

CONFOUNDING AND CORNFIELD: BACK TO THE FUTURE

Milo Schield

Augsburg University, Minneapolis, MN, USA

Schield@Augsburg.edu

Most students see less value in statistics after taking the intro course than they did before. Most students are in disciplines that deal with observational studies and confounding. Introductory statistics doesn't deal with observational studies. Most intro statistics textbooks never mention confounding. Confounding wasn't one of McKenzie's top 30 concepts. Why are we so silent? Perhaps we aren't comfortable teaching confounding. To change the future, we need to go back to when Jerome Cornfield argued that smoking caused cancer. Fisher disagreed arguing it might be caused by a genetic confounder. Cornfield refuted Fisher by deriving the Cornfield conditions: one of the greatest contributions of statistics to human knowledge. We need to teach multivariate statistics, confounding and the Cornfield conditions so students will appreciate statistics.

Key words: Statistical literacy, effect size, confounder, lurking variable, Simpson's paradox

STATISTICAL EDUCATION ENJOYS SUCCESS

Statistical education has had success. Most college graduates study statistics. Schield (2016a) estimates that 57% of US four-year college students graduating in 2012 studied statistics. This estimate assumes that all students in the Table 1 majors must take statistics. This assumption will over-estimate the number in math, biology and history, but it omits many other majors where statistics may be required. ASA (2016).

The number of US high school students taking the AP statistics exam has more than doubled in 9 years: from 98 thousand in 2007 to 206 thousand in 2016. Wikipedia (2018a)

Table 1: Distribution Stat 101 by Major

%	Major
38%	Business or Economics
19%	Social Science or History
13%	Health
10%	Psychology
9%	Engineering
9%	Biological Science
2%	Math or Statistics
100%	All students in these majors

Table 2: Student Math Aptitudes by Major

SAT MATH SCORE					
Score	Pctile	MAJOR	Score	Pctile	MAJOR
613	80	Math/Stats	550	61	Social Sc.
575	72	Physical Sc.	522	51	Bus; English
579	70	Engineering	498	43	Commun.
554	62	Computer Sc.	489	40	Psychology
551	61	Biological Sc.	482	38	Education
Business Insider (2014)			College Board (2015)		

STATISTICAL EDUCATION HAS ONE BIG PROBLEM

Students see less value in statistics after taking the course than they did before. Schield and Schield (2008). Statistical educators seem unwilling to survey their students on this criterion. This topic is almost absent in statistical education papers or conferences. It's not a stated goal. It's as though statistical educators don't want to know. If we don't measure it, it doesn't exist.

Why don't students appreciate the value of statistics despite being deluged by a sea of data? Perhaps it is an acquired taste like Scotch. Maybe students won't see the value until they start working; maybe their dislike of math carries over to statistics. Here are some more explanations:

One explanation is a student-teacher math-aptitude gap. This gap can be seen in Table 2. (Schield 2016a). Most statistical educators majored in mathematics or statistics with an average math SAT score at the 80th percentile among college-bound SAT takers. The biggest group of students taking statistics majored in business where the average math SAT score is the 51st percentile. The average math SAT for psychology and education majors is at the 39th percentile. Maybe statistics teachers think and teach at a more abstract level than their students.

A second explanation is a student-teacher subject-interest gap. A student's subject interest is typically related to their chosen major. Table 1 shows the choice of majors among those taking statistics. The majors in Table 1 can be classified into two groups: those in the social science and those in the physical sciences with math/stat included in the latter. Since statistics teachers tend to

come from math/stat or the physical sciences, perhaps they can't relate to the subject interests of their students in the social sciences.

A third explanation involves a student-teacher prediction-causation gap. Richard de Veaux argued that "We are teaching the *wrong* things in the *wrong way* and in the *wrong order*." Rossman (2016). Perhaps it is the kinds of data we are – or are not – teaching. Normally we introduce causation using random assignment. If we teach observational data, it only involves predictive modeling – not causal exploration or explanation. Confounding isn't a problem in predicting.

Students majoring in the social sciences are more likely to use observational data to understand causation. Confounding is a major problem. Multivariate data is needed to deal with confounding. Teachers may avoid those topics except to say "Association is not causation."

