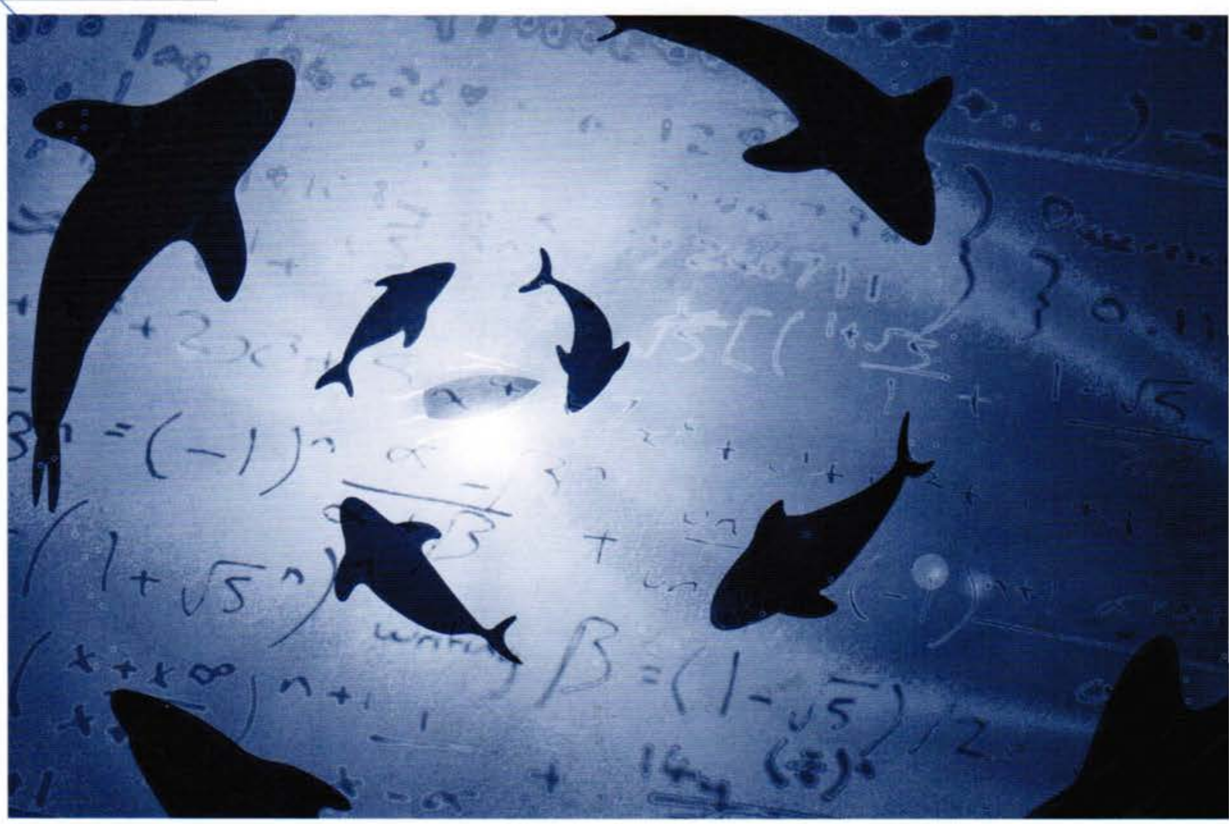
THE MAGAZINE FOR STUDENTS OF STATISTICS

FALL 2006 • ISSUE # 46

STATS



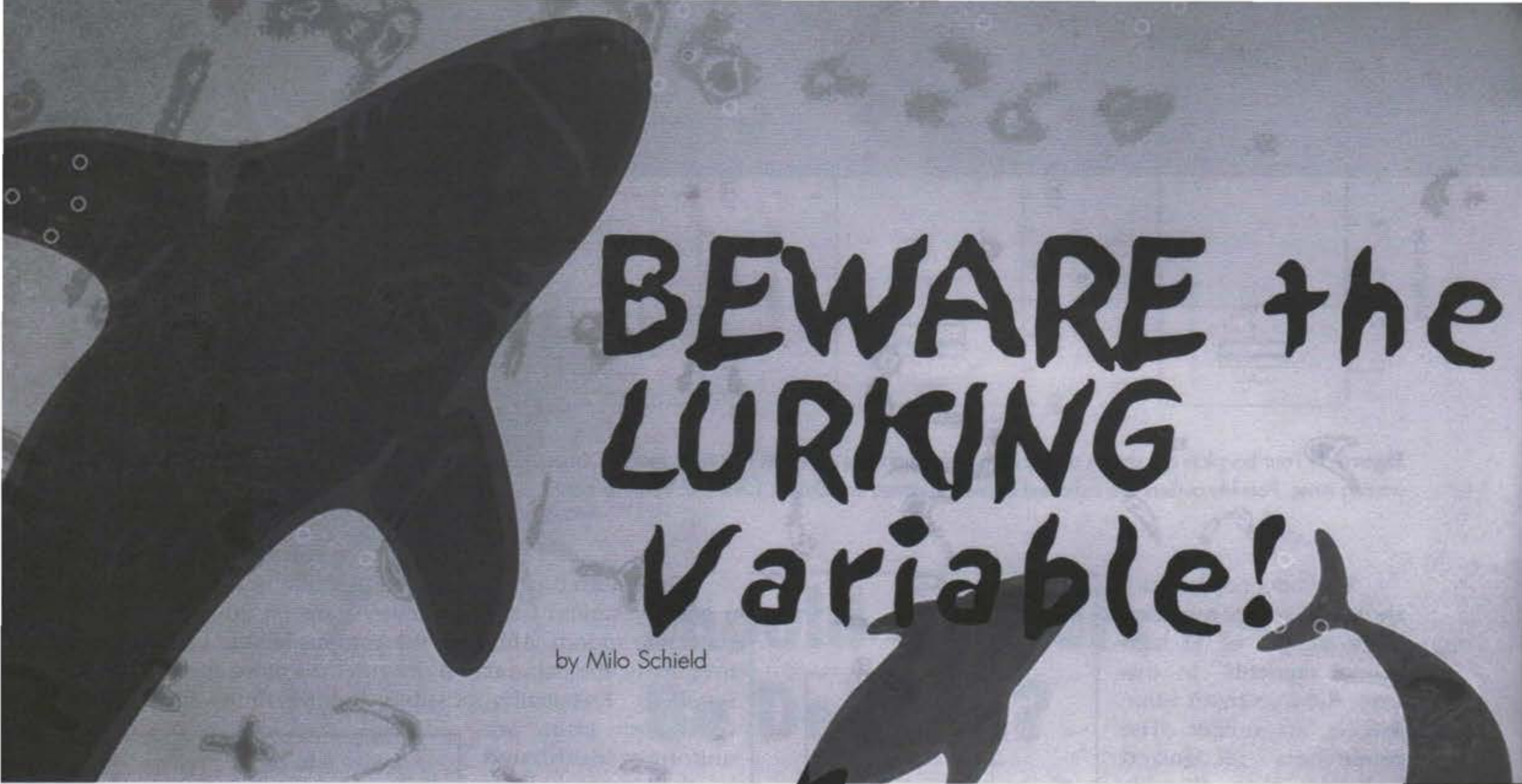
Beware the LURKING Variable

Understanding Confounding from Lurking Variables Using Graphs

Random Numbers from Nonrandom Arithmetic

Observational Studies: The Neglected Stepchild in the Family of Data Gathering

Non Profit Org.
U.S. Postage
PAID
Permit No. 361
Alexandria, VA



BEWARE the LURKING Variable!

by Milo Schield

Understanding Confounding from Lurking Variables Using Graphs

Did you know the United States has a higher death rate than Mexico? It's a fact. In 2003, the death rate was 80% higher in the United States than in Mexico (8.4 per 100,000 compared to 4.7 per 100,000).

