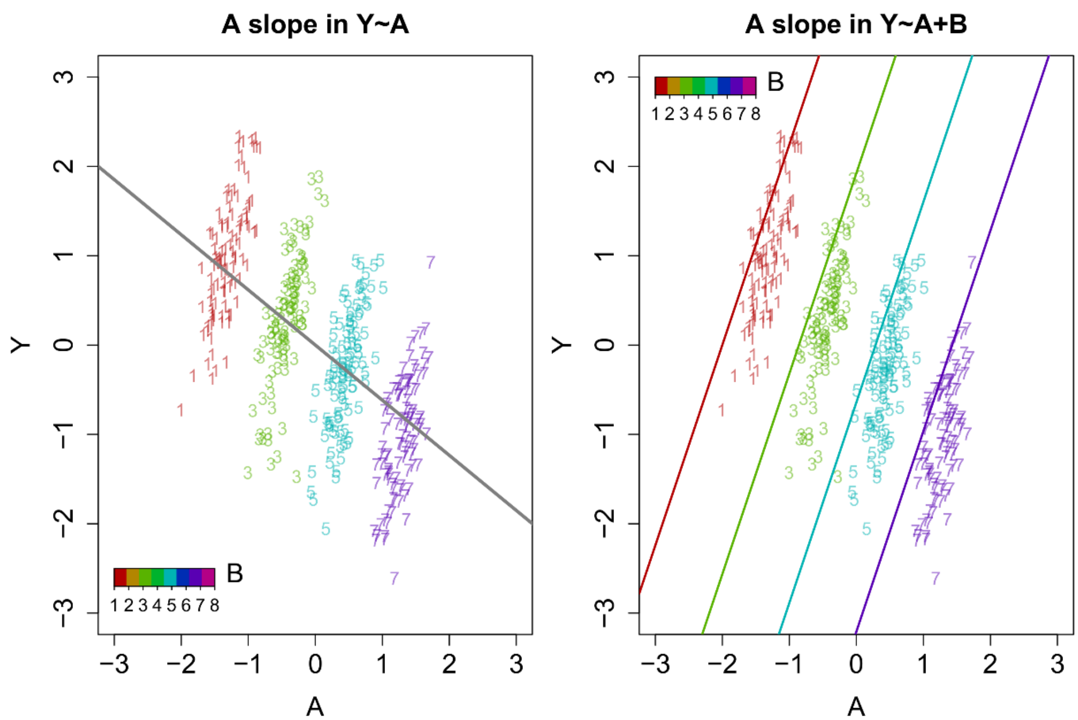
Simpson’s paradox and causality

E-mail distributed on 30-01-2023

Dear all,

I have previously discussed in these mailings Lord's paradox, which occurs when two equally valid models produce apparently contradictory results for the same interventional data. Lord's paradox is a special case of the more widely known **Simpson's paradox**. This paradox occurs when an effect, e.g., Y~A, disappears or even reverses direction when controlled for a covariate, e.g., Y~A+B. While such reversals may be a sign of model instability under multicollinearity, it can also arise under other conditions.

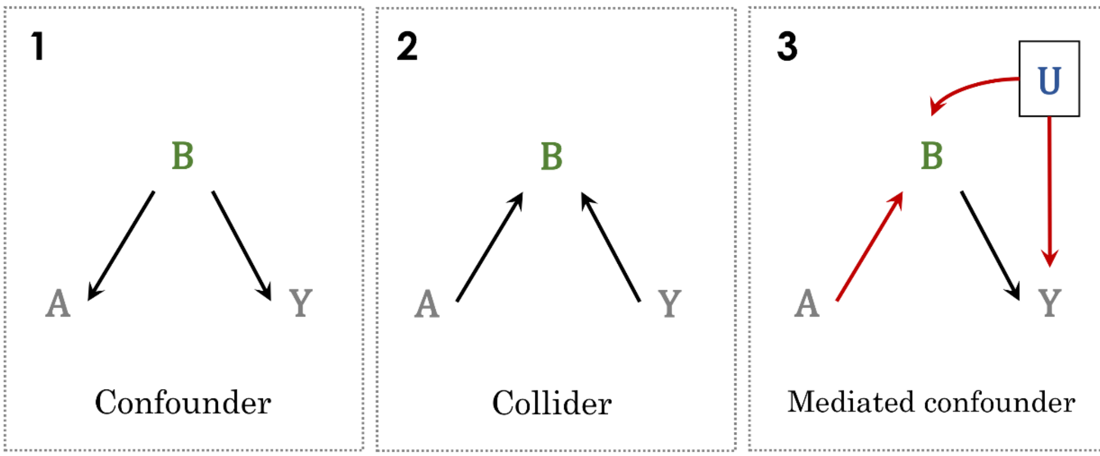


Confounders and colliders

The attached plot shows the example for three continuous variables (A,B,Y). In these data, the marginal AY association (=total effect) is negative, whereas the conditional AY association (=direct effect) is positive. Intuitively, this makes no sense, since if AY is positive for every level of B, we would also expect it to be positive when collapsed across those levels. Moreover, both effects are highly significant, with both $p < 0.0001$!