By making multivariate data a goal for intro statistics, the 2016 update to the GAISE guidelines was a major step forward. Mentioning observational studies and confounding was a big step. Schield (2017) But confounding was not mentioned at the highest level. Our unwillingness to talk about observational causation, confounding and strength of evidence is arguably the primary reason our students' see little value in the introductory statistics normally taught in Stat 101.

STATISTICAL EDUCATORS: CAUSATION AND CONFOUNDING

Statistical educators are encouraged to deal with causation and confounding in intro stats.

- Moore (1997, 1998) identified "Beware the lurking variable" as a key component of statistical literacy.
- Tintle et al (2014) identified "confounding and variation" as "two substantial hindrances to drawing conclusions from data".
- Kaplan (2010). "Saying 'Correlation is not causation' is like saying 'Sex is not love!'"

Statistical educators seem unwilling to deal with causation in the introductory course.

- Pearl (2018) has written a most comprehensive critique of why statisticians are unwilling to deal with causation. Pearl's book, *The Book of Why: The New Science of Cause and Effect*, is arguably the most important book on causal statistics since Cornfield debated Fisher on whether smoking caused lung cancer.

Statistical educators seem unwilling to deal with confounding in the introductory course.

- Confounding was not listed in McKenzie's (2004) top-30 topics in intro statistics.
- Most statistical educators use confounding to show that association is not causation, but never mention the topic thereafter: a classic case of "bait and switch"!
- Confounding is not mentioned in the index of many (most?) intro textbooks. (Schield, 2018). This finding was based on a convenience sample of 80 statistics textbooks. Of these 65% (52/80) contained no index reference to confounding or lurking variable, 25% (20/80) contained a few index entries and 10% (8/80) contained substantial content.
- Statistical educators disregarded Moore's call (1997) for two separate introductory statistics courses: statistical competence (a first course for those who must deal with data in their work); statistical literacy (what every educated person should know).

So why are statistical educators unwilling to discuss confounding in the intro course?

- Content. Moore (1997) wondered if a statistical literacy course was really "statistics".
- Mathematical allegiance. Causation is not a mathematical idea or operation. Avoiding causation and confounding maintains a closer allegiance to mathematics. Schield (2013)
- Statistical allegiance. Statistical educators may share Fisher's total disregard for observational studies. Freedman et al (1997) quoted Fisher's comment on observational studies, "That's not an experiment you have there, that's an experience."
- Concern. "When students see how easily statistics are influenced by other factors, this may bring our discipline into disrepute." Anonymous statistical educator.
- Ignorance. Most statistical educators never heard of Cornfield's minimum effect size needed to ward off nullification an observed association.

This paper addresses the last two explanations. Concern and ignorance.

BACK TO THE FUTURE: BATTLE OF THE TITANS

There is a similarity between statistical education today and the parents in the 80s hit movie: *Back to the Future*. Wikipedia (2018b). In that movie, Marty's dad and mom are not communicating. His parents had great plans in the past, but now they've stopped listening. Similarly, statistical educators have stopped listening to their students. They didn't listen when students said they saw less value in statistics after taking the course than they did before.

To change the future, we need to go back to 1959 and see Jerome Cornfield arguing that smoking caused cancer – based on data from observational studies. We need to see Sir Ronald Fisher arguing that the observationally-based association between smoking and cancer could be due to genetics: a confounder. Yes, Fisher was a smoker, but he produced data from a German twins study showing an association between the type of twinship (identical vs. fraternal) and smoking preference. We need to see how Cornfield derived the minimum effect size needed for a confounder to nullify or reverse an observed association. See Schield (1999).

Cornfield noted that smokers were at least 10 times as likely to develop lung cancer as non-smokers. Fisher's association had a relative risk of only three. Cornfield proved that Fisher's confounder was too small to nullify the observed association. Yes, there is no test for an unobserved confounder, but there is a necessary condition for an unobserved confounder to nullify an observed association. See Schield (1999), Gastwirth et al. (2000) and Cornfield (2018).