What does this statistic mean? Does Mexico have better health care than the United States? That seems unlikely. Yet it is difficult to claim that this unexpected result is due to chance, error, or bias. The populations being studied are large, and death is definite, therefore usually counted accurately. You may be perplexed further when you learn that death rates are even lower in Ecuador and Saudi Arabia (4.3 per 100,000 and 2.7 per 100,000).

A possible explanation is ***confounding***: "a situation in which the effects of two processes are not separated," according to John M. Last's *A Dictionary of Epidemiology*. Confounding can be due to a ***lurking variable***. Often referred to as a confounder, Last says a lurking variable "can cause or prevent the outcome of interest ... and [is] associated with the factor under investigation."

Lurking variables are called "lurking" because they are not recognized by the researcher as playing a role in the study. Although they can influence the outcome of the process being studied, their effect is mixed in with the effects from other variables.

In comparing the death rates in the United States and Mexico, a lurking variable may be the difference in the age distributions within each population. Mexico has a much younger population than the United States. In 2003, there were 59% more people under 15 in Mexico than in the United States (32% of the Mexican population, compared to 21% of the United States population). In addition, there were more than twice as many people 65 or older in the United States as in Mexico (12% compared to 5%).

It's a fact that older people are much more likely to die than younger people. Unless we take age into account, a comparison of the crude (not accounting for age) death rates may be misleading. Mexico's comparatively low death rate is more likely due to its youthful population, rather than to its health care system.

So how can we untangle this confusion? How can we take into account the influence of a lurking variable that confounds an association?

Standardizing

Standardization is used in demography to 'take into account' the distribution of ages within a population. It can take into account the influence of a related factor

when comparing ratios for two groups so we are not comparing "apples to oranges." When the death rates of Mexico and the United States are standardized for age, the death rate in Mexico is higher than that in the United States.

Standardization also can take into account the influence of a related factor when comparing ratios over time for the same group. For example, according to the *2001 United States Statistical Abstract*, the crude death rate due to pneumonia was 7.4% higher in 1996 than in 1990 (33.4 per 100,000 compared to 31.1 per 100,000). But when standardized on the 1940 United States population distribution, the age-adjusted death rate due to pneumonia was 5.1% lower in 1996 than in 1990 (13.0 per 100,000 compared to 13.7 per 100,000). In this case, standardizing actually reversed the direction of the association.

Standardizing Ratios Graphically

To 'see' standardization, it would be nice to have a simple technique—ideally graphical—that will take into account or 'adjust for' the influence of a lurking variable.

In an article that appeared in *The Roles of Representation in School Mathematics*, Lawrence Lesser featured a graphical technique for showing how an association can be influenced when the lurking variable has just two values. The graph shows how a weighted average can be obtained easily without algebra. Howard Wainer did the same in a 2002 *CHANCE* article, "The BK-Plot: Making Simpson's Paradox Clear to the Masses." Milo Schield used this technique to illustrate standardization in "Three Graphs To Promote Statistical Literacy," presented at the 2004 International Congress on Mathematical Education. To see how it works, let's consider some examples.

Patient Death Rates by Hospital

Table 1 and Table 2 present the underlying data (hypothetical) for two hospitals: Rural Hospital and City Hospital. Patients in good condition can walk in; patients in poor condition are carried in.

Table 1. Death Rates of Patients by Hospital and by Condition

Death Rate	Patient Condition		
	Good	Poor	All
Rural	2.0%	7.0%	3.5%
City	1.0%	6.0%	5.5%
All	1.5%	6.5%	

We want to analyze the association between hospital (predictor) and death rate (outcome). First, we plot the data from Table 1 in Figure 1. City Hospital has a death rate of 6% for patients in poor condition and 1% for patients in good condition. Connecting these data values gives the heavy dashed line. Rural Hospital has a death rate of 7% for patients in poor condition and 2% for patients in good condition. Connecting these data points gives the light dashed line.

