Directed acyclic graphs

Directed acyclic graphs (DAGs) are diagrams which depict the hypothesized causal relations connecting the primary exposure of interest and the outcome.1 The primary utility of a DAG is to hypothesize the underlying causal mechanism of interest between the exposure and the outcome, identify potential sources of bias in the estimate, and thoughtfully plan the study design and statistical analysis in order to best mitigate these biases. DAGs display non-parametric information through non-looping (acyclic) pathways.

*Terminology*

Figure 1a displays a DAG for the effect of the primary exposure (X) on the outcome under study (Y). Causal effects are denoted by directed paths; i.e. non-crossing and nonrepeating sequences in which all connections flow head to tail. All other paths are non-causal. In Figure 1a, two directed paths lie between X and Y; the path from X to Y and the path from X to another variable (M) to Y. Parents of a variable Y are variables that directly affect Y (X is a parent of Y).

*Mediators, confounders and colliders*

Variables that lie on a directed path are mediators (e.g. M in Figure 1a). Confounders are factors that are causally associated with the exposure and outcome among the unexposed, and are not affected by the exposure or outcome (e.g. Z is a confounder of the X-Y relation in Figure 1a).2 In a DAG, a confounding path is often termed a “back-door path”, and begins with an arrow pointing into the exposure (X) and ends with an arrow into the outcomes (Y). A variable that has two arrows pointing into it is termed a collider (e.g. C is a collider on the path X->C<-Y). Colliders are path specific, and a variable can be both a mediator and a collider depending on which path is examined (e.g. M is a mediator on the path X ->M->Y and a collider on the path X->M->R in Figure 1B).

*Pathways and control of bias*

In order to estimate the causal effect of the exposure on the outcome, we must consider three types of paths in the DAG. Causal paths, where all directions flow from head to tail, are non-biasing paths. Causal paths are by their nature unbiased and require no control (through regression based approaches, restriction or matching) of variables along the path to estimate the effects of the exposure on the outcome through that path. Back-door paths (a non-directed path through a confounder) are biasing pathways, and the primary objective is to close them (either via restriction in the study design or control in the statistical analysis through regression-based approaches of a variable on the back-door path). The third type of path is one through a collider. If not action is taken, these paths are “closed” at the collider, with no association between the parent variables adjacent to the collider through that specific path. However, control of a collider through regression-based approaches or restriction opens the path between the parent variables, inducing a spurious association between these variables. Thus, in Figure 1a, to estimate the causal effect of X on Y, the only biasing pathway requiring control (statistical or other) is the backdoor path from X to Y through Z.

References

1 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37–48.

2 Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology, Third ed.* Wolters Kluwer Health/Lippincott Williams & Wilkins.; 2008.