V1 2015 StatChat2 1

Confounding: A Big Idea

Milo Schield, Augsburg College

Member: International Statistical Institute

US Rep: International Statistical Literacy Project

Director, W. M. Keck Statistical Literacy Project

VP. National Numeracy Network

Editor: www.StatLit.org

August 1, 2016

www.StatLit.org/pdf/2016-Schield-ASA-Slides.pdf

V1 2015 StatChat2 2

Core Concepts in Intro Stats

McKenzie (2004): Survey of Educators

Goodall@RSS (2007) Big Ideas in Statistics

Garfield & Ben Zvi (2008): Big Ideas of Statistics

Gould-Miller-Peck (2012). *Five Big Ideas*

Blitzstein@Harvard (2013): *10 Big Ideas Stat110*

Stigler (2016): *Seven pillars of statistical wisdom*

V1 2015 StatChat2 3

Ambiguity of "Importance"

	Classifying Important Ideas		
	Logically	Socially	Cognitively
Topic	{Common}		
Claim		**Good/Bad***	**Fallacy**
	* Contribution or catastrophe		

Topic (randomness) or a claim: ME ~ 1/sqrt(n)

This paper focuses on claims or relationships having substantial social or cognitive consequences.

V1 2015 StatChat2 4

1A: Fallacies

1. Confusion of the inverse: P(A|B) = P(B|A)
2. Conjunction fallacy: P(A&B) > P(A)
3. P(A&B |C) > P(A |B&C): Three-factor fallacy
4. **Individual fallacy**
5. **Ecological fallacy**
6. **Simpson's Paradox**

V1 2015 StatChat2 5

Contributions of Statistics to Human Knowledge

Statistics are numbers in a context
Association is not causation

RANDOMNESS and CAUSATION
- Chance, independence and sampling distributions
- Margin of error, hyp tests & statistical significance
- *Random assignment and causation (Fisher: RCT)*

CONFOUNDING and CAUSATION
- Comparisons, ratios, models and study designs
- Epidemiological causation (Bradford-Hill)
- *Confounder conditions for nullification (Cornfield)*

V0.7 Conditional probability, medical tests and Bayesian reasoning
Coincidence, Simpson's Paradox and regression to the mean

V1 2015 StatChat2 6

#2A: Butterfly Fallacy

One should never trust a statistical association generated by an observational study.

An unknown or unmeasured confounder – regardless of size (a small as a butterfly) – can nullify or reverse an observed association.

V1 2015 StatChat2 7

Smoking Causes Cancer: Fisher's Argument

Observational data: Smokers are **10 times** as likely to develop lung cancer as are non-smokers.

Some statisticians wanted to support the claim that smoking "caused" lung cancer.

Sir Ronald Fisher (1958) noted that "association was not causation" and that there was a difference (**factor of two**) in smoking preference between fraternal and identical twins.

V1 2015 StatChat2 8

Smoking Causes Cancer: Cornfield's Reply

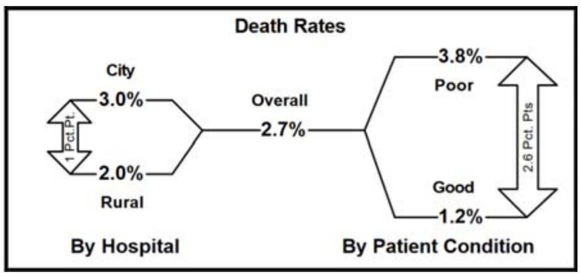
Cornfield et al (1959) argued that to nullify or reverse the observed association, the relative risk of a confounder must exceed the relative risk of that association.

Fisher never replied.

"Cornfield's minimum effect size is as important to observational studies as is the use of randomized assignment to experimental studies." Schield (1999)

V1 2015 StatChat2 9

Cornfield Condition for Nullification or Reversal



Schield (1999) based on realistic data

V1 2015 StatChat2 10

Confounder Distribution: Simple One-Parameter Model

How to deal with unknown or unmeasured confounders?

Assume: RR of confounders is distributed exponentially with a minimum RR of one and a mean RR of two.

