MAKING SENSE OF THE EPIDEMIOLOGICAL LITERATURE: PROBLEMS WITH EPIDEMIOLOGICAL STUDIES OF ECOLOGICAL DESIGN

D. Wayne Berman, Ph.D.
Aeolus, Inc.
Albany, CA

And

Louis Anthony Cox, Ph.D.
Cox Associates
Denver, CO
I would like to thank NSSGA for funding the original research.
A Class of Epidemological Studies of ecological design that use regression to evaluate associations

The Problem:
- In such studies, proximity to a source or sources of a particular toxin is used as a surrogate for direct estimates of exposure
- Thus, such studies involve evaluation of spatial distributions
- Regression analyses are non-spatial statistical procedures.

Approximately 100 studies of similar design addressing a variety of toxins and disease end points have been published and continue to appear
Illustrate the problem with using regression in epidemiological studies of ecological design

Summarize the characteristics of real population distributions

Summarize findings and conclusions from the Berman, Cox, and Popken papers

Define some useful criteria for detecting these problems

Identify some useful references describing how to conduct these types of analyses properly
Mesothelioma risk is positively associated with proximity to ultramafic rocks in California.

Risk decreases 6% with each 10 km increase in distance.

Mesothelioma risk is positively associated with proximity to ultramafic rocks in California. Risk decreases 6% with each 10 km increase in distance.

Unfortunately, the associations observed in this study have nothing to do with causality.....
Mesothelioma risk is positively associated with proximity to ultramafic rocks in California.

Risk decreases 6% with each 10 km increase in distance.

Unfortunately, the associations observed in these studies have nothing to do with causality.....

.......by the end of this talk, I hope this will be intuitively obvious
Cases and Controls dispersed throughout:

- 100s of communities
- 1000s of work locations
By definition, valid statistical tests show positive results due to chance no more than 5% of the time (meaning of 5% significance).
By definition, valid statistical tests show positive results due to chance no more than 5% of the time (meaning of 5% significance).

Correspondingly, to infer causality, tests for associations cannot detect non-causal (random) associations more than 5% of the time.
Two Dense Communities:
(Perfectly Segregated (0:100) and Unrealistic)
Two Dense Communities: (Perfectly Segregated (0:100) and Unrealistic)
Two Dense Communities:
(Perfectly Segregated (0:100) and Unrealistic)
Two Dense Communities:
(Perfectly Segregated (0:100) and Unrealistic)

X Coordinate
\[ P = 0.005 \]
Two Dense Communities:
(Perfectly Integrated (50:50) and Unrealistic)

X Coordinate

P = 0.47
Two Dense Communities: (Imperfectly Integrated (60:40) but Realistic)
Two Dense Communities:
(Imperfectly Integrated (70:30) but Realistic)
Two Disperse Communities:
(Perfectly Segregated (0:100) and Unrealistic)
Two Disperse Communities:
(Perfectly Integrated (50:50) and Unrealistic)

X Coordinate

Y Coordinate

$P = 0.72$
Two Disperse Communities:
(Imperfectly Integrated (60:40) but Realistic)

X Coordinate
P = 0.005
Two Disperse Communities:
(Imperfectly Integrated (70:30) but Realistic)

X Coordinate

P = 0.002
# Observed Ratios of "Cases" vs. "Controls" for Indicated Characteristics Among the 7,049 Census Tracts and 58 Counties in California

<table>
<thead>
<tr>
<th>Ratios:</th>
<th>Fraction of Observed Ratios at Least as Extreme as Indicated Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Higher vs. Lower Income</td>
</tr>
<tr>
<td><strong>Census Tract Data</strong></td>
<td></td>
</tr>
<tr>
<td>80/20</td>
<td>23%</td>
</tr>
<tr>
<td>75/25</td>
<td>35%</td>
</tr>
<tr>
<td>70/30</td>
<td>48%</td>
</tr>
<tr>
<td>60/40</td>
<td>73%</td>
</tr>
<tr>
<td>50/50</td>
<td>99%</td>
</tr>
<tr>
<td><strong>County Data</strong></td>
<td></td>
</tr>
<tr>
<td>80/20</td>
<td>0%</td>
</tr>
<tr>
<td>75/25</td>
<td>0%</td>
</tr>
<tr>
<td>70/30</td>
<td>5%</td>
</tr>
<tr>
<td>60/40</td>
<td>47%</td>
</tr>
<tr>
<td>50/50</td>
<td>100%</td>
</tr>
</tbody>
</table>
What’s going on?
Cases vs. Controls with Alternate 70 to 30% Enrichment
Epidemiological studies with spatial ecological designs typically produce invalid conclusions if regression models are used to interpret exposure-response associations; causality cannot be reasonably inferred from these studies.

Note: at least 100 studies of this design have been published and virtually all inappropriately suggest causality.
Useful Criteria: What to Look For in a Study

- Does the study link outcome with exposure explicitly?
- If not, does the study employ appropriate methods of spatial statistics?
- If not, does the study incorporate appropriate negative and positive controls?
- If not, does the study at least attempt to control for all reasonable factors that affect where individuals choose to live?


