Presenting Confounding and Standardization Graphically

Milo Schield[1]

Did you know that the US has a higher death rate than Mexico? It's a fact. In 2003, the death rate was 80% *higher* in the US than in Mexico (8.4 per 100,000 versus 4.7).

What does this statistic mean? Does Mexico have better health care than the US? That seems very unlikely.

Yet it is difficult to claim that this unexpected relationship is due to chance, error or bias. The populations being studied are large; death is definite and usually counted accurately.

You may be perplexed, confused or confounded when you learn that death rates are even lower in Ecuador and Saudi Arabia (4.3 and 2.7).

An alternate explanation is *confounding.* Last (1995) defines confounding as "a situation in which the effects of two processes are not separated." Confounding reflects the influence of a *lurking variable*. A lurking variable is often referred to as a confounder which Last defines as "a variable that can cause or prevent the outcome of interest … and is associated with the factor under investigation."

In comparing these death rates, a lurking variable may be the difference in the age distributions. Mexico has a much younger population than the US. In 2003, people under 15 were 50% more prevalent in Mexico than in the US (32% compared to 21%). People 65 and older were more than twice as prevalent in the US 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 distribution into account, a comparison of these crude (unadjusted) death rates may be misleading. Mexico's comparatively low death rate is most 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.

Standardization can take into account the influence of a related factor when comparing ratios for two different groups at the same time. When the death rates of Mexico and the US are standardized for age, the death rate in Mexico is *higher* than that in the US.

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

Standardization can even take into account the influence of a related factor when comparing two ratios that have different times and places.

Standardizing Ratios Graphically

Standardizing is typically a complex process. When the lurking variable is continuous (e.g., age), standardization typically involves multivariate regression. When the lurking variable has multiple levels (e.g., age groups), standardizing typically involves some complex algebra.

To really "see" standardization it would nice to have a simpler technique – ideally graphical – that will "adjust for" or "take into account" the influence of a lurking variable.

Lesser (2001) and Wainer (2002) featured a new graphical technique for showing how an association can be influenced when the lurking variable has just two values. This new graph shows how a weighted average can be easily obtained without algebra. Schield (2004a) used this graph to illustrate standardization. To see how this graphical technique works, consider some examples.

Patient Death Rates by Hospital

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

[1] Milo Schield, (schield@augsburg.edu) has taught statistics and statistical literacy for over 20 years at Augsburg College in Minneapolis, MN. With a BS from Iowa State, an MS from the University of Illinois and a Ph.D. from Rice University he has pursued a variety of professional interests including Operations Research at a large property-casualty insurance company. Milo is the Director of the W. M. Keck Statistical Literacy Project at Augsburg College. See Schield (2004b) and www.StatLit.org for more details.

DEATH RATE	PATIENT CONDITION		
	GOOD	POOR	ALL
RURAL	2%	7%	3.5%
CITY	1%	6%	5.5%
ALL	1.5%	6.5%	

Table 1: Death Rates of Patients

We want to analyze the association between hospital (predictor) and death rate (outcome). But we notice that association is tangled up (confounded) by patient condition (a lurking variable).

First we plot the data from Table 1 in Figure 1. City hospital has a death rate of 6% for patients in poor condition (1% for patients in good condition). Connecting these two data values gives the bottom weighted-average line. Rural hospital has a death rate of 7% for patients in poor condition (2% for patients in good condition). Connecting these two data points gives the top weighted-average line.

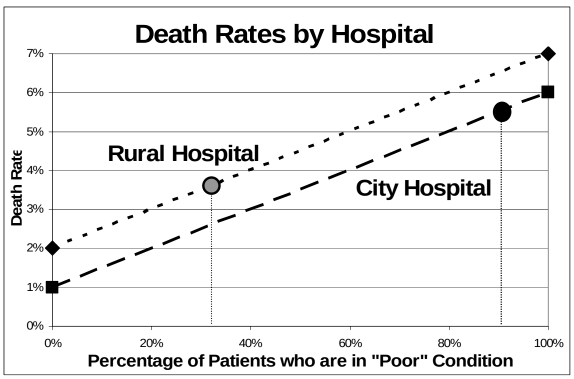

Figure 1: Hospital Death Rates Plotted

From Table 2 you can see that patients in poor condition are 90% of those at City (30% of those at Rural). Plotting these percentages in Figure 1 gives the death rates at City and Rural hospitals.

The death rate is *much higher* at City (5.5%) than at Rural (3.5%). Based on this, Rural seems like a better hospital than City. But notice that City has a lower death rate than Rural for patients in good condition and for 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 subgroups.

Consider the distribution of patients in Table 2.

NUMBER OF PATIENTS	PATIENT CONDITION		
	GOOD	POOR	ALL
RURAL	700	300	1,000
CITY	100	900	1,000
ALL	800	1,200	2000

Table 2: Number of Patients

Note that the percentage of patients who are in poor condition is higher for City than Rural (30% vs. 90%). Before we shut down City hospital as "the hospital of death", consider whether this higher death rate could be confounded by a related factor: patient condition. Note that patient condition is associated with the outcome of interest (death) and is associated with the predictor (hospital).