Table 2. Number of Patients by Hospital and by Condition

Number of Patients	Patient Condition		
	Good	Poor	All
Rural	700	300	1,000
City	100	900	1,000
All	800	1,200	2,000

From Table 2, we can see that 90% of the patients in City Hospital are in poor condition, while only 30% of those at Rural Hospital are in poor condition. Plotting these percentages in Figure 1 gives the death rates at City Hospital and Rural Hospital.

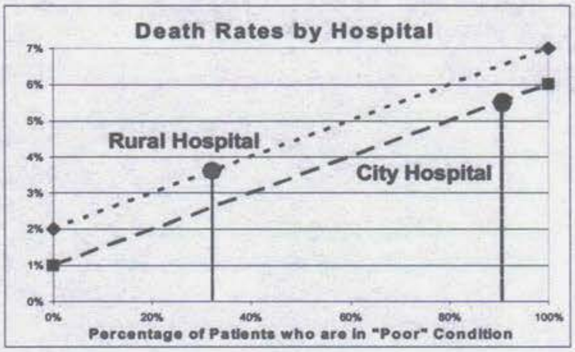


Figure 1. Hospital death rates by percentage of patients in poor condition

The death rate is much higher at City Hospital (5.5%) than at Rural Hospital (3.5%). Based on this, Rural Hospital would seem like a better hospital than City Hospital. But notice that City Hospital has a lower death rate than Rural Hospital for patients in good condition and those in poor condition. This is an example of Simpson's Paradox. Simpson's Paradox occurs when an association has one direction at the group level, but the opposite direction in each of the sub-groups.

Before we shut down City Hospital as "the hospital of death," we need to consider whether City's higher death rate could be confounded by patient condition. Note that patient condition is associated with the outcome of interest (death) and with the predictor (hospital). Being in poor condition is positively linked with dying. Dying is more likely for patients in poor condition (6.5%) than for those in good condition (1.5%). See Table 1. Being in poor condition is positively linked with City Hospital. The percentage of patients who are in poor condition is greater at City (90%) than at Rural (30%). See Table 2.

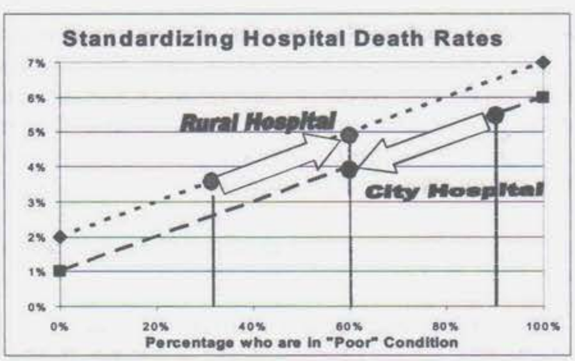


Figure 2. Hospital death rates standardized based on patient condition

To make a fairer comparison of these hospitals, we need to standardize their mix of patients. Let's standardize both hospitals on their combined mix (60%). Using the group average as the standard emulates the desired outcome in a randomized experiment where the goal is for each group (exposure and control) to have the same percentage of confounder as found in the overall population.

Standardizing the mix in both groups at 60% increases the expected death rate at Rural Hospital and decreases it at City Hospital, as shown in Figure 2. The standardized death rate is lower for City Hospital than for Rural Hospital (4% compared to 5%). In this case, the direction of the association between the standardized rates is the reverse of that between the crude rates—and we have a fair comparison of the two hospitals; we are comparing "apples and apples."

Family Incomes by Race

Here is another case. Suppose that in the United States in 1994, mean family income was 66% more for whites than for blacks ($54,500 compared to $32,900, as estimated based on the *United States Statistical Abstract*). (See Table 3) Is the black-white income gap fully explained by only race? The $21,600 white-black income gap could be confounded by a related factor: family structure.

Table 3. Estimated Family Incomes by Race and Family Structure

Family Income	Head of Family		
	Unmarried	Married	All
White	$26,700	$60,600	$54,500
Black	$14,000	$53,900	$32,900
All	$23,000	$60,100	$51,900

Family income is higher for married-couple families ($60,100) than for single-parent families ($23,000). In order to standardize data, we need the distribution of families by family structure within each race, as shown in Table 4.

