

# Directed acyclic graphs: A tool to identify confounders in orthodontic research, Part I

Thikriat S. Al-Jewair,<sup>a</sup> Nikolaos Pandis,<sup>b</sup> and Yu-Kang Tu<sup>c</sup>

Buffalo, NY, Bern, Switzerland, Corfu, Greece, and Taipei, Taiwan

Confounding is a bias that threatens the validity of causal inferences in a study. Rothman and Greenland<sup>1</sup> defined confounding: “A confounding factor must not be affected by the exposure or disease. In particular, it cannot be an intermediate step in the causal pathway between the exposure and the disease.”

Controlling for confounders is particularly vital for observational studies, where subjects are not randomly assigned to different groups.<sup>2</sup> A recent cross-sectional study of 4 major orthodontic journals reported that only 17% of the studies adjusted for confounders.<sup>3</sup>

There are 3 principal methods for assessing confounders: traditional approach (also called standard approach), noncollapsibility approach, and causal diagrams through the use of directed acyclic graphs (DAGs).

In the traditional approach, a variable is a confounder if it is associated with the exposure (E), is an independent risk factor to the outcome (D), and is not on the causal pathway between E and D.<sup>4</sup> The problem with this approach is that some variables meet these criteria, but they are not genuine confounders. This could confuse the association between E and D and introduce bias.

The noncollapsibility concept states that a confounder is identified when the crude or unadjusted estimate for the relationship between E and D is different from the estimate calculated after adjusting for potential confounders.<sup>5</sup> This concept can also lead to inappropriate adjustment and invalid inferences, since the adjusted estimate is not always a less biased one than the unadjusted estimate. For instance, the change in estimate could still occur when an

adjustment is implemented for a covariate that is not a confounder because it is on the causal pathway between E and D.<sup>6</sup> This inappropriate adjustment is likely to result in overestimation or underestimation of the true association.

Selection bias is a systematic and nonrandom error that can occur when the methods of subject enrollment into a study or allocations to treatment groups are systematically different among the selected groups. An example of selection bias is when an investigator plans to study retrospectively the effectiveness of functional appliances in the treatment of adolescents with Class II Division 1 malocclusion. The investigator decides to include only successfully treated patients, thus introducing selection bias into the study.

The third approach to control for confounders is the use of causal diagrams or DAGs. DAGs are directed (1 direction), acyclic (no cycles, cannot go back in circle), and graphic (diagrammatic) representations of the causal relationship among variables in a model.<sup>7-9</sup>

Both the traditional and noncollapsibility approaches depend only on statistical associations among variables with no attention to the background knowledge of the study subject.<sup>10</sup> Failure to consider background knowledge can result in either incomplete adjustment for confounding or overadjustment by inclusion of unnecessary variables in the analysis.

Causal diagrams (DAGs) address the above limitations by combining statistical associations (quantitative approach) with the a priori subject-matter knowledge of the causal relationships among E, D, and the confounders (qualitative approach).<sup>10</sup> DAGs provide a quick and simple visual representation of the structural associations at the design stage of a study, aid in determining which variables to adjust for,<sup>11</sup> and distinguish a genuine confounder from nonconfounders.

However, DAGs are still not very suitable to assess effect modifications (also known as statistical interactions among causes) and strengths of associations between variables (correlation coefficients), or for parametric structural modeling.<sup>12</sup>

<sup>a</sup>Department of Orthodontics, School of Dental Medicine, State University of New York, Buffalo, NY.

<sup>b</sup>Associate Editor of Statistics and Research Design; Department of Orthodontics and Dentofacial Orthopedics, School of Dental Medicine/Medical Faculty, University of Bern, Bern, Switzerland; private practice, Corfu, Greece.

<sup>c</sup>Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan.

Am J Orthod Dentofacial Orthop 2017;151:419-22

0889-5406/\$36.00

© 2017 by the American Association of Orthodontists. All rights reserved.