RR	CDF		RR-1	CDF
1	0.00		0	0.00
2	0.63		1	0.63
3	0.86		2	0.86
4	0.95		3	0.95
5	0.98		4	0.98

V1 2015 StatChat2 11

Effect Sizes: Relative Risk 95% Confounder Resistant: Exp20

Obese vs. non-Obese

RR of Health Outcome	Women	Men
Type 2 Diabetes	12.7	5.2
Hyptertension	4.2	2.6
Heart attack	3.2	1.5
Colon Cancer	2.7	3.0
Angina	1.8	1.8
Gall-bladder	1.8	1.8
Ovarian Cancer	1.7	
Osteoarthritis	1.4	1.9
Stroke	1.3	1.3
Average	3.4	2.4

V1 2015 StatChat2 12

Conclusion

Students should be exposed to the major contributions of statistics to human knowledge.

Including multivariate thinking in the intro course means discussing confounding.

Introducing confounding means dealing with
- the Butterfly fallacy,
- the Cornfield conditions and
- ranking the resilience of an association to unknown or unmeasured confounders.

V1 2015 StatChat2 1

Confounding: A Big Idea

Milo Schield, Augsburg College

Member: International Statistical Institute

US Rep: International Statistical Literacy Project

Director, W. M. Keck Statistical Literacy Project

VP. National Numeracy Network

Editor: www.StatLit.org

August 1, 2016

www.StatLit.org/pdf/2016-Schield-ASA-Slides.pdf

Core Concepts in Intro Stats

McKenzie (2004): Survey of Educators

Goodall@RSS (2007) Big Ideas in Statistics

Garfield & Ben Zvi (2008): Big Ideas of Statistics

Gould-Miller-Peck (2012). *Five Big Ideas*

Blitzstein@Harvard (2013): *10 Big Ideas Stat110*

Stigler (2016): *Seven pillars of statistical wisdom*

V1 2015 StatChat2 3

Ambiguity of "Importance"

	Classifying Important Ideas		
	Logically	Socially	Cognitively
Topic	{Common}		
Claim		Good/Bad*	Fallacy
	* Contribution or catastrophe		

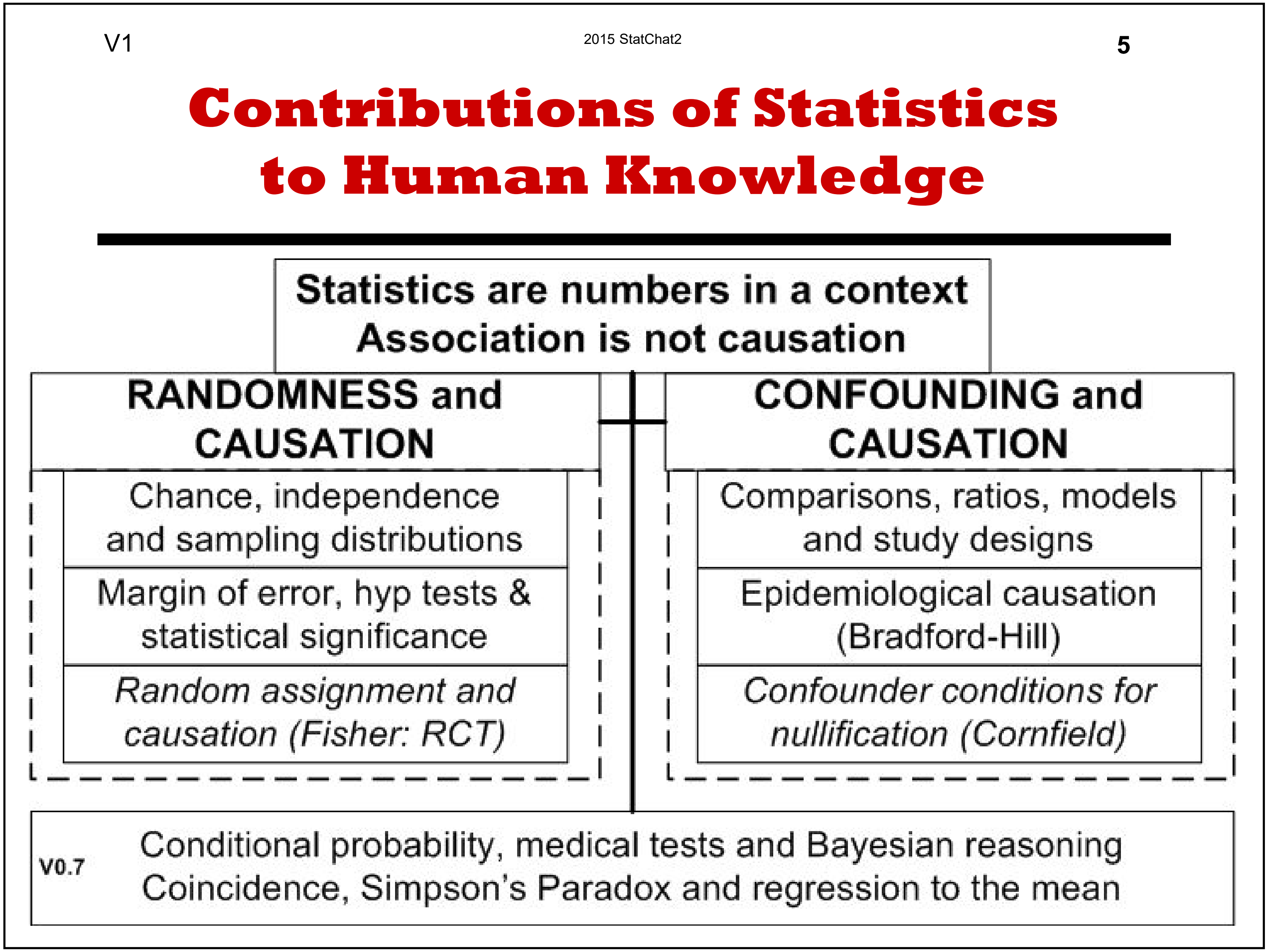
Topic (randomness) or a claim: ME ~ 1/sqrt(n)

