8/29/06

Constantine, N.A. (forthcoming). Simpson's paradox. In S. Boslaugh (Ed.), *Encyclopedia of Epidemiology*. Thousand Oaks, CA: Sage Publishers.

Simpson's Paradox

Simpson's paradox is an extreme form of confounding, where the association between two variables in a full group is in the opposite direction of the association found within every subcategory of a third variable. This paradox was first described by G. U. Yule in 1903, and later developed and popularized by E. H. Simpson in 1951.

By way of example, consider a new drug treatment that initially appears to be effective, with 54 percent of treated patients recovering, as compared to 46 percent of placebo patients. However, when the sample is divided by gender, it is found that 20 percent of treated males recover compared to 25 percent of placebo males, and 75 percent of treated females recover as compared to 80 percent of placebo females. So the apparent paradox is that the drug is found to be more effective than the placebo in the full group, but less effective than the placebo in each of the two gender-specific subgroups that fully comprise the combined group.

The key to unraveling this puzzle involves the gender confound - differing numbers of patients of each gender receiving the treatment versus placebo, combined with differing overall recovery rates for males versus females. The table below shows that in this example males are 1.6 times more likely to receive the placebo than the treatment, whereas females are 1.6 times more likely to receive the treatment than the placebo. At the same time, females are more than three times as likely to recover as males within both the treatment group and the placebo group. In other words, females are relatively easy to cure. So the fact that the placebo is more effective than the treatment in both groups is obscured when the groups are combined, due to the disproportionate number of easy-to-cure females in the treatment group.

	Treatment	**Placebo**
Male	10/50 (20%)	20/80 (25%)
Female	60/80 (75%)	40/50 (80%)
All patients	70/130 (54%)	60/130 (46%)

Table. Recovery rates

In this particular example, it would be commonly agreed that the correct conclusion involves the subgroup-specific results – the drug is not effective – and that the apparent effectiveness found in the combined group is merely a statistical artifact of the study design due to the gender confound.

Simpson's paradox can be problematic when not recognized, leading to naïve and misleading conclusions regarding effectiveness or other relations studied. Perhaps more ominously,

knowledge of Simpson's paradox can be intentionally used to present or emphasize results that support a desired conclusion, when that conclusion is not valid. More generally, Simpson's paradox has been shown to have implications for the philosophical study of causation and causal inference. In practical terms, it is prudent for both researchers and research consumers to be on guard for this potentially perilous paradox.

Norman A. Constantine
Public Health Institute, and University of California, Berkeley, School of Public Health

See also Confounding, Ecological Fallacy

Further Readings

Reintjes, R., de Boer, A., van Pelt, W., & Mintjes-de Groot, J. (2000). Simpson's paradox: An example from hospital epidemiology. *Epidemiology, 11*, 81-83.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Association, 13,* 238-241.

Wainer, H., & Brown, L. M. (2004). Two statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing data. *American Statistician, 58*, 117-123.

Website

http://plato.stanford.edu/entries/paradox-simpson/