Tier II

- Local community-based:
  - Surveys
  - Geographically referenced
  - Small area
  - Community-defined

- State or national level:
  - Assessment
  - Discrete
  - Secondary data sources
  - Small sample statistics
Search for Local Clustering

Case-to-case Nearest Neighbor Sentinel Events
Pleural Cancer Cases in Charleston S.C.
CUSUM

Pleural Cancer and Asbestos Deaths in the Tri-County Area
(1969 - 1997) - also includes new pleural cancer cases in 1996 and 1997

P-VALUE
< .05

TEXAS

Pleural Cancer and Asbestos Deaths in the Tri-County Area
(1969 - 1997) - also includes new pleural cancer cases in 1996 and 1997

P-VALUE
0.0000008

POISSON

Pleural Cancer and Asbestos Deaths in the Tri-County Area
(1969 - 1997) - also includes new pleural cancer cases in 1996 and 1997

P-VALUE
0.0057

NEGATIVE BINOMIAL

Pleural Cancer and Asbestos Deaths in the Tri-County Area
(1969 - 1997) - also includes new pleural cancer cases in 1996 and 1997

P-VALUE
0.0010

SCAN

Pleural Cancer and Asbestos Deaths in the Tri-County Area
(1969 - 1997) - also includes new pleural cancer cases in 1996 and 1997

P-VALUE
0.0057

*Tri-County area includes Berkeley, Charleston, and Dorchester counties.
*Filled bars denote the 13 pleural cancer cases from 1969 and 11 pleural cancer cases from 1996 identified through the SCOCR.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Best Est.</th>
<th>Acceptable Ests.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobacco</td>
<td>30</td>
<td>25 - 40</td>
</tr>
<tr>
<td>Diet</td>
<td>35</td>
<td>10 - 70</td>
</tr>
<tr>
<td>Infection</td>
<td>10?</td>
<td>1 - ?</td>
</tr>
<tr>
<td>Reproductive &amp; Sexual History</td>
<td>7</td>
<td>1 - 13</td>
</tr>
<tr>
<td>Occupation</td>
<td>4</td>
<td>2 - 8</td>
</tr>
<tr>
<td>Geophysical</td>
<td>3</td>
<td>2 - 4</td>
</tr>
<tr>
<td>Alcohol</td>
<td>3</td>
<td>2 - 4</td>
</tr>
<tr>
<td>Pollution</td>
<td>2</td>
<td>1 - 5</td>
</tr>
<tr>
<td>Medicines/Medical Procedures</td>
<td>1</td>
<td>0.5 - 2</td>
</tr>
<tr>
<td>Industrial Products</td>
<td>1</td>
<td>1 - 2</td>
</tr>
</tbody>
</table>

Derived from Doll & Peto, J.N.C.I., June '81
SENTINEL HEALTH EVENT (SHE):

A preventable disease, disability, or untimely death whose occurrence serves as a warning signal.
A Sentinel Event Surveillance Strategy Requires

- A Reportable List of Rare Health Events
- Population-Based Case Ascertainment
- Innovative Statistical Methods and Criteria
Figure 4. Three Tiered Information Capability.

1. RISK ASSESSMENT DESIGN
   BASELINE MEASURES
   SIMULATION STUDIES

2. MONITOR AND SENTINEL SYSTEMS
   ALERT LEVEL-
   ROUTINE SYSTEMS

3. DETAILED SPECIALIZED STUDIES
   ACTION LEVEL-
   EPIDEMIOLOGICAL STUDIES

"The reason for collecting, analyzing, and disseminating information on a disease is to control that disease. Collection and analysis should not be allowed to consume resources if action does not follow."

Cancer Clusters: The Myth and the Method

Reports of Cancer clusters have buffeted cancer registries for decades. Many public health officials disparage cancer cluster investigations because of their lack of productivity for identifying *bona fide* environmental risks: the myth. Nonetheless, several states have established response protocols, of which statistical methods are a prominent part. This talk will highlight a battery of solutions for assessing spatial and temporal aggregates of disease; examples will be provided. Several of these analysis techniques are useful for searches for generalized clustering [e.g., small area surveillance, or program evaluation] as well as studies of local clustering. The advent of Geographic Information Systems [GIS] technology has also augmented disease cluster studies for central cancer registries. Therefore, a judicious response mechanism to cancer cluster reports can serve an immense disease control function by promoting community-level goodwill and action: the method. Likewise, sound statistical methods for spatial clustering studies may garner immense public health benefit.
‘It’s amazing what you can see when you look...”  Yogi Berra

Contact:  Tim E. Aldrich, Ph.D., M.P.H.
Dept. Epidemiology and Biostatistics
Univ. So. Carolina, School Public Health
Sumpter and Green Streets
Columbia, SC 29208

Voice: 803-777-5331    Fax: 803-434-7529

E-mail: tim.aldrich@palmettohealth.org