This paper focuses on claims or relationships having substantial social or cognitive consequences.

V1 2015 StatChat2 4

1A: Fallacies

1. Confusion of the inverse: $P(A|B) = P(B|A)$
2. Conjunction fallacy: $P(A\&B) > P(A)$
3. $P(A\&B\ |C) > P(A\ |B\&C)$: Three-factor fallacy
4. **Individual fallacy**
5. **Ecological fallacy**
6. **Simpson's Paradox**


V1
2015 StatChat2
5
Contributions of Statistics to Human Knowledge
Statistics are numbers in a context
Association is not causation
RANDOMNESS and CAUSATION
Chance, independence and sampling distributions
Margin of error, hyp tests & statistical significance
Random assignment and causation (Fisher: RCT)
CONFOUNDING and CAUSATION
Comparisons, ratios, models and study designs
Epidemiological causation (Bradford-Hill)
Confounder conditions for nullification (Cornfield)
V0.7
Conditional probability, medical tests and Bayesian reasoning
Coincidence, Simpson's Paradox and regression to the mean

V1 2015 StatChat2 6

#2A: Butterfly Fallacy

One should never trust a statistical association generated by an observational study.

An unknown or unmeasured confounder – regardless of size (a small as a butterfly) – can nullify or reverse an observed association.

Smoking Causes Cancer: Fisher's Argument

Observational data: Smokers are **10 times** as likely to develop lung cancer as are non-smokers.

Some statisticians wanted to support the claim that smoking "caused" lung cancer.

Sir Ronald Fisher (1958) noted that "association was not causation" and that there was a difference (**factor of two**) in smoking preference between fraternal and identical twins.

