**Path Analysis, Structural Equation Modeling, and Causal Inference**

For a complete and comprehensive guide to path analyses and structural equation models see Bill Shipley’s book, “Cause and Correlation in Biology.”

*Introduction*

How do we obtain causal inference from correlation? Skeptics of path analyses often retort that correlation ≠ causation, and so how can we possibly make inference about causal relationships using correlational data? Shipley evokes an interesting and useful analogy of the Malaysian tradition called *Wayang Kulit*, which is an ancient theatrical art of casting shadows of puppets onto a canvas. Although the audience never sees the puppets themselves (only their shadows), the story is nonetheless conveyed. The correlations among observations and measurements we make of biological systems are like the shadows, and the causal structures are the puppets. The logic follows that if we know what the puppets looks like (the causal relationships), then we know what the correlational relationships should look like (the shadows). In other words, causation structures imply a completely *resolved* correlation structure. The reverse is not true however. Correlation only implies an *unresolved* causation structure. Back to the analogy, a shadowy object with rounded features can be generated from a wide variety of objects, e.g., ball, bowl, the bottom of a cup, etc. This is the foundation of structural equation models. Once a causal structure is proposed, mathematical proofs can be used to derive the correlation structure, and thus these observations and their correlations can be used to test hypotheses about causal relationships among variables. It is very well possible that more than a single causal structure generates the *same* correlational structure, and acknowledging this is part of the SEM inferential process.

Before going any further, it is worth illustrating other approaches for discerning causality in order to highlight where SEMs are particularly useful. When we want to test the relationship between two variables, we intervene by implementing a treatment on the system. For example, if we are interested in the effects of herbivory on plant defenses, we might artificially damage or introduce herbivores to some plants while leaving others undamaged. Moreover, we randomize the plants that receive our treatments to minimize covariation between our treatment and some other variable. For instance, if this is a greenhouse experiment, there will be microhabitat variability within the greenhouse. A fan on one side of the greenhouse might dry nearby plants out more than plants further away. Assigning all the plants near the fan the herbivore treatment would make it impossible to known whether relative humidity or herbivory is responsible for the observed changes in plant defense. Thus, randomization creates causal independence between your treatment and other attributes of the system, allowing you to assess the causal relationship between X and Y. Another way to say this is that we are interested in the effect of X on Y. But how do we know that what we are observing (correlation between X and Y) is due to 1) effect of X on Y vs 2) effect of Y on X vs 3) confounding variable common to X and Y. It can’t be 2 because we implemented the damage. The plants were otherwise healthy and undamaged before we intervened. It shouldn’t be 3 because we decided (randomly) what the treatments were. This leads to the most likely answer being that X is affecting Y. Now, it is possible that we’ve gotten incredibly unlucky and every plant in the herbivory treatment just so happened to be growing under similar microhabitat conditions. As Shipley notes eloquently, “in any empirical investigation, experimental or observational, all we can do is to advance an argument that is beyond reasonable doubt, not a logical certainty.”

Without belittling the power and importance of randomization for causal inferences, it is worth dwelling on a major limitation of randomization. When we randomize and implement a treatment, we can assess the causal relationship between our treatment and the experimental unit. However, in biology (and other fields), these *units* are often exactly that: *units* comprised of complex interrelating parts, the interconnections of which we are particularly interested in quantifying. However, we cannot (usually) manipulate elements of the unit. Shipley gives the example of the hypothetical causal effects of fertilizer addition on crop yield. The prediction is that fertilizer enhances yield through increased nitrogen absorption, which leads to higher photosynthetic rates, which in turn increases yield. Yield is excluded here to simplify the example. However, the experimental units, plants, are comprised of individual components that are not manipulated in this experiment. So, the relationships among components *within the unit* can only be inferred from their correlations. The correlation between N absorption rates and photosynthetic rates could be generated by the following 3 scenarios (non-exhaustive):

Photosynthetic enzymes

N absorption

N fertilizer

Photosynthetic enzymes

N absorption

N fertilizer

N absorption

Photosynthetic enzymes

N fertilizer

Thus, one cannot infer causality for the sorts of relationships we care about from a randomized experiment like the fertilizer example described above. These limitations lend themselves nicely to SEMs, which allow biologists to develop a better understanding of the response of biological units (leaves, plants, communities, etc.) to experimental treatments.

**References (will add more later)**

DAGs—<https://www.dagitty.net/>

[piecewiseSEM in R Gitbook](https://jslefche.github.io/sem_book/index.html) – Jonathan Lefcheck

Bill Shipley’s book, “Cause and Correlation in Biology.”

<https://jonlefcheck.net/2014/07/06/piecewise-structural-equation-modeling-in-ecological-research/>

<https://jebyrnes.github.io/semclass/schedule.html>

<https://ojs.library.queensu.ca/index.php/IEE/article/view/13427>

<https://pubs.usgs.gov/publication/70140273>

<https://oneecosystem.pensoft.net/article/72780/>

<https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.12512>

<https://web.pdx.edu/~newsomj/semclass/ho_improper.pdf>

<https://oneecosystem.pensoft.net/article/50452/>