Causal diagrams for empirical legal research: a methodology for identifying causation, avoiding bias and interpreting results

TYLER J. VANDERWEELE†
Associate Professor, Departments of Epidemiology and Biostatistics,
Harvard School of Public Health, Boston, MA 02115, USA

AND

NANCY STAUDT
Edward G. Lewis Chair in Law, University of Southern California School of Law, Los Angeles, CA 90089, USA

[Received on 13 September 2010; accepted on 19 August 2011]

In this paper, we introduce methodology—causal directed acyclic graphs (DAGs)—that empirical researchers can use to identify causation, avoid bias, and interpret empirical results. This methodology is popular in a number of disciplines, including statistics, biostatistics, epidemiology and computer science, but has not yet appeared in the empirical legal literature. Accordingly, we outline the rules and principles underlying this methodology and then show how it can assist empirical researchers through both hypothetical and real-world examples found in the extant literature. While causal DAGs are not a panacea for all empirical problems, we show that they have potential to make the most basic and fundamental tasks, such as selecting covariate controls, relatively easy and straightforward.

Keywords: Causal diagrams; confounding; direct effects; empirical research; regression; voting rights act.

1. Introduction

Empirical legal scholars spend significant time and energy seeking to identify cause and effect relationships in their data. Even a brief review of the extant legal literature suggests that state-of-the-art statistical analysis is now routine; every stage of empirical research—from data collection to model choice to presentation of findings—is more advanced and sophisticated than was typical in the empirical literature just a decade ago.

Surprisingly, one aspect of causality that researchers have spent very little time exploring is the precise nature of the underlying relationships between and among the variables of interest. Understanding this basic structural framework is essential for a number of empirical tasks, such as specifying sound statistical models, avoiding bias and confounding and accurately interpreting results. In fact, pursuing an empirical project without a map of the cause and effect relations is a bit like undertaking a construction project without a detailed blueprint: success is possible, but the likelihood of confusion and error increases quite a bit absent a good plan.

† Corresponding author. Email: tvanderw@hsph.harvard.edu
In this paper, we show that researchers can use causal directed acyclic graphs (DAGs) as a means to theorize about cause and effect relationships and to identify qualitative assumptions about the data. Pearl (1995) introduced these diagrams into the causal inference literature and showed that they could be useful for reasoning about causal structures and for determining the control variables necessary for an investigator to answer specific causal queries. These diagrams are similar to (and sometimes even used synonymously with) Bayes’ nets or influence diagrams (Edwards, 1991), but causal DAGs specifically allow for causal or counterfactual interpretation, as we clarify below; we will demonstrate how the graphs can assist researchers in the interpretive process. As we discuss, this causal DAG methodology generalizes and formalizes, within a causal context, ideas from the structural equation modeling and path analysis literature that have been popular in the legal and social sciences.\(^1\) While building on existing ideas, the causal DAG methodology also offers innovations and advantages for empirical scholars seeking to make causal claims and for this reason has become popular in statistics, biostatistics, epidemiology and computer science—we argue here that it could be of use in empirical legal research as well.\(^2\)

To give just one example of how causal diagrams can aid empirical researchers, consider a study of judicial behaviour by Cox and Miles (2008) investigating the effects of individual judges’ characteristics on federal judicial decision making in the voting rights context. For their project, the authors collected data on background characteristics (ideology, gender, race, age, education and employment experience prior to the bench), along with case characteristics, and the final judicial decision in the legal controversy. They were particularly interested in the effects of ideology and race\(^3\) on judicial decisions but they also comment on the effects of various other demographic characteristics of the judges. The causal relationships of these variables might be as depicted in Fig. 1 below, suggesting that each of the variables has a direct and unmediated effect on the unit of analysis, the judicial decision, but are not related to each other except as ‘parents’ of the decisions themselves. As will be seen below, this structure would warrant the analytic approach taken by Cox and Miles in the presentation and interpretation of their findings.