Firstly, the plot illustrates that the reversal occurs because B is strongly associated with both A and Y. That is, when increasing one level of B, we automatically increase a level of A and decrease a level of Y.

Secondly, the question arises which A effect is correct, Y~A or Y~A+B? In their very lucid paper on the subject, Hernán, Clayton and Keiding (2011) argue that the answer cannot be made on purely statistical grounds, but requires a clarification of the **causal relation** among the variables. That is, if B causes both Y and A (=confounder), then the conditional effect Y~A+B is appropriate. On the other hand, if A and Y both cause B (=collider), then the marginal effect Y~A is appropriate, since adjusting for colliders may introduce selection bias (see Diagrams 1 and 2).



The term *collider* is less familiar to social scientists but was popularized by Pearl (2009), a philosopher of causality who introduced (Bayesian) graphical networks (i.e., so-called directed acyclic graphs) to formalize causal inference. Critically, the graph representation reveals why identical statistical models (Y~A+B) can be correct *and* incorrect depending on the causal assumptions. For the collider example, consider an experiment where A and B were manipulated orthogonally, with both influencing Y. The model B~A+Y would be non-sensical, since it conditions the AB association on their common outcome. AB may in fact be significant when conditioned on identical outcome values, but spuriously so, since AB was orthogonal by design. While this mistake is unlikely to occur in experimental data, in observational data it is frequent enough to be known as Berkson's paradox.

Birth weight paradox

A special historical example of Simpson's paradox is the birth weight paradox, which purported to demonstrate that *smoking* by expecting mothers (A) reduced *infant mortality* (Y) when controlling for *infant birth weight* (B). This subverted the marginal finding that maternal smoking increased infant mortality, as one would expect. In this scenario the causal problem is not apparent on the surface, since birth weight is neither a collider nor a confounder, but a mediator (Diagram 3). Adjustment for B would seem acceptable in the analysis, but overlooks that this may introduce potential confounder variables for the BY association (U). That is, by adjusting for the mediator (A-B-Y), we have inadvertently adjusted for a collider (A-B-U), opening up the confounding path A-B-U-Y!

This version of the paradox highlights not only the importance of clarifying causal relationships in analyses, but also the assumption of **no unmeasured confounders** in causal analyses! Historically, the birth weight paradox was used by tabacco companies to discredit scientific studies showing a causal relationship between smoking and negative health outcomes. As such, it is also a cautionary tale about the pitfalls of inferring causality from observational data, especially when the assumed causal model is overly simplistic.

Finally, note that the classical Simpson's paradox is more often discussed as a problem of categorical association in 2×2×2 tables (e.g., A×Y×B), where the marginal AY odds ratio may contradict the conditional odds ratios AY_{B1} and AY_{B2}. However, for categorical data, this contradiction can be a consequence of **non-collapsibility** as well as confounding. That is, for multiplicative statistics like the odds ratio, the group OR cannot be collapsed to a weighted sum of the subgroup ORs, even when A and Y are independent of B. Hernan et al. (2011) address this distinction in their paper.

Take-aways

- Effect changes/reversals require critical attention, with as first step a check for multicollinearity.
- When the model is otherwise stable, the choice between adjusted and unadjusted effects should be guided by causal assumptions. Mere correlation between A and B does not reveal if B is a confounder, collider, or mediator!
- Adjusting for confounder variables is generally appropriate. Adjusting for collider variables is generally inappropriate.
- For mediators, care should be taken that important sources of confounding for the mediating relationship are measured and controlled for!

Best,
Ben
--
Ben Meuleman, Ph.D.
Statistician
Swiss Center for Affective Sciences
University of Geneva | Campus Biotech
Chemin des Mines 9 | CH-1202 Genève
ben.meuleman@unige.ch | +41 (0)22 379 09 79

References

- Hernán, M.A., Clayton, D., & Keiding, N. (2011). The Simpson's paradox unraveled. *International Journal of Epidemiology*, *40(3)*: 780–785. https://doi.org/10.1093/ije/dyr041.
- Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd edition). Cambridge University Press.
- Pearl, J. (2014). Understanding Simpson's Paradox. *The American Statistician*, *68(1)*: 8–13. doi:10.2139/ssrn.2343788.
- Pearl, J. (2016). Lord's paradox revisited–(oh Lord! Kumbaya!). *Journal of Causal Inference*, *4(2)*: 20160021. https://doi.org/10.1515/jci-2016-0021.