Cornfield's minimum effect size is as important to observational studies as is the use of randomized assignment to experimental studies. No longer could one refute an ostensive causal association by simply asserting that some new factor (such as a genetic factor) might be the true cause. Now one had to argue that the relative prevalence of this potentially confounding factor was greater than the relative risk for the ostensive cause. Schield (1999)

Cornfield's minimum effect size is one of the greatest contributions of statistics to human knowledge alongside the central limit theorem and Fisher's use of random assignment to statistically-control for pre-existing confounders. Cornfield's minimum size confounder eliminates the concern that a small confounder can overwhelm any observed association – regardless of size.

Consider the case of second-hand smoke. In a pooled analysis, the risk of lung cancer among spouses exposed to second-hand smoke is about 1.3 times as much as that among spouses not exposed and that this relative risk was statistically-significant. US Surgeon General (2006). Fisher's genetic confounder (RR=3) might easily nullify – and thereby totally explain – this weak association.

Consider the loss of smell (anosmia) among those taking Zicam: a homeopathic remedy for the common cold. With less than 100 cases among millions of users, the relative risk was extremely small and therefore not statistically-significant. Schield (2011). In their 2010 appearance before the US Supreme Court, the manufacturer (Matrox) argued that "statistical significance should be required as a threshold element of materiality in 10(b) cases based on undisclosed Adverse Effect Reports (AERs)". The US Supreme Court disagreed saying that when the adverse effect is so extreme, a statistically-significant association should not be used as a requirement for a claim against the manufacturer. Although relative risk was not cited as an issue, the presence of such an extreme adverse effect might justify treating a much lower relative-risk as supporting a causal connection.

Rosenbaum (2017) has an excellent treatment of quasi-experimental devices (Ch 8), sensitivity to [confounder] bias (Ch 9) and design sensitivity [to confounder bias] (Ch 10).

For some reason other statisticians seem to be united in ignoring Cornfield's game-changing contribution. Schield (2018) noted that most of those who reviewed Cornfield's contributions to statistics never mentioned his derivation of the minimum effect size: the Cornfield condition.

EFFECT SIZE AND CONFOUNDING

Effect size is another area where confounding is conspicuously absent. Effect size is the subject of recent books by Ellis (2010), Cumming (2011), and Grissom and Kim (2012). But none of these books included confounding in the index. Those touting effect sizes certainly recognize their importance. They may be aware that a large effect size was Hill's (1953) first criterion for identifying a causal relationship. But they may not understand *why* a big effect size is so important.

The bigger the association effect size, the bigger the confounder effect-size it can resist (the greater the number of confounders it can resist). A bigger effect size is more confounder resistant.

Schild and Burnham (2003) argued that when the association predictor is binary, the ability to ward off the influence of a given size binary confounder is directly related to relative risk (RR). This allowed Cornfield to argue that Fisher's genetic confounder (RR=3) was inadequate to nullify or overturn the observed effect size between smoking and lung cancer (RR=10).

EFFECT SIZE: HOW BIG?

The question, "How big is a strongly-resistant relative risk?" is similar to "How small is a statistically-significant p-value?" Neither has a mathematical answer. Both require a rule. Statistical significance used "one chance in 20" from Fisher. Confounder significance has no such foundation.

Furthermore, the shape of the sampling distribution is known whereas the shape of the confounder distribution is typically unknown. As a result, any confounder cutoff will be arbitrary. However, there are two kinds of arbitrary: totally arbitrary and somewhat – but not totally – arbitrary. E.g., A speed limit of 750 mph is totally arbitrary; a speed limit of 75 mph is somewhat arbitrary.

To avoid being totally arbitrary, it is best to use known information. (1) Relative risk (RR) measures effect size when predictor and outcome are binary. (2) Relative risk is never negative. (3) Taubes' (1995) list of 25 cancer associations had an average relative risk of 2.23.

None of these say anything about the distribution of confounder relative risks. But if we assume that the distribution of confounders is similar to the distribution of observed associations we can argue that (4) the distribution of confounder relative risks should have a maximum at one and decrease monotonically as RR increases and (5) the distribution of confounder relative risks should have an average relative risk of two (since small relative risks were presumably ignored).

