Dec 2010 1

Statistical Literacy: Confounding

MILO SCHIELD,
Augsburg College
Director, W. M. Keck Statistical Literacy Project
Vice President, National Numeracy Network

University of St. Thomas
December 10, 2010

Slides at www.StatLit.org/pdf/2010SchieldUST6up.pdf

Dec 2010 2

Statistical Literacy

Statistical literacy is **the ability to read and interpret summary statistics in everyday life**.

Statistical Literacy studies

(1) the relation between statistical associations and causation, and

(2) the full-range of influences on a statistic or on a statistical association. [Take CARE]

Dec 2010 3

Take CARE: Context

The influence of factors **taken into account by**

- data broken out by subgroups in tables and graphs
- averages, ratios and comparisons of averages and ratios
- epidemiological models (cf., deaths attributed to obesity)
- regression models and
- the study design (cf., longitudinal vs. cross-sectional; experiment vs. observational study).

The influence of related factors (confounders)
not taken into account in the study and
not blocked by the study design.

Dec 2010 4

Controlling for a confounder can DECREASE an association

MN has 3.8 times as much prison expense as ME

State	Total	# Inmates	Per Inmate
MN	$184M	4,865	$37,825
ME	$48M	1,424	$33,711

MN has 3.4 times as many inmates as ME

MN has 25% more prison expense *per inmate* than ME

Dec 2010 5

Controlling for a confounder can NULLIFY an association

MD has 3 times as much prison expense as KS

State	Total	# Inmates	Per Inmate
MD	$481M	21,623	$22,250
KS	$159M	7,148	$22,250

MD has three times as many inmates as KS

MD has the same prison expense *per inmate* as KS

Dec 2010 6

Controlling for a confounder can REVERSE an association

CA has 50% more prison expense than NY

State	Total	# Inmates	Per Inmate
CA	$2.9B	136K	$21,385
NY	$1.9B	69K	$28,426

CA has almost twice as many inmates as NY

CA has 25% less prison expense *per inmate* than NY

Dec 2010 7

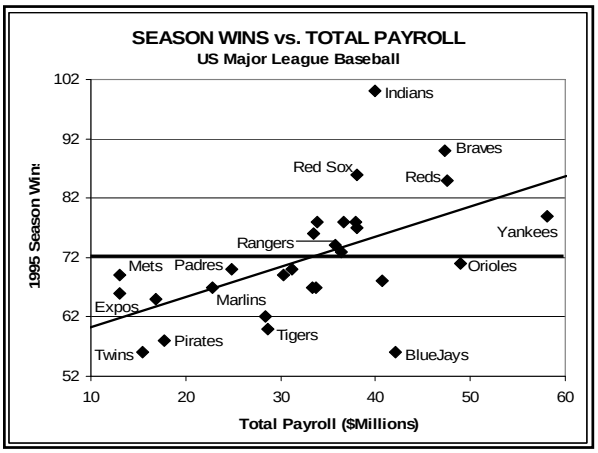
Controlling for a confounder can INCREASE an association

MN has 27% more prison expense than IA

State	Total	# Inmates	Per Inmate
MN	$184M	4,865	$37,825
IA	$144M	5,929	$24,286

MN has 18% fewer inmates than IA

MN has 56% more prison expense *per inmate* than IA



Dec 2010 9

Adjusting for Land Size: Standardize on Average Lot

Dec 2010 10

SAT VERBAL SCORES: FLAT

GROUP	1981	2002	CHANGE
White			
Black			
Asian			
Mexican			
Puerto Rican			
American Indian			
ALL Test takers	**504** (100%)	**504** (100%)	**ZERO**

Dec 2010 11

Multivariate Analysis can be Complex

To simplify, consider cases with
- a binary outcome,
- a binary predictor and
- a binary confounder.

What are the necessary conditions for nullification or a reversal?

See Schield (1999) and Schield and Burnham (2003)

Dec 2010 12

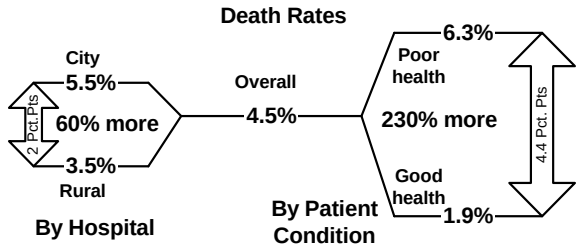
City Hospital: Hospital of Death??

Hospital	Total	Died	Death Rate
City	1,000	55	5.50%
Rural	1,000	35	3.50%
Both	2,000	90	4.50%

Condition	Total	Died	Death Rate
Good	800	15	1.90%
Poor	1,200	75	6.30%

Dec 2010 13

Can this confounder nullify or reverse this association?