Alternatively, the variables in Cox and Miles’ study could be related as depicted in Fig. 2, implying a far more complex (and perhaps more realistic) set of relationships. The unit of analysis is the case and the case characteristics now affect both judicial decisions and the likelihood that litigation will take place in courts with judges of a particular race, gender, age or ideology—i.e. plaintiffs are apt to file claims with judges deemed friendly to their legal claims. The various variables are essentially viewed as characteristics of the case itself. However, causal relationships may exist between and among the judge’s race, gender, ideology, education and employment as thus these are also indicated in the diagram. Gender, race and age have direct effects on the variables education, ideology, employment and judicial decisions, as well as indirect effects on ideology, employment

---

\(^1\) One key distinction between the diagrams used in structural equation modelling and those that we discuss here, e.g. is that in the former context the diagrams depict the variables that ‘will be included’ in a statistical model, whereas in our context, the diagrams help researchers identify ‘which variables to include and exclude’ as controls. Many other differences will become apparent in our discussion below.

\(^2\) Researchers have noted and utilized mapping techniques, some that are akin to DAGs, to understand and explain a range of legal issues, such as forensic evidence in traffic accidents (e.g. Davis, 2003) and legal reasoning and argument (van Gelder, 2007). We have, however, been unable to identify any articles or studies that formally explain the DAG methodology and its usefulness for empirical legal scholars seeking to make causal claims.

\(^3\) We note that some authors (e.g. Holland, 1986; Hernán, 2005) argue that ‘effects of race’ are potentially ill-defined because there is no conceivable hypothetical intervention to change race; though see VanderWeele and Hernán (forthcoming) for further discussion.
and judicial decision as mediated through education. Education has a direct effect on ideology, employment and judicial decisions and an indirect effect on judicial decisions as mediated through ideology and employment. Indeed, among these variables, only ideology and employment have a direct unmediated effect on judicial decisions in both Figs 1 and 2.

In sum, we suggest that both case and individual judge characteristics directly affect the Cox and Miles outcome unit of analysis, i.e. the judicial decision. Further, case characteristics directly affect only litigation-relevant judge characteristics along with the final judicial decision. Several of
the background judge characteristics do not affect each other because they are unchanging, such as race and gender.

Deciding which diagram better describes the structural relationships between these variables requires that investigators rely on theory and qualitative assumptions, and the decision is key for purposes of making good modelling decisions and interpreting empirical results correctly. If, on the one hand, Fig. 1 reflects the true structure of the data, then, as will be made clear below, the total and direct effects on judicial decisions are equal for each of the seven variables, and the choice of statistical controls for assessing these effects is relevant only for purposes of precision (that is to say, the controls are useful for decreasing the standard error of the estimate). If Fig. 2 is the true model, on the other hand, then the estimates of the total and direct effects of race, gender or age will in general diverge; indeed, they could have different signs—the direct effect could be negative, while the overall total effect could be positive. Parsing total effects and direct effects of, say, race, in the context of Fig. 2 would require the authors to adjust for education, employment, ideology, gender, age and case characteristics for the direct effect, but only case characteristics for the total effect. Moreover, and perhaps more importantly, the estimated total effect of ideology, education and employment are likely to be biased if Fig. 2 is correct and the authors fail to control for gender, race and age for the reasons just described. After our exposition of the theory of causal DAGs, we will return to these and other issues and show how the assumptions made by empirical scholars could have notable effects on results.4

We would like to point out an important fact that is often overlooked by researchers before continuing: interpreting regression coefficients is never free of assumptions vis-à-vis the structural relationships between and among variables. Whenever a coefficient is interpreted causally, assumptions are necessarily being made about how the variables in the regression are related; causal diagrams, as we will show below, are useful because they make explicit what scholars often assume implicitly and sometimes inadvertently. Our goal in this article is to demonstrate how causal diagrams can aid empirical scholars in clarifying their theory about the data and in the subsequent task associated with interpreting empirical results. This will increase the likelihood that investigators will make sound qualitative assumptions, pursue the best modelling strategy for identifying causal effects and interpret results in a useful and precise manner.5

Before outlining the ways in which causal diagrams can advance the goals and aims of empirical legal research, we first describe the formal rules and principles for constructing and reasoning about these diagrams. Section 2 describes the mechanics of constructing causal graphs and shows why the graphs are useful for clarifying qualitative modelling assumptions, a necessary step for addressing

---

4 It might be argued that to estimate a target parameter, investigators should simply adjust for all available variables that affect the cause and outcome of interest and that potentially mediate the effects in order to assure precise estimates, avoid bias and allow for the identification of direct effects. But this common approach, as noted above, does not allow the researcher to distinguish between total and direct effects, a difference that is often of great importance in empirical research. Furthermore, as will be seen below, such an all-inclusive approach to modelling has the potential to induce bias and confounding.