Based on Table 4, families headed by a married couple is more likely among whites than among blacks (82% compared to 47.5%). Figure 3 summarizes this data so it can be standardized graphically.

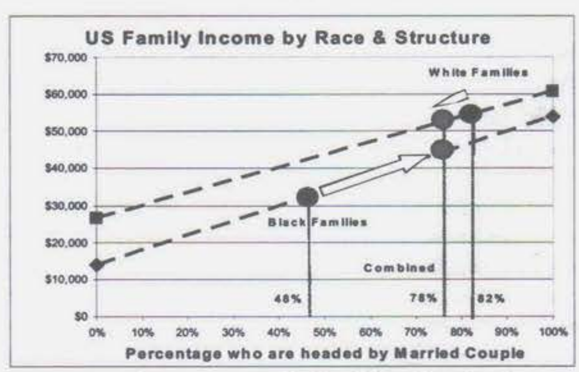


Figure 3. Family incomes standardized based on family structure

To take into account the influence of family structure, let's standardize the mix of family types to a standard mix: the overall percentage of families who

Table 4. Number of Families by Race and Family Structure

Families, 1994	Head of Family		
	Unmarried	Married	All
White	10,539	47,905	58,444
Black	4,251	3,842	8,093
All	14,790	51,747	66,537

are married (78%). Standardized family income is 18% more for whites ($53,000) than for blacks ($45,000). Standardizing on family structure decreases the black-white income gap by 65%, from $21,600 to $8,000. Thus, 65% of the black-white family income gap is explained by family structure.

Auto Death Rates by Airbag Presence

We generally think airbags are good. That conclusion is supported by the data in Table 5, which appeared in Mary C. Meyer and Tremika Finney's *CHANCE* article, "Who Wants Airbags?". For the occupants of automobiles in accidents, the death rate is lower for those with an airbag than for those without (37 per 10,000 compared to 60 per 10,000).

Table 5. Death Rate per 10,000 Automobile Accident Occupants

Death Rate	Seatbelt Used		
Airbag	No	Yes	All
No	105	26	60
Yes	122	18	37
All	111	21	

But wait! For those not using a seatbelt (left column), the death rate was higher for those with an airbag than for those without (122 per 10,000 compared to 105 per 10,000). The association between airbags and death rate may be confounded by seatbelt usage. Consider the distribution of automobile accident occupants as shown in Table 6.

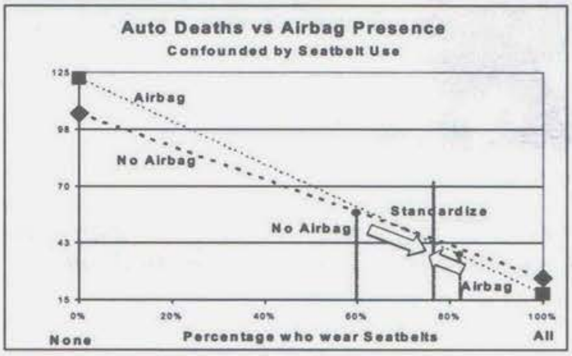


Figure 4. Automobile accident death rates with and without an airbag, standardized based on seatbelt usage

Using a seatbelt is positively associated with a lower death rate, as the death rate was much higher for those who didn't use a seatbelt at all than for those who used one (111 per 10,000 compared to 21 per 10,000 in Table 5). And using a seatbelt is positively associated with having an airbag, as the percentage using a seatbelt is greater among people in cars with airbags than in cars without (85% compared to 60% in Table 6).

Let's standardize on the overall percentage of people in accidents who were wearing a seatbelt (73%), as shown in Figure 4.