Being in poor condition is positively linked with dying. Dying is more likely for patients in poor condition than for those in good condition. See Table 1 (6.5% vs. 1.5%).

Being in poor condition is positively linked with the City Hospital. The percentage of patients who are in poor condition is greater at City than at Rural. See Table 2 (90% vs. 30%).

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%) as shown in Figure 2.

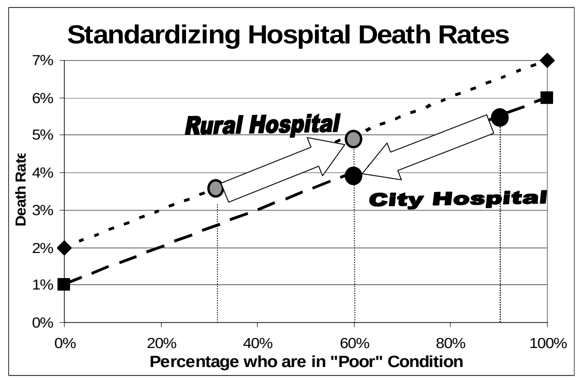

Figure 2: Hospital Death Rates Standardized

Standardizing the mix in both groups at 60% increases the expected death rate at the rural hospital and decreases it at the city hospital.

The standardized death rate is *lower* for City than for Rural (4% vs. 5%). In this case the direction of the association between the standardized rates is the reverse of that between the crude rates.

Family Incomes by Race

Here is another case. Suppose that in the US in 1994, mean family income was 66% *more* for whites than for blacks ($54.5K compared to $32.9K).[2] Is the

[2] Mean family incomes were estimate by multiplying median family incomes by (4/3). The factor of (4/3) was obtained from the following data. In Table 726 of the 1995 US Statistical Abstract, the 1993 mean household income was $41,428. In Table 727, the 1993 median household income was $31,241. The mean-median ratio for 1993 household incomes was 1.33. Mean incomes were obtained from the1994 median incomes in the 1997 US Statistical Abstract: $40,884 for all whites and $24,698 for all blacks while those for families headed by married couples were $45,474 for whites and $40,432 for blacks.

black-white income gap evidence of racial discrimination?

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

Table 3: Estimated Family Incomes

This $21,600 white-black income gap could be confounded by a related factor: family structure. Family income is *higher* for married couple families ($60.1K) than for single parent families ($23K). In order to standardize data, we need the distribution of families by family structure within each race as shown in Table Table 4: Number of Families.

FAMILIES US, 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

Table 4: Number of Families

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

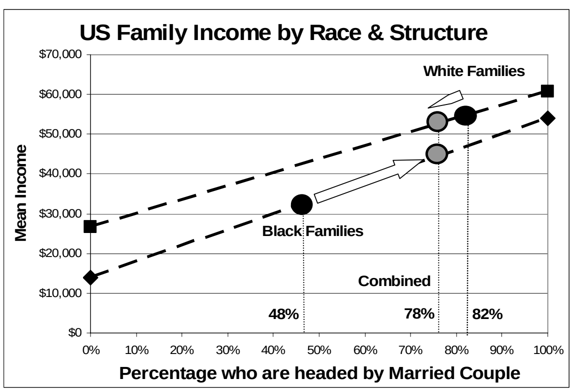

Figure 3: Standardized Family Incomes

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 are married (78%).

Standardized family income is 18% ($8K) *more* for whites ($53K) than for blacks ($45K).

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.

Some of the black-white income difference may still be due to discrimination but after taking into account family structure, it is now $8,000 instead of $21,600.

Auto Death Rates by Airbag Presence

You probably think that airbags are a good thing. That conclusion is supported by the data in Table 5 from Meyer and Finney (2005). For the occupants of autos in accidents, the death rate is *lower* for those with an airbag than for those without an airbag (37 to 60).

DEATH RATE	SEATBELT USED		
AIRBAG	NO	YES	ALL
NO	**105**	**26**	**60**
YES	**122**	**18**	**37**
ALL	**111**	**21**	

Table 5: Death Rate/10,000 car accident occupants

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 compared to 105). The association between airbags and death rate may be confounded by seatbelt usage. Consider the distribution of car accident occupants as shown in Table 6.

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**

Table 6: Car accident occupants

Using a seatbelt is positively associated with a lower death rate since the death rate was *much higher* for those who didn't use a seatbelt at all than for those who used one. See Table 5: 111 vs. 21. And using a seatbelt is positively associated with having an airbag since the percentage who use a seatbelt is *greater* among those in cars with airbags than those in cars without. See Table 6: 85% vs. 60%.