5 The causal structures like those presented in Figs 1 and 2 are not only important for empirical studies in the academic context but also have practical use. To see this, suppose an investigator seeks to identify the effects of race not on judicial outcomes, but on employment decisions for purposes of identifying race discrimination. In this context, the total effect of race is not the target; rather, the interest lies in indentifying if and how race directly influenced the hiring decision irrespective of the applicants’ education. In the words of one prominent judge, ‘The central question in any employment-discrimination case is whether the employer would have taken the same action had the employee been of a different race (age, sex, religion, national origin, etc.) and everything else had remained the same’. This standard requires the researcher to control for education (to identify the direct effects) and for gender (to avoid confounding) if Fig. 2 is the true model. However, no adjustments are necessary if Fig. 1 reflects the underlying structure of the data.
problems of confounding and selection bias. Section 3 explains how and why the graphs enable researchers to determine what independence and conditional independence relations hold among variables. Section 4 turns directly to the issues of confounding and selection and illustrates how investigators can use causal graphs to examine problems of confounding and selection bias and to estimate the causal effects of an intervention, treatment or policy change from nonexperimental data. Section 5 discusses the way in which DAGs can address the issues associated with controlled directed effects.

In Section 6, we return to the extant legal literature. Section 6.1 demonstrates how to apply the causal DAG framework to real-world research with the help of a case study that focuses on the Cox and Miles (2008) investigation outlined above. Section 6.2, thanks to Cox and Miles’ generous decision to share their data set, refits the data to new statistical models that account for the bias and confounding that we believe exists in their study. Through this re-estimation process, we demonstrate the ways in which causal diagrams could have improved and made more precise the empirical results reported by Cox and Miles; specifically, we note that the authors systematically report both inflated and deflated coefficients in their study of judicial decisions. The authors’ claim with respect to ideology is perhaps overstated, but their causal claims with respect to race, age and employment are quite a bit stronger than their empirical results suggest. Section 7 offers concluding remarks and notes the variety of empirical legal studies that could benefit from the methodology that we present here.

Finally, we would like to note that our presentation here is intended to be an introduction to the theory and uses of causal diagrams. We seek to demonstrate how legal scholars are able to use the graphs for the most basic questions that emerge in all empirical studies. There are many other empirical issues that can be addressed through the use of causal graphs that we do not discuss here, such as issues involving mutual causation, interaction and qualitative information about signs. We encourage scholars to consult the broad, emerging literature in statistics, biostatistics, epidemiology and computer science for more-in-depth discussion of these and other issues.

2. Causal theory and DAGs

The first step in creating a causal DAG requires the construction of a network or diagram representing the investigator’s understanding of the relationships and dependencies between and among variables. The graph consists of a set of nodes (the variables) and a set of directed edges (arrows) that link the nodes. A path is an unbroken nonintersecting sequence of edges that may go along with or against the arrows. A directed path is a path that follows the edges in the direction of the arrows. Relationships such as \( A \leftarrow B \rightarrow C \) and \( A \rightarrow B \rightarrow C \) are both paths but only the latter is ‘directed’ path as it follows the edges in the direction indicated by the graph’s arrow. A node \( X_i \) that has a directed edge into node \( X_j \) indicates the former is the ‘parent’ (or ‘direct cause’) of the latter; in this case, \( X_j \) is said to be a ‘child’ of \( X_i \). A node \( X_i \) is an ‘ancestor’ (or ‘indirect cause’) of \( X_j \) if there is a directed path from \( X_i \) to \( X_j \); in this case, \( X_j \) is said to be a ‘descendant’ of \( X_i \). If there is no node on the graph that has a directed path back to itself, then the graph is said to be acyclic. In this essay, we consider graphs that are both directed and acyclic and for this reason are called DAGs (Greenland et al., 1999; Pearl, 2009).