<http://dx.doi.org/10.1016/j.ajodo.2016.11.008>



**Fig 1.** DAG representation of independent relationship between exposure (OT) and outcome (GLR).

### DEFINITION

A DAG is a nonparametric<sup>8</sup> structural method to identify potential confounders through the encoding of assumptions (subject-matter knowledge) about the causal effects in a study and the description of the associations and causations among variables in the form of a graph.<sup>6</sup> DAG is nonparametric, because it does not specify how the relationships among variables should be estimated or the probability distribution of the variables; a DAG presents these relationships in a graphic way instead.

A DAG consists of 3 main elements: nodes (variables), directed arrows (also called edges) connecting the nodes to represent the temporal relationship between the variables (arrows move in 1 direction, from left to right only, since the future cannot cause the past); and the researcher's subject-matter knowledge about the causal relationships among the variables drawn in the DAG.<sup>13</sup>

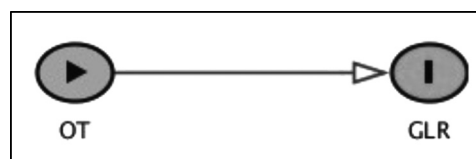
### BASIC DAG TERMINOLOGY

In this example, we want to investigate the effects of full fixed orthodontic appliances (intervention) on the occurrence of labial gingival recession (outcome) in a case-control study. The cases are patients who have labial gingival recession (outcome), and the controls are those who do not have recession (outcome), whereas the exposure is full fixed orthodontic treatment vs no orthodontic treatment.

The first step is to identify the variables of interest. From the above example, we identified 4 binary variables. "OT" denotes the exposure of interest: full fixed orthodontic appliance treatment; "GLR" denotes the outcome: gingival labial recession; "C" is the confounder: which in this study is patient's age; and "S" stands for selection bias: bias during the selection process into the case or control group.

The second step is to draw a DAG to model the causal assumptions presented (DAGitty software<sup>14</sup> was used in this example [<http://www.dagitty.net/>], but other software programs are also available). We now explain the key concepts behind DAGs using our example.

1. In **Figure 1**, the omission of an arrow connecting the exposure "OT" to the disease "GLR" indicates that



**Fig 2.** DAG representation of direct dependence between exposure (OT) and outcome (GLR).

the probability of developing the disease is the same in the treated and untreated patients. In other words, "OT" does not have a causal association with "GLR," and the 2 variables are independent.

- As shown in **Figure 2**, the arrow between "OT" and "GLR" indicates a direct causal association (direct dependence)—ie, the probability of developing "GLR" depends on whether a patient receives "OT."
- A path is the sequence of arrows connecting 2 or more variables regardless of the direction. A directed path is the arrow-based pathway directed, in **Figure 3**, from ancestor to descendent. In this example, it is the path between "OT" (the starting variable) and "GLR" (the ending variable). "OT" is an ancestor of "GLR," and "GLR" is a descendent of "OT."
- Intermediates on the path are variables that intercept the path but are not on the end of it. For example, mandibular incisor proclination "L1" can be a mediating variable between "OT" and "GLR," as in **Figure 4**.
- A *collider* is a variable in which 2 arrowheads converge and appear to collide on the node. It is considered a "common effect" or an outcome of E and D. The collider blocks the path between E and D; hence, they become statistically independent or unassociated with each another. If I select to my study only subjects who develop tooth mobility "S" after "GLR," assuming that "S" can occur as a result of both "OT" and "GLR," "S" is now a common effect and a collider (**Fig 5**).
- A "common cause" is defined as a variable that is associated with both E and D. A "common cause" is what we call a confounder. An association between E and D can simply occur because they both share a "common cause" and not because of a true association between them, hence confounding bias. Thus, it is imperative to control confounding. As shown in **Figure 6**, patient's age, "C," in our example, is the confounder, and it is a "common cause" of E and D.
- A path is open or unblocked if the variable is a noncollider, and closed or blocked if it is a

Download English Version:

<https://daneshyari.com/en/article/5637579>

Download Persian Version:

<https://daneshyari.com/article/5637579>

[Daneshyari.com](https://daneshyari.com)