A simple model of confounder relative risks should have a shape determined by the fewest parameters. Plausible candidates include the exponential distribution and a straight line. The exponential has a longer tail for a given average, so it is more conservative. That seems preferable.

To talk about confounder significance, we need to agree on a standard confounder distribution. Consider this proposal. A standard confounder distribution is an exponential distribution of confounder relative risks where the predictor values are exhaustive, the mean RR is two and the data is obtained from a cross-sectional study. See Figure 1.

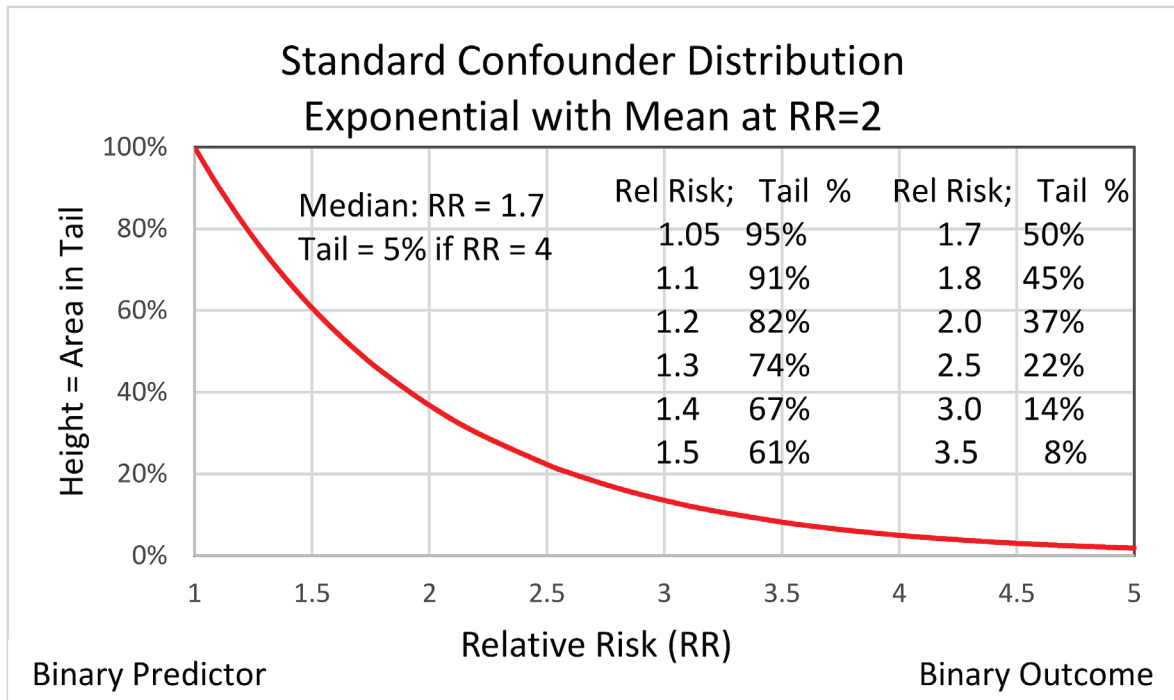


Figure 1: Standard Confounder Distribution

The minimum size confounder that can nullify an association must have a relative risk that is at least as large as the relative risk in the association. Schild and Burnham (2003).

If we adopt a "one in 20" standard for confounder resistance (or confounder significance), then a relative risk of four is required. Yes, this choice is arbitrary – perhaps doubly arbitrary in that both the distribution and the one-in-20 are arbitrary. But, a standard is needed to facilitate translating the continuous measure of relative risk into a binary classification: confounder resistant (or confounder significant) versus confounder non-resistant (or confounder non-significant).

If associations in scientific studies involving cross-sectional data need to be classified into five grades, consider using $RR > 4$ (A: significant), $3 < RR < 4$ (B: strong), $2 < RR < 3$ (C: moderate or medium), $1.5 < RR < 2$ (D: weak), $1 < RR < 1.5$ (F: ignorable).