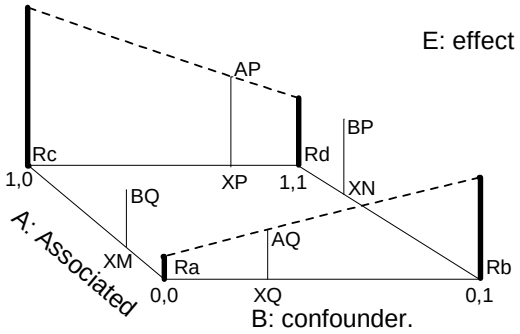
Dec 2010 14

Confounder Reverses; City Hospital is Better

Condition	Hospital	Total	Died	Death Rate
Good	City	100	1	1.00%
	Rural	700	14	2.00%
	Total	800	15	1.90%
Poor	City	900	54	6.00%
	Rural	300	21	7.00%
	Total	1,200	75	6.30%

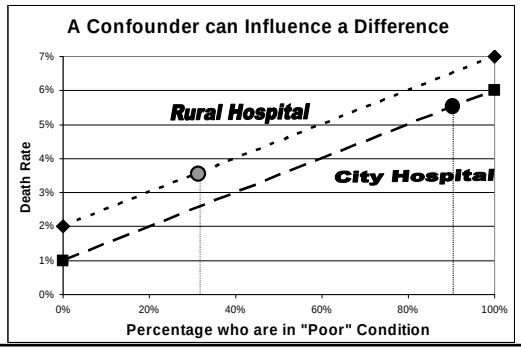
Dec 2010 15

Two-Group Rates with a Binary Confounder



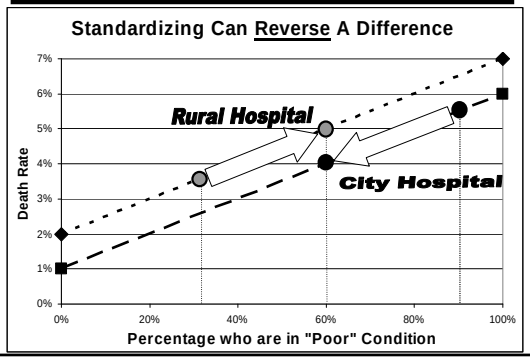
Dec 2010 16

Compare Hospital Death Rates Confounder: Patient Condition

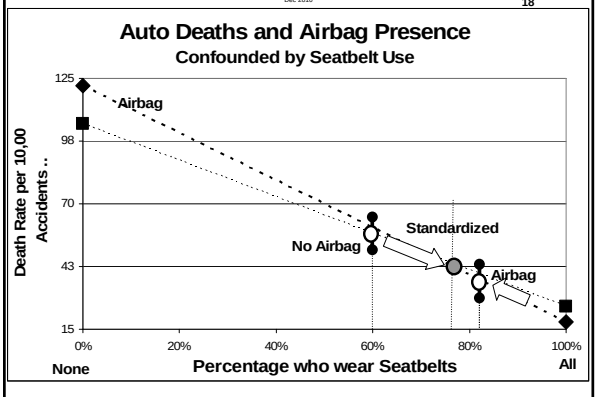


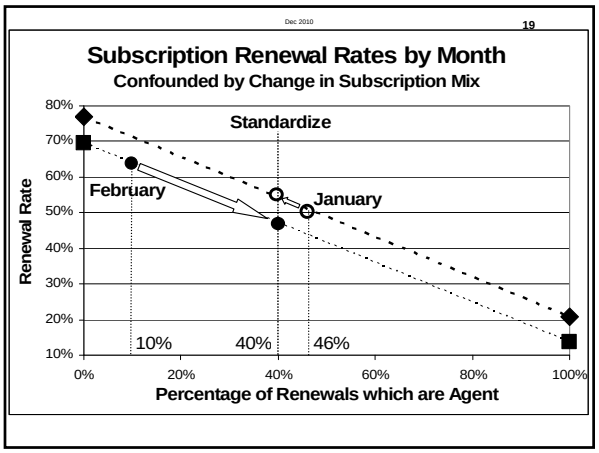
Dec 2010 17

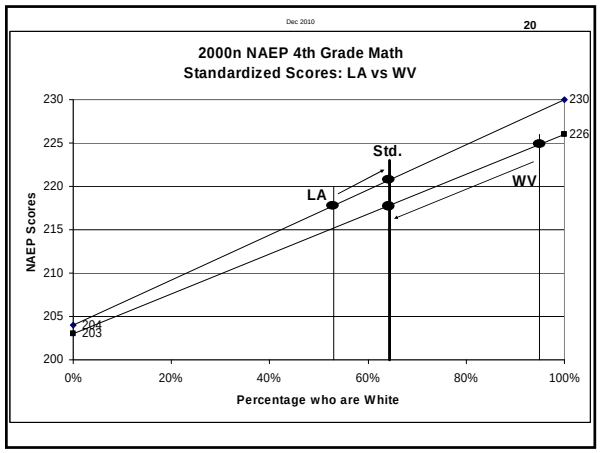
Standardize on combined confounder percentage

