

Simpson's Paradox in Meta-Analysis

To the Editor:

Simpson's paradox is used in the teaching of epidemiology to dramatize the effects of confounding. As Reintjes *et al*¹ note, however, there are few published real life instances of Simpson's paradox in epidemiology itself. Thus, despite the several engaging examples from other areas,² many teachers illustrate it using hypothetical data.

The example described by Reintjes *et al*¹ has a "double twist" that could give an ambiguous pedagogic message. We agree with their suggestion that the stratum-specific effect estimates are in the "wrong" direction and that the crude effect estimate reflects "the true causal" direction. They suspect that the stratum-specific "contrary results" stem from confounding by indication, and that there must be some second "unknown factor" which—in the crude data—cancels out and even reverses the direction seen within each stratum.

We now report a simpler, but equally striking, example of Simpson's paradox in epidemiology. The data come from a meta-analysis of case-control studies that examined the role of high voltage power lines in the etiology of leukemia in children. Five published reports contained sufficient data to reconstruct the numbers of cases and controls who resided within 100 meters of high voltage overhead power lines. These are given in the Table 1 below. The study-specific odds ratios range from 1.0 to 2.8, but the "crude" odds ratio from the aggregated cell frequencies is 0.7, outside of this range and in the opposite direction. The Mantel-Haenszel summary odds ratio is 1.3.

Usually, confounding is induced by subjects (exposed subjects are more likely or less likely than unexposed subjects to have other characteristics that independently contribute to a higher risk of disease) or by patients' physicians ("patients at high risk for urinary tract infection were more likely to be given prophylaxis"³). Here, in contrast, the distortion is investigator-induced. The investigators of studies 2³ and 3⁴ chose subpopulations living close to power lines (thereby augmenting the percentage who live within 100 m of the lines), while those of studies 1,⁵ 4,⁶ and 5⁷ chose entire populations. But studies 2³ and 3⁴ also had the lowest ratio of cases to controls! This co-occurrence leads to the extreme artifact in the odds ratio calculated from the aggregated frequencies.

In a logistic regression of pooled subject-level data, including covariates, from the five studies, the varying case-control ratios can be dealt with by using the logs of the case-control ratios as "offsets," *ie*, as study-specific intercepts with known values.

In meta-analyses of randomized clinical trials, it is common to derive an effect estimate directly from the aggregated frequencies. This practice produces little distortion, since most trials employ a 50:50 allocation. As our example shows, one cannot be as complacent in the meta-analysis of data from non-experimental studies.

James A. Hanley
Gilles Thériault

Departments of Epidemiology and Biostatistics and of Occupational Health
McGill University
1020 Pine Avenue West
Montreal, Quebec
H3A 1A2 Canada
(address correspondence to: James A. Hanley)

TABLE 1. Data from Five Case-Control Studies on the Role of High Voltage Power Lines in the Etiology of Leukemia in Children

Study (ref)		Distance to Lines		Total	Odds Ratio
		<100 m	>100 m		
1 ³	Cases	18	162	180	1.1
	Controls	25	252	277	
2 ⁴		43	414	457	1.6
	Cases	12	26	38	
	Controls	123	431	554	
3 ⁵		135	457	592	2.8
	Cases	27	121	148	
	Controls	104	475	579	
4 ⁶		131	596	727	2.0
	Cases	7	21	28	
	Controls*	3	25	28*	
5 ⁷		10	46	56	0.7
	Cases	14	353	367	
	Controls	7	359	366	
Total		21	712	733	0.7
	Cases	78	683	761	
		262	1542	1804	
		340	2225	2565	

* Exposure data available on entire population base; simulated as case-control study with 28 controls; the odds ratio from aggregated frequencies (0.7) remains essentially the same if one simulates multiple controls per case for this study.

References

1. Reintjes R, de Boer A, van Pelt W, Mintjes-de-Groot J. Simpson's paradox: an example from hospital epidemiology. *Epidemiology* 2000;11:81–83.
2. Moore DS, McCabe GP. *Introduction to the practice of statistics*. 3rd ed. New York: W.H. Freeman, 1999.
3. Myers A, Clayden AD, Cartwright RA, Cartwright SC. Childhood cancer and overhead powerlines: a case-control study. *Br J Cancer* 1990;62:1008–1014.
4. Feychting M, Ahlbom A. Magnetic fields and cancer in children residing near Swedish high-voltage power lines. *Am J Epidemiol* 1993;138:467–481.
5. Tynes T, Haldorsen T. Electromagnetic fields and cancer in children residing near Norwegian high-voltage lines. *Am J Epidemiol* 1997;145:219–226.
6. Li V-Y, Lee W-C, Lin RS. Risk of leukemia in children living near high-voltage lines. *J Occup Environ Med* 1998; 40:144–147.
7. McBride ML, Gallagher RP, Thériault G, Armstrong BG, Tamaro S, Spinelli JJ, Deadman JE, Fincham S, Robson D, Choi W. Power-frequency electric and magnetic fields and risk of leukemia in children. *Am J Epidemiol* 1999;149:831–842.

Two of the Authors Reply:

Confounding can play an important role in the analysis of epidemiologic studies. If not recognized, it can distort results and inferences. It is important, therefore, especially in teaching epidemiology, that sufficient attention be given to confounding. Epidemiologists should maintain critical skepticism of their data, especially as it is difficult to identify all confounders. It is important to realize that with epidemiologic studies, we are only able to describe a simplified version of the complex world. Simpson's paradox is a handy phenomenon for drawing attention to and facilitating the understanding of the principles and risks of confounding. Real live examples from published studies can be more illustrative than hypothetical examples. Therefore we are pleased to learn about another example of Simpson's paradox from a meta-analysis.