The directed graphs that are acyclic preserve the notion that causes must precede their effects. Systems exhibiting mutual causation can be handled on a causal DAG by representing the same variable at different times by different nodes. Representing systems in which causation among variable is mutual and simultaneous is more difficult but some progress can be made even in these settings (White and Chalak, 2009).
A DAG is said to be a *causal DAG* if two conditions are met. First, the arrows on the graph must have a causal interpretation in the sense that interventions on a parent node should affect the values of the child node. Second, for a graph to be a *causal DAG*, it is necessary that any common cause of two variables on the graph is also on the graph. If there is a variable that is only a cause of one other variable on the graph, it may be included or omitted from the graph; however, if the variable is a cause of two or more other variables on the graph, then it must be included. As will be seen below the requirement that any common cause of two variables on the graph must be on the graph is important for reasoning about confounding and causal relationships.

Figure 3 presents a modified example of a causal graph from Judea Pearl’s 2009 book, *Causality: Models, Reasoning, and Inference* (Pearl, 2009, p. 15). The graph indicates the relationship among and between five separate variables. Assume, for purposes of discussion, that it represents the relations among the seasons of the year ($X_1$), sprinkler systems ($X_2$), rainfall ($X_3$), wet pavement ($X_4$) and accidents ($Y$). The DAG shows that $X_1$ is a parent of both $X_2$ and $X_3$; that $X_2$ and $X_3$ are parents of $X_4$; and that $X_4$ is a parent of $Y$. In the terminology given above, we could also describe $X_1$, $X_2$, $X_3$ and $X_4$ as ancestors of $Y$; we could describe $X_2$, $X_3$, $X_4$ and $Y$ as descendants of $X_1$. All these concepts will become useful in our discussion below. Moreover, it is easy to see that $X_2 \leftarrow X_1 \rightarrow X_3$ and $X_1 \rightarrow X_2 \rightarrow X_4 \rightarrow Y$ are both paths but only the latter is a directed path.

The graph reflects our intuitions, understanding and beliefs about the world and it is meant to convey underlying assumptions of analysis. The absence of a direct link between $X_1$ and $Y$, e.g. captures our understanding that the influences of seasonal variation on sidewalk accidents is mediated through various other conditions and are not direct causes of accidents. Springtime, e.g. does not directly cause one to slip on the sidewalk; rather, springtime leads to more rain or higher levels of sprinkler use, which, in turn, leads to wet pavement, which is the direct and proximate cause of observed accidents in this model. The intuitions represented in the graph also coincide with independence conditions. As will be seen below, the graph implies that knowing $X_4$ renders $Y$ independent of the set {$X_1$, $X_2$, $X_3$}. If further thought and consideration lead us to believe that the variables are related in a different manner, then the graph must be updated to reflect this new understanding. We show how slight modifications in the graph reflect different assumptions in our discussion below.

More extended discussion of causal DAGs can be found in the work of Pearl (1995, 2009) and Greenland et al. (1999). In the online appendix, we discuss some more technical material on how causal DAG are related to structural equation modelling. Importantly, these causal DAGs can generalize structural equation modelling ideas in the legal and social sciences without imposing assumptions about linearity or normality; causal DAGs in fact do not impose any assumptions concerning functional form or distributions (see the online appendix for further details). This considerable increased generality, however, comes at a price. Unlike traditional structural equation modelling.

![Figure 3. A directed acyclic graph with five variables (Pearl, 2009, p. 15).](https://academic.oup.com/lpr/article-abstract/10/4/329/962265)
techniques, the use of causal DAGs generally does not entail estimating a path coefficient. Because causal DAGs do not make assumptions about functional form, it will often be impossible to characterize the relationships between variables as a single path coefficient. Causal DAGs rather are conceptual tools, which allow the researcher to draw conclusions about confounding and to clarify structural assumptions. The conclusions drawn will apply irrespective of the functional form relating the variables. However, to conduct empirical analyses, the use of causal DAGs will be supplemented by regression analyses and thus also with additional assumptions about functional form.