Based on this standard confounder distribution, these five grades would have this distribution: Significant (5%), Strong (10%), Moderate (23%), Weak (24%) and Ignorable (38%).

These conservative levels may be lowered when making non-critical decisions, when using longitudinal before-after cohort data with a control group or when multiple studies indicate this distribution is overly conservative. These levels may need to be lowered when the consequences or risks are severe and immediate action is required.

Never forget that a large effect size is never sufficient by itself to conclude the association is causal. Pearl (2018) prompted this example. Suppose we divide people by shoe size (child or adult) and by the size of their vocabulary (child or adult), and calculate the percentage who have an adult vocabulary in the two groups. The resulting ratio may exceed a factor of ten. But this large factor is weak evidence for saying that a large shoe size causes a large vocabulary.

Cornfield's 1959 paper is a good example. Cornfield looked broadly and deeply at the evidence supporting the claim that smoking caused cancer. He noted the lab studies on rats and the dose-response relationship. He also looked at the evidence that opposed the claim. All the data supported smoking as a cause; no other explanation was plausible. As a statistician, Cornfield knew that an unknown super-confounder could be the underlying explanation. But, he noted, there was no evidence for this. Thus the claim that a super-confounder might exist was arbitrary – totally arbitrary. Cornfield, a history major, recognized that the totally arbitrary had no status in inductive reasoning.

TEACHING CONFOUNDING AND THE CORNFIELD CONDITIONS:

Students find that thinking about plausible confounders (hypothetical thinking and minimum effect size) is very different from working a traditional math problem. For details on how this can be taught and used to work testable problems, see Schield (2006 and 2016a).

Most importantly, students see value in studying observationally-based statistics since most of the statistics they see in the news or at work are observationally-based. Of the 105 students who recently took Augsburg's observationally-based statistical literacy course, 61% agreed that this course should be required by all college students for graduation. Schield (2016b). This result aligns with the MacNaughton (2004) goal for introductory statistics: to give students "a lasting appreciation for the value of statistics."

CONCLUSION:

Ideally, the GAISE guidelines will someday be updated to promote a statistical literacy course for decision makers and citizens that focuses on observational studies, confounding, strength of evidence and causation. In that course, statistical educators should teach the "Smoking causes cancer" debate, identify Cornfield's necessary conditions, and present Cornfield's condition as one of the greatest contributions of statistics to human knowledge.

Teaching a confounder-based introductory course will revolutionize the teaching of statistics. It will help students think critically about everyday statistics, identify spurious associations and even detect fake news. Teaching a confounder-based course will result in students seeing more value in statistics after completing an introductory course than they did before.

REFERENCES

- ASA (2016). College Majors Requiring Statistics. American Statistical Association 2016. www2.amstat.org/misc/CollegeMajors.pdf
- Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* 22 (1):173-203. Copy at <https://academic.oup.com/ije/article/38/5/1175/666926>