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

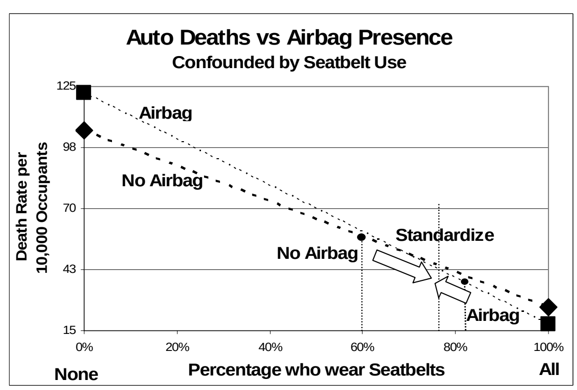

Figure 4: Standardized Auto Death Rates

The standardized death rate of occupants in auto accidents is *slightly higher* for those with airbags than for those without (47 vs. 46). So, do airbags save lives? Not on average for this mix of occupants. This situation is complex. Airbags save lives for those using seatbelts, but take lives for those who do not. The main point is that seatbelts save lives!

Analysis of Confounding

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

Notice what is common to the three examples shown in this essay. In each case, the researcher was passive. Researchers had no power to assign patients to a particular hospital, to determine which families were headed by a married couple or to 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 is a major problem – if not the main problem – in the social sciences. See Lieberson (1985).

Confounding can also arise in any study – observational or experimental – where a related factor is observed at a single level and the choice of level influences the association.

Since most studies in the news are observational, understanding confounding is absolutely necessary to being statistically literate.

To check your understanding of this graphical technique, try working out this sample problem. ☺

Sample Problem: Baseball

Ted and Sam were on the same baseball team. Sam was the better player. If the pitcher was weak, Sam hit 50% of his times at bat while Ted hit 40%. If the pitcher was strong, Sam hit 20% while Ted hit 10%. Since Sam was better than Ted, their Coach had Sam play more against strong pitchers and had Ted play more against weak pitchers.

When facing weak pitchers, Sam had 10 hits in 20 tries while Ted had 32 hits in 80 tries. When facing strong pitchers, Sam had 16 hits in 80 tries while Ted had 2 hits in 20 tries.

- Q1. What is Sam's batting average? What is Ted's?
- Q2. Who has the higher batting average?
- Q3. Does the better player have the higher average?
- Q4. Is this an instance of Simpson's Paradox?
- Q5. What is the lurking variable?
- Q6. What are their standardized averages if each faced weak pitchers 50% of their times at bat?

Work this problem and check your answers with those at the end of this article.

Conclusion

The influence of context on comparisons of ratios is profound. Context is an essential difference between statistics and mathematics. If you really want to understand the influence of context on a statistic or a statistical association, you must understand confounding.

Confounding is typically the reason that "association is not causation" in large observational studies. Confounding – the influence of context – is what links statistics and statistical literacy to the social sciences and the humanities. See Schield 2004b. Confounding and standardization are two of the most important ideas in statistics.

Once you recognize that standardizing (taking into account confounding) can change the size of a comparison and may even reverse the direction (Simpson's Paradox), you will have taken a big step toward being statistically literate. You will have a stronger reason to "Take Care" in using statistical associations as evidence for causal connections.

Acknowledgements

To the W. M. Keck Foundation for their grant "*to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts*" and to Tom Burnham, Cynthia Schield and Marc Isaacson for editorial assistance.

References

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, pp. 129-145 [chapter in 2001 NCTM yearbook].

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 T. Finney (2005). *Who Wants Airbags?* Chance Magazine, Spring 2005, p. 3-16.

Schield, M. (1999). *Simpson's Paradox and Cornfield's Conditions.* 1999 ASA Proceedings of the Section on Statistical Education, p. 106-111.[3]

Schield, M. (2004a). *Three Graphs to Promote Statistical Literacy*. 2004 ICME, Copenhagen.[4]

Schield, M. (2004b). *Statistical Literacy and Liberal Education at Augsburg College*. 2004 Summer Issue of Peer Review, AACU.[5]

Wainer, H. (2002). "*The BK-Plot: Making Simpson's Paradox Clear to the Masses.*" Chance Magazine Vol. 15, No. 3, p. 60–62.

[3] www.StatLit.org/pdf/1998SchieldASA.pdf
[4] www.StatLit.org/pdf/2004SchieldICME.pdf
[5] www.StatLit.org/pdf/2004SchieldAACU.pdf

Answer to Baseball Problem

Q1. What are their batting averages?
A1. Sam had a 260 batting average (26 hits in 100 tries) while Ted had a 340 batting average (34 hits in 100 tries).

Q2. Who has the higher batting average?
A2. Ted

Q3. Does the better player have the higher average?
A3. No

Q4. Is this Simpson's Paradox?
A4. Yes

Q5. What is the lurking variable?
A5. The quality of the pitcher.

Suppose that both players faced weak pitchers 50% of their times at bat.

Q6. What are their standardized averages?
A6. 350 (35%) for Sam; 250 (25%) for Ted.

Conclusion: After taking into account (controlling for) the quality of the pitcher, the batting average is higher for Sam than for Ted (35% vs. 25%).