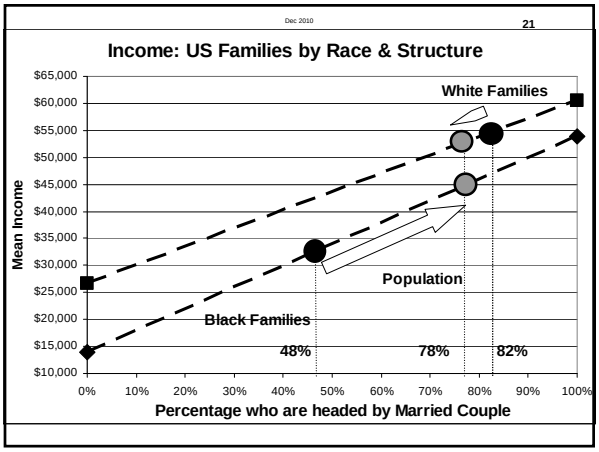


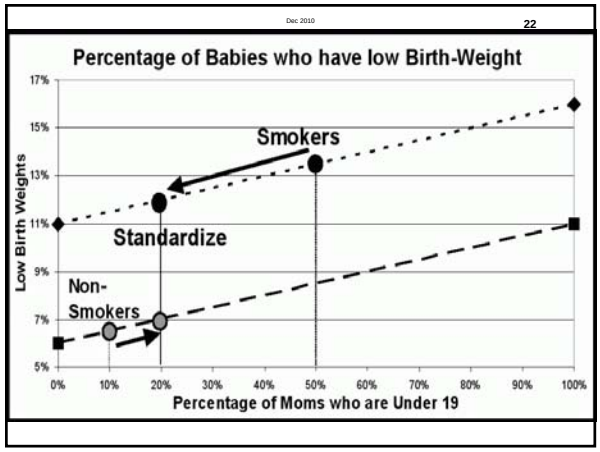
Dec 2010 18

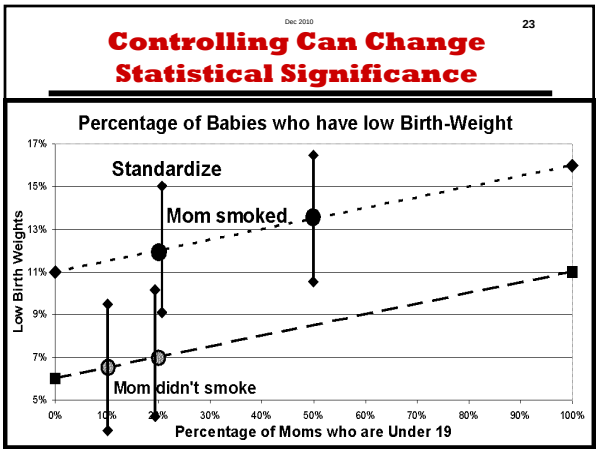


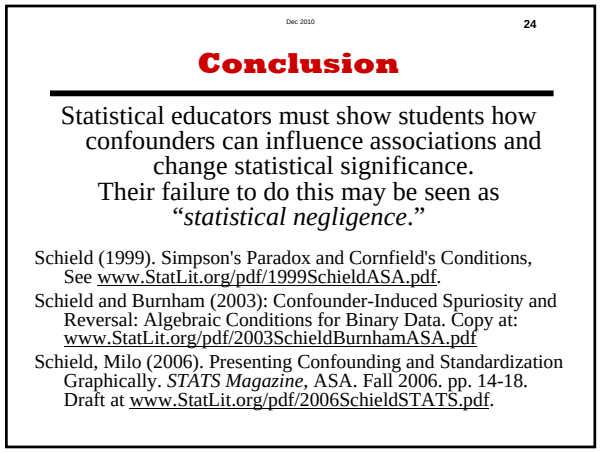

Dec 2010
19
Subscription Renewal Rates by Month
Confounded by Change in Subscription Mix
Standardize
February
January
Renewal Rate
10%
40%
46%
Percentage of Renewals which are Agent


Dec 2010
20
2000n NAEP 4th Grade Math
Standardized Scores: LA vs WV
Std.
LA
WV
NAEP Scores
Percentage who are White


Dec 2010
21
Income: US Families by Race & Structure
White Families
Mean Income
Population
Black Families
48%
78%
82%
Percentage who are headed by Married Couple


Dec 2010
22
Percentage of Babies who have low Birth-Weight
Smokers
Standardize
Non-
Smokers
Low Birth Weights
Percentage of Moms who are Under 19


Dec 2010
23
Controlling Can Change
Statistical Significance
Percentage of Babies who have low Birth-Weight
Standardize
Mom smoked
Mom didn't smoke
Low Birth Weights
Percentage of Moms who are Under 19


Dec 2010
24
Conclusion
Statistical educators must show students how confounders can influence associations and change statistical significance.
Their failure to do this may be seen as "statistical negligence."
Schield (1999). Simpson's Paradox and Cornfield's Conditions, See www.StatLit.org/pdf/1999SchieldASA.pdf.
Schield and Burnham (2003): Confounder-Induced Spuriosity and Reversal: Algebraic Conditions for Binary Data. Copy at: www.StatLit.org/pdf/2003SchieldBurnhamASA.pdf
Schield, Milo (2006). Presenting Confounding and Standardization Graphically. STATS Magazine, ASA. Fall 2006. pp. 14-18. Draft at www.StatLit.org/pdf/2006SchieldSTATS.pdf.