3. Causal diagrams and independence relations

It is easy to see that causal DAGs provide an intuitive and straightforward means for expressing qualitative assumptions about the relationships of the variables of interest. The graphs, however, can also highlight hidden complexities about these relationships that researchers ignore in the absence of the DAG. Consider the variable $X_4$ in Fig. 3 and note that $X_2$ and $X_3$ both have directed edges into $X_4$. Because $X_4$ is the effect of two separate causes, we will say that $X_4$ is a ‘collider variable’ on the path $X_2 \rightarrow X_4 \leftarrow X_3$. Colliders take on special importance because they reflect the fact that two parents can be marginally independent but can become dependent if we condition on their common effect.

It may seem surprising and perhaps counter-intuitive that conditioning on a node could actually create dependence rather than block it. Observations of the common consequences of two independent variables tend to render those causes dependent because information about one of the causes tends to make the other more or less likely given the consequences that have occurred (Pearl, 2009; Morgan and Winship, 2007). Stephen Morgan and Christopher Winship provide a useful example to illustrate the point. Consider a team of researchers who plan to study the applicant pool of a particular university. The admission criteria at the university calls for either a high SAT score or a high level of extracurricular activities. Even if these two factors are independent in the population of interest, they will likely be negatively correlated in the population of admitted students (because e.g. if a student is admitted but does not have a high level of extracurricular activities, then she must have had a high SAT score). Thus, if the researchers decide to adjust for the admission decision in their study, they would uncover a strong correlation between SAT scores and extracurricular activities when in fact no such correlation exists in the applicant pool generally (Morgan and Winship, 2007, pp. 66–67). This pattern is sometimes referred to as selection bias in the legal and social science literature and can be induced by adjustment for collider variables in some contexts—a problem we discuss further below.

Formally, a collider is defined as a node on a particular path such that both the preceding and the subsequent nodes on the path have directed edges going into that node. Note that a collider is specific to a path. Thus, $X_4$ is a collider on the path $X_2 \rightarrow X_4 \leftarrow X_3$ but $X_4$ is not a collider on the path $X_2 \rightarrow X_4 \rightarrow Y$. A path between two nodes, $A$ and $B$, is said to be blocked conditional on some set of variables $Z$ if either there is a variable in $Z$ on the path that is not a collider or if there is a collider on the path such that neither the collider itself nor any of its descendants are in $Z$. If all paths between $A$ and $B$ are blocked given $Z$, then it can be shown that $A$ and $B$ are independent conditional on $Z$. Thus, in Fig. 3, $X_2$ and $X_3$ will be conditionally independent given $X_1$ since the path $X_2 \leftarrow X_1 \rightarrow X_3$ is blocked by $X_1$, which is in the conditioning set and the path $X_2 \rightarrow X_4 \leftarrow X_3$ is blocked by $X_4$, which is a collider on the path. Thus, all paths between $X_2$ and $X_3$ are blocked conditional on $X_1$ and so $X_2$ and $X_3$ are conditionally independent given...
However, using these rules of conditional independence, we can see that $X_2$ and $X_3$ will not be conditionally independent given $X_1$ and $X_4$. When we condition on $X_4$, the path $X_2 \rightarrow X_4 \leftarrow X_3$ is no longer blocked conditional on $X_1$ and $X_4$ because we are now conditioning on the collider $X_4$. These rules concerning blocked paths allow us to assess any independence relation on the graph. Essentially, statistical association between two variables $A$ and $B$ can arise in one of three ways. First, $A$ might be a cause of $B$ (or $B$ a cause of $A$) and this creates association between the two variables. Second, $A$ and $B$ might be statistically associated because of some common cause $C$ of both $A$ and $B$, even if neither $A$ nor $B$ is a cause of the other. Third, $A$ and $B$ might be associated if we condition on a common effect of $A$ and $B$ (or, more generally, if we condition on a common effect of two variables one of which is associated with $A$ and one of which is associated with $B$).