Table 6. Automobile Accident Occupants Using Seatbelts and/or Having Airbags

Number (1,000)	Seatbelt Used		
Airbag	No	Yes	All
No	1,952	2,903	4,855
Yes	871	4,872	5,743
All	2,823	7,775	10,598

The standardized death rate of occupants in auto accidents is slightly higher for those with air-bags than for those without (47 per 10,000 compared to 46 per 10,000). So, do airbags save lives? Not on average for this mix of occupants. This situation is complex because there is an interaction between having airbags and using seatbelts. We can see this because the lines cross. The main point is that seatbelts make a bigger difference in saving lives! Without taking into account the effect of seatbelts, the effect of airbags is almost masked due to the confounding and interaction.

Analysis of Confounding

Now that we have seen how a lurking factor can confound our understanding of a statistical association, it is good to reflect on what causes these situations and what we can do to avoid them.

Notice what is common to the three examples we have examined. In each case, the researcher was an observer. The researchers did not (and could not) assign patients to a particular hospital, determine which families were headed by a married couple, or determine which car owners bought cars with an airbag. Studies in which the researcher is passive in assigning subjects to exposure and control groups are called observational studies. While the influence of chance decreases as sample size increases, the influence of a confounder remains unchanged in observational studies. The influence of confounding can be a major problem—if not the main problem—in social sciences research, according to Stanley Lieberson in *Making It Count*.

Confounding also can arise in any study—observational or experimental—where the response due to a factor is observed or measured at a single level and the choice of level influences the association. Because most studies in the news are observational, understanding confounding is absolutely necessary to being statistically literate.

A Problem from Baseball

To test your understanding of this graphical technique, try working out this problem from baseball.

Ted and Sam are on the same baseball team. Both players have been to bat 100 times. Sam had 26 hits and Ted had 34 hits. So, Sam's batting average is .260 (26%) and Ted's is .340 (34%). But the coach thinks Sam is the better hitter. Could this be due to Simpson's Paradox? Could the strength of the pitcher be a factor? Could the percentage of times each player faced a strong pitcher be a lurking variable? If the pitcher was weak, Sam hit 50% of his times at bat while Ted hit 40% of the time. When facing strong pitchers, Sam hit 20% of the time while Ted hit only 10% of the time. Who is the better hitter? To answer this question, standardize their averages as if each faced strong pitchers 50% of the time. After you have worked this problem, check your answer with the answer on Page 21.

Conclusion

The influence of context on comparisons of ratios can be profound. Context is an essential difference between statistics and mathematics. To understand the influence of context on a statistic or a statistical association, it helps to understand the confounding effect of lurking variables.

Confounding from lurking variables is the reason that "association is not necessarily causation." With this understanding, we have a stronger reason to be careful in using statistical association as evidence for casual connections. A statistical association is only the first step in establishing causation.

Confounding and standardization are two of the most important ideas in statistics. Once we recognize that standardizing (taking into account confounding) can change the size of a comparison—and may even reverse the direction (Simpson's Paradox)—we have taken a big step toward being statistically literate.

Viewing confounding as the influence of context increases our statistical literacy and provides a link between statistics and other areas of study, including the social sciences and the humanities. For more about this, see Schield's "Statistical Literacy and Liberal Education at Augsburg College," available at *www.StatLit.org/pdf/2004SchieldAACU.pdf.* ■

Editor's Note: *The author would like to thank the W. M. Keck Foundation for their grant "to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts" and Tom Burnham, Cynthia Schield, and Marc Isaacson for editorial assistance.*

Additional Reading

Lesser, L. (2001). "Representations of Reversal: Exploring Simpson's Paradox." in Albert A. Cuoco and Frances R. Curcio (Eds.) *The Roles of Representation in School Mathematics (2001 Yearbook)*. National Council of Teachers of Mathematics, 129-145.

Last, J. (1995). *A Dictionary of Epidemiology*. Third Edition, Oxford University Press.

Lieberson, S. (1985). *Making It Count*. University of California Press.

Meyer, M. and Finney, T. (2005). "Who Wants Airbags?" *CHANCE*, 18(1):3-16.

Schield, M. (1999). "Simpson's Paradox and Cornfield's Conditions." *Proceedings of the 1999 Joint Statistical Meetings*, Section on Statistical Education, 106-111.