- Cornfield, Jerome (2018). Links to publications by and articles on. www.StatLit.org/Cornfield.htm.
- Cumming, Geoff (2011). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge Press.
- Ellis, Paul D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge University Press.
- Freedman, David, Robert Pisani and Roger Purves (1997). *Statistics*, 3rd ed. W. W. Norton & Co.
- Gastwirth, J, A; Krieger and P. Rosenbaum (2000). Cornfield's Inequality in *Encyclopedia of Epidemiological Methods*. P 262-265. Wiley. www.StatLit.org/cp/2000-Gastwirth-EEM.pdf
- Grissom, Robert J. and John J. Kim (2012). *Effect Sizes for Research*. 2nd edition. Routledge.
- Hill, A. Bradford (1953). Observation and Experiment. *New England Jrnl. Medicine* 248 995-1001
- Kaplan, Danny (2010). Our Abstinence-Based Curriculum and other Stories from Statistics Education. Slide presentation. Copy at www.statlit.org/pdf/2010KaplanASA-MN6up.pdf
- MacNaughton, Donald (2004). Goals of Your Introductory Statistics Course. Copy at <http://www.matstat.com/teach/p0002.htm>
- McKenzie, John (2004). Conveying the Core Concepts. *ASA Proceedings of the Section on Statistical Education*, P 2755-2757. Copy at www.statlit.org/pdf/2004McKenzieASA.pdf
- Moore, David (1997). Statistical Literacy and Statistical Competence in the 21st Century. IASE in South Korea. Slides. Copy at www.statlit.org/pdf/1997MooreASAslides.pdf
- Moore, David (1998). Statistical Literacy and Statistical Competence in the 21st Century. Handout. MSMESB. U. Iowa. Copy at <http://www.statlit.org/pdf/1998MooreMSMESB.pdf>
- Pearl, Judea (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.
- Rosenbaum, Paul (2017). *Observation and Experiment*. Harvard University Press.
- Rossmann, Alan (2016). Interview with Richard De Veaux. *Journal of Statistics Education*. Volume 24, 2016 - Issue 3. Page 157-168. Published online in 2017.
- Schild, Milo (1999). Simpson's Paradox and Cornfield's Conditions, *1999 ASA Proceedings of Section on Statistical Education*, pp. 106-111. See www.StatLit.org/pdf/1999SchildASA.pdf.
- Schild, Milo (2006). Understanding Confounding from Lurking Variables using Graphs. *STATS Magazine ASA*. Fall 2006. pp. 14-18. See www.StatLit.org/pdf/2006SchildSTATS.pdf
- Schild, Milo (2013). Statistics Education: Steadfast or Stubborn? *2013 ASA Proceedings of Section on Statistical Education*, pp. 1005-1017. See www.StatLit.org/pdf/2013-Schild-ASA.pdf.
- Schild, Milo (2011). Statistical Significance is Not Necessary for Causation (slides). 2011 ASA Invited Late-Breaking Session. See www.statlit.org/pdf/2011Schild-ASA-Zicam6up.pdf
- Schild, Milo (2016a). Stat 102: Social Statistics for Decision Makers. Invited paper. *Proceedings of the Roundtable Conference of IASE*. Go to www.StatLit.org/pdf/2016-Schild-IASE.pdf
- Schild, Milo (2016b) Augsburg Students Evaluations of STAT 102. *ASA Proceedings of the Section on Statistical Education*, p. 3741-3750. www.StatLit.org/pdf/2016-Schild-ASA.pdf
- Schild, Milo (2017). GAISE 2016 Promotes Statistical Literacy. *Statistics Education Research Journal (SERJ)* 16(1), 46-50. Copy at www.StatLit.org/pdf/2017-Schild-SERJ.pdf
- Schild, Milo (2018). Confounding and Cornfield: Prevalence in textbooks and articles. Research paper. Copy at www.StatLit.org/pdf/2018-Schild-ICOTS2.pdf
- Schild, Milo and Cynthia Schild (2008). Student Attitudes toward Statistics at Augsburg College. *2008 American Statistical Association Proceedings of the Section on Statistical Education*. [CD-ROM] P. 2150-2155. See www.StatLit.org/pdf/2008SchildASA.pdf
- Schild, Milo and Thomas Burnham (2003). Confounder-induced Spuriousity and Reversal. *2003 American Statistical Association Proceedings of the Section on Statistical Education*. [CD-ROM], 3690 - 3697. See www.StatLit.org/pdf/2003SchildBurnhamASA.pdf
- Taubes, G. (1995). Epidemiology Faces Its Limit. *Science, New Series*, Vol 269, #5221, 146-169. <https://www.mcgill.ca/cancerepi/files/cancerepi/TaubesScience269-164-Letters.pdf>
- Tintle, N., B. Chance, G. Cobb, A. Rossman, S. Roy, T. Swanson and J. VanderStoep (2014). Challenging the state of the art in post-introductory statistics. *Proceedings 59th ISI World Statistics Congress*. P. 295-300. <http://2013.isiproceedings.org/Files/IPS032-P1-S.pdf>
- US Surgeon General's Report (2006). *The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General*. <https://www.ncbi.nlm.nih.gov/books/NBK44330/>
- Wikipedia (2018a). AP Statistics. Number of examinees by year.
- Wikipedia (2018b). Back to the Future. A review of the movie.