V1 2015 StatChat2 8

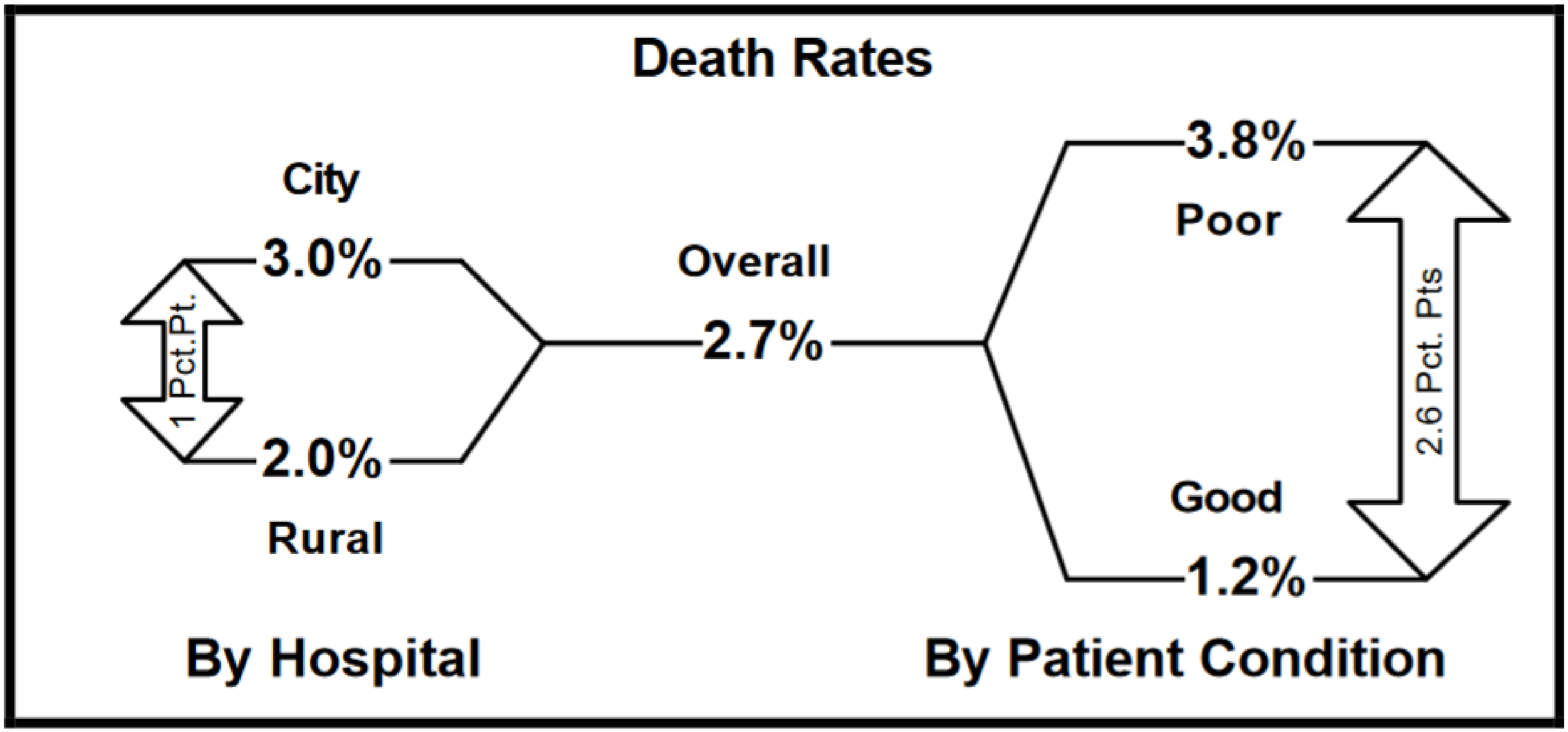
Smoking Causes Cancer: Cornfield's Reply

Cornfield et al (1959) argued that to nullify or reverse the observed association, the relative risk of a confounder must exceed the relative risk of that association.

Fisher never replied.

"Cornfield's minimum effect size is as important to observational studies as is the use of randomized assignment to experimental studies." Schield (1999)

Cornfield Condition for Nullification or Reversal



Schield (1999) based on realistic data

V1 2015 StatChat2 10

Confounder Distribution: Simple One-Parameter Model

How to deal with unknown or unmeasured confounders?

Assume: RR of confounders is distributed exponentially with a minimum RR of one and a mean RR of two.

RR	CDF		RR-1	CDF
1	0.00		0	0.00
2	0.63		1	0.63
3	0.86		2	0.86
4	0.95		3	0.95
5	0.98		4	0.98

V1 2015 StatChat2 11

Effect Sizes: Relative Risk 95% Confounder Resistant: Exp20

Obese vs. non-Obese

RR of Health Outcome	Women	Men
Type 2 Diabetes	12.7	5.2
Hyptertension	4.2	2.6
Heart attack	3.2	1.5
Colon Cancer	2.7	3.0
Angina	1.8	1.8
Gall-bladder	1.8	1.8
Ovarian Cancer	1.7	
Osteoarthritis	1.4	1.9
Stroke	1.3	1.3
Average	3.4	2.4

Conclusion

Students should be exposed to the major contributions of statistics to human knowledge.

Including multivariate thinking in the intro course means discussing confounding.

Introducing confounding means dealing with

- the Butterfly fallacy,
- the Cornfield conditions and
- ranking the resilience of an association to unknown or unmeasured confounders.