Hanley and Thériault state that the double twist found in the results of our paper could give an ambiguous pedagogic message. Our experiences from teaching epidemiology in different countries have shown that this so-called double twist catches the attention of students and is thus of increased pedagogic value. In our paper we proposed an intuitive expla-

nation of the paradox, showing in our Figure that the overall data set can reveal a negative correlation between exposure and outcome while within each of several subsets of the data (strata) a positive correlation exists. In Hanley and Thériault's example five independent case-control studies were combined. Confounding is especially important in meta-analysis, as the selection of studies adds another potential confounding factor. This new factor could have an additive effect, but could also work in the opposite direction, even causing a reversal (Simpson's paradox), as shown by Hanley and Thériault. This example illustrates an aspect of confounding that would be helpful to students of epidemiology, especially in the context of meta-analysis.

Ralf Reintjes
Annette de Boer

Dezernat Infektionsepidemiologie
Landesinstitut für den Öffentlichen Gesundheitsdienst NRW
PO Box 3809
48151 Münster
Germany
(address correspondence to: Ralf Reintjes)

Oral-Contraceptive Use, Anovulatory Action, and Risk of Epithelial Ovarian Cancer

To the Editor:

It is established that parity lowers the risk of developing epithelial ovarian cancer, and that the risk reduction associated with each pregnancy is greater than what would be expected based solely on the suppression of ovulation.¹ Oral-contraceptive (OC) use is also known to reduce the risk; it is uncertain, however, whether such protection is consistent with the amount of anovulation caused by these products.¹ Siskind *et al*² recently examined this question. In their large and well-conducted case-control study, Siskind *et al* found that the magnitude of risk decrease for each year of OC use is 7% (95% CI = 4–9%), matched on age and adjusted for parity, smoking, hysterectomy, tubal ligation, and calculated lifetime number of ovulatory cycles. Because of the multivariate control of number of ovulatory cycles, the authors concluded that OCs have a protective influence on ovarian cancer beyond the anovulatory action.² While this conclusion could be true, it does not follow from their analysis at all, and in fact the statistical support for this conclusion from their data is low.

To compare the observed effect of OC use with what would be expected according to the duration of anovulation, it is required to compare the risk amount per year of OC use (7%) with the risk amount per year of ovulation. The latter value, while not published by Siskind *et al*,² may be inferred from their data to be about 2%, based on the observed change in the effect of OC use with and without adjustment for number of ovulations and parity. Since their regression model containing lifetime number of ovulations also included additional terms for parity and OC use, the statistical information determining the 2% value comes essentially from age at menopause/diagnosis/interview less age at menarche. This variable is much less precisely related to risk than either parity or OC use. That is, its 95% confidence interval is wide and includes the 7%. Therefore, the risk reduction of 7% per year of OC use is consistent with the 2% per year of ovulation.

How large should the increase in risk be per year of ovulation? For most women, ovulations occur over at least 20 years. Thus, as we have observed,¹ the risk reduction for a year of anovulation

should be no greater than 5% or so. This amount also is consistent with the 95% confidence interval (4–9%) of the 7% for each year of OC use. Thus, there are no grounds to conclude from the study of Siskind *et al*² that OC use conveys a magnitude of protection beyond that from an equivalent duration of anovulation. On the other hand, contrary to the authors' assertion, a full treatment of this question has been given by Risch *et al*,³ where the magnitude of protection per year of OC use (11.8%) differed from the effect per year of ovulation (2.9%).

Harvey A. Risch

Department of Epidemiology and Public Health
Yale University School of Medicine
60 College Street, PO Box 208034
New Haven, CT 06520-8034
(address for correspondence)

References

1. Risch HA. Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone. *J Natl Cancer Inst* 1998;90:1774–1786.
2. Siskind V, Green A, Bain C, Purdie D. Beyond ovulation: oral contraceptives and epithelial ovarian cancer. *Epidemiology* 2000;11:106–110.
3. Risch HA, Weiss NS, Lyon JL, Daling JR, Liff JM. Events of reproductive life and the incidence of epithelial ovarian cancer. *Am J Epidemiol* 1983;117:128–139.

The Authors Reply:

We are grateful to Risch for giving us the opportunity to clarify further the question of oral contraceptives versus anovulation in conferring protection against epithelial ovarian cancer. We are currently undertaking an analysis concentrating on ovulatory life and its contributory factors in relation to ovarian cancer, which we hope will address in depth the issues raised by Risch.

In the interim, we shall respond to Risch's comments by pointing out that if oral contraceptive (OC) use influences ovarian cancer risk solely by suppressing ovulation, the inclusion of an estimate of lifetime ovulations ("ovulatory life") in the analytic model should wholly nullify the OC effect. In other words the expected value of the relative risk of oral contraceptive exposure, however parameterized, should be unity in the presence of the ovulatory life variable. In our paper, we estimated that a year of OC use was associated with a 7% reduction in ovarian cancer risk (95% confidence interval 4–9%) in a model that included a carefully constructed estimate of ovulatory life. We argued further that even if errors in computing the latter value had led to imperfect control of confounding, the effect of OC use would not be seriously overestimated.

We apologize for our oversight in claiming priority for our analysis including both OC use and ovulatory life, which properly belongs to Risch *et al*'s 1983 paper.¹ We were aware of its existence but had overlooked their results owing to differences in terminology and analytic method.

Victor Siskind
Adèle Green
David Purdie

The Queensland Institute of Medical Research
Royal Brisbane Hospital
Queensland 4029
Australia

Christopher Bain

Department of Social and Preventive Medicine
University of Queensland
Herston
Australia