Schield, M. (2004a). "Three Graphs To Promote Statistical Literacy." Presented at the 2004 International Congress on Mathematical Education, Copenhagen. Available at *www.StatLit.org/pdf/2004SchieldICME.pdf.*

Schield, M. (2004b). "Statistical Literacy and Liberal Education at Augsburg College," Peer Review, 6(3). Available at *www.StatLit.org/pdf/2004SchieldAACU.pdf.*

Wainer, H. (2002). "The BK-Plot: Making Simpson's Paradox Clear to the Masses." *CHANCE*, 15(3):60–62.

Milo Schield

Milo Schield (schield@augsburg.edu) has taught statistics and statistical literacy for more than 20 years at Augsburg College in Minneapolis, MN. With a BS from Iowa State, an MS from the University of Illinois, and a PhD from Rice University, he has pursued a variety of professional interests, including operations research at a large property-casualty insurance company. Currently, he is the director of the W. M. Keck Statistical Literacy Project at Augsburg College. See Schield (2004b) and www.statlit.org *for details.*

study has been instrumental in developing the current understanding of heart disease and has led to more than 1,200 publications about risk factors for coronary disease. The cost of this study over the years, however, is measured in the tens of millions of dollars.

QUALITY OF INFORMATION FROM OBSERVATIONAL STUDIES

The money spent on observational studies can be well worth it if the studies produce sound results. Otherwise, if confounding is an overwhelming problem, the studies might be leading us in the wrong direction. Several recent efforts have been made to evaluate the quality of results from observational studies, and, in general, the results are promising. It seems that with well-designed observational studies, the risks of confounding can be limited. Concerns have been expressed that observational studies tend to exaggerate treatment effects. A *New England Journal of Medicine* article by J. Concato, N. Shah, and R. Horowitz, "Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs," indicates that "well-designed observational studies (with a cohort or case-control design) did not systematically overestimate the magnitude of the associations between exposure and outcome as compared with the results of randomized, controlled trials." In fact, it seems there can be more variability in the outcomes of the randomized, controlled trials than in the observational studies.

For our example, appendectomies, there have been a number of studies that have compared the results of laparoscopy and open surgery. The consensus wisdom is that laparoscopic surgery, the less invasive alternative, is the preferred method for straightforward cases. This consensus has been established by a combination of both observational studies and randomized, controlled trials. A comparison by K. Benson and A. Hartz of eight observational studies and 16 randomized, controlled trials revealed that seven of the eight observational studies found an advantage for laparoscopy (and the eighth no difference), while, among the 16 randomized trials, eight favored laparoscopy, three favored open surgery, and the remaining five had very close results.

Conclusion

To use a randomized, controlled trial, the research question must be relatively mature—with some confidence that a treatment is meaningful—before the cost of the trial can be justified. An observational study can set the stage. In many situations, an observational study is a first step toward understanding by roughly identifying associations that can be more closely examined with either a more rigorous observational study or a randomized, controlled trial. The issue of confounding with observational studies is real, but we can benefit from an observational study by having a more developed view that places the value of a randomized, controlled trial in a more meaningful context. ■

Additional Reading

Benson, K. and Hartz, A. (2000). "A Comparison of Observational Studies and Randomized, Controlled Trials." *New England Journal of Medicine*, (342): 1878-1886.

Concato, J. Shah, N. and Horowitz, R. (2000). "Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs." *New England Journal of Medicine*, (342), 1887-1892.

National Heart, Lung, and Blood Institute. (2002). Framingham Heart Study. Available at *www.nhlbi.nih.gov/about/framingham/index.html*.

Pai, M. (Accessed 2006). "Observational Study Designs," available at *http://sunmed.org/Obser.html*.

Answer to Baseball Problem

from Page 18

The players' standardized batting averages are .350 (35%) for Sam and .250 (25%) for Ted. After taking into account (controlling for or conditioning on) the strength the pitcher, Sam's batting average is higher than Ted's. So, the Coach is right – Sam is the better hitter.