Algebraic Conditions for Binary Spuriousity

MILO SCHIELD
Augsburg College  Department of Business Administration
Director, W. M. Keck Statistical Literacy Project

THOMAS V.V. BURNHAM
Cognitive Consulting

August 4, 2003

www.augsburg.edu/ppages/~schield  schield@augsburg.edu
Associations Confounded
No test for Confounding

In observational studies, associations are often *confounded* (tangled up).

E is Effect of interest
A is associated predictor
B is confounder

Sex

Weight

Height

Confounding Factor

A —— B —— E

A is associated predictor
B is confounder
E is Effect of interest
Categorical Cube: Three Binary Variables

B: confounder.
Quantitative Rate Cube
Non-Planar Data

AQ = \( P(E|A=0, B=XQ) = Rb \times XQ + Ra \times (1-XQ) \)

\( XQ = P(B|A=0) \)

A: Associated

B: confounder.

E: effect
Criteria for Spuriousity: A has “no effect” on E

Cornfield & Gastwirth used a cross-A rate equality model:
• \( P(E|A \text{ and } B) = P(E|B) = P(E|\text{non-}A \text{ and } B) \)
• \( P(E|A \text{ and non-}B) = P(E|\text{non-}B) = P(E|\text{non-}A \text{ and non-}B) \)

We used two regression models:
• A non-interactive model: \( E = b_0 + \underline{b1}A + b2B \)
• An interactive model: \( E = b0 + (b1 + b3B) \times A + b2B \)

A-E association is spurious if underlined factor is zero.
As viewed from confounder perspective: B-E
• Non-interactive model: B line \parallel A line
• Interactive model: Rate lines intersect at prevalence of B.
Non-Interactive Model: AP:AQ line and BP:BQ line
Non-interactive Spuriousity
Projected on B:E Face

Non-Interactive Spuriousity

Percentage who are E

Percentage who are B

All Non-B

Percentage of Subjects who are B

All B
Standardizing Shows Influence of Confounder

Standardizing Can Decrease A Difference

Mean Income

Percentage of Families with Two Parents

White Families

Black Families
Standardizing Shows Simpson’s Paradox

Standardizing Can **Reverse A Difference**

![Graph showing the relationship between the percentage of people in "Poor" condition and death rate for rural and city hospitals.](image_url)
Interactive Spuriousity via Standardizing

Interactive Spuriousity: Non Planar Data

Percentage who are E

Percentage who are B

All Non-B

All B

Percentage of Subjects who are B

0% 20% 40% 60% 80% 100%
Spuriousity Results: New Necessary Condition

Gastwirth-Cornfield: $\text{RR}(E:B) > \text{RR}(E:A)$

New: $\text{RR}(E:B) - 1 > \left[\text{RR}(E:A) - 1\right]\left[\frac{P(A)}{P(B)}\right]$ 

What cancer-gene effect size is necessary to make association between smoking and cancer spurious?

$\text{RR}(E:A) = 9$ for cancer among smokers vs. non.
$P(B) = 10\%$. 10% of adults have a cancer gene
$P(A) = 40\%$. 40% of adults smoke, then

- Gastwirth-Cornfield: $\text{RR}(E:B) > 9$.
- New: $\text{RR}(E:B) > 33$
Conclusions

Spuriousity depends on model.  
Cornfield conditions more-generally valid.  
Standardizing illustrates interactive model.  
Spuriousity conditions for non-interactive and interactive models overlap.  
New equations for non-interactive spuriousity.  
New inequality for non-interactive model:
\[ \text{RR}(E:B)-1 > \frac{\text{RR}(E:A)-1 \cdot P(A)}{P(B)} \]