The next section describes the relationship between causal DAGs and the estimation of causal effects, shows how it is possible to use the graph to identify causal effects and relates this discussion to methods and criteria for addressing problems of confounding and selection biases.

4. Using DAGs for the identification of causal effects

Suppose we have observational data and that we have constructed a causal diagram representing the relationships and dependencies of the variables of interest as discussed above. We want to estimate the causal effect of a specific intervention, policy program or medical procedure (set $X = x$) on an outcome of interest ($Y$). The question arises as to what adjustments must be made to avoid confounding. Adjustments are essentially equivalent to dividing the population into groups that are homogenous relative to some factor, say $Z$, and assessing the effect of the intervention on the outcome in each homogenous group and then averaging the results (Pearl, 2009, p. 78). Such a procedure is often carried out in conjunction with modelling by means of regression techniques.

Confounding occurs when an unblocked ‘backdoor’ path exists from the treatment or intervention to the outcome, thereby allowing additional factors, other than treatment, to affect the dependent variable of interest. A backdoor path from $X$ to $Y$ is a path that begins with an edge going into $X$. The effect of $X$ on $Y$ is unconfounded when all backdoor paths between the treatment and the outcome are blocked by the set of pretreatment variables for which adjustment is made. More formally, let $Y_x$ denote the value of $Y$, which would be obtained, if possibly contrary to fact, there were an intervention to set $X$ to the value $x$. Variables of the form $Y_x$ are sometimes referred to as counterfactual variables or potential outcomes (Rubin, 1974). We cannot in general draw inferences about counterfactual variables for a particular individual but under certain assumptions, we can draw inferences about the probability distributions or expectations over a population of counterfactual variables. We say that the effect of $X$ on $Y$ is unconfounded given $Z$ if $P(Y_x|Z = z) = P(Y|Z = z, X = x)$. The quantity $P(Y_x|Z = z)$ is not generally empirically observable from observational data since we do not in general know what would happen under interventions to set $X$ to $x$. In contrast, the quantity $P(Y|Z = z, X = x)$ is empirically observable. When the effect of $X$ on $Y$ is unconfounded given $Z$, we can estimate counterfactual probabilities $P(Y_x|Z = z)$ using observed probabilities $P(Y|Z = z, X = x)$. The condition referred to as unconfoundedness is sometimes also referred to as ‘conditionally ignorable treatment assignment’ in the literature on causation (Rubin, 1990; Greenland et al., 1999).
These concepts become transparent and easy to understand with the help of a causal DAG and Judea Pearl’s ‘backdoor criterion’, a simple graphical test that researchers can use to determine whether the effect of \( X \) on \( Y \) is confounded. We first set out a general version of the result and then provide some specific examples of its application. A set of variables, \( Z \), satisfies the backdoor criterion relative to the ordered pair \( X \) and \( Y \) in a DAG if

(i) no node in \( Z \) is a descendant of \( X \); and

(ii) \( Z \) blocks every path between \( X \) and \( Y \) that begins with an arrow into \( X \) (i.e. all backdoor paths from \( X \) to \( Y \)).

If the set of variables \( Z \) satisfies this criterion, then the causal effect of \( X \) on \( Y \) is identified and given by the formula

\[
P(Y_x) = \sum_z P(y|x, z) P(z).
\]

To understand how the backdoor criterion operates, reconsider Fig. 3 and assume, we seek to identify the effects of wet pavement (\( X_4 \)) on the likelihood of accidents (\( Y \)). Because all paths to \( Y \) must enter through \( X_4 \), we easily can see that there is no backdoor path from \( X_4 \) to \( Y \) and thus, it is possible to identify the causal effects of the treatment without controlling for any variables whatsoever. Thus, we do not need to make any adjustments to the model, and the expected value of the causal effect of the treatment on the outcome of interest is equal to \( E(Y_{x4}) = E(Y|X_4 = x_4) \). The total effect of the wet pavement on accidents, comparing wet pavement to dry pavement, would then be given by

\[
E(Y_{x4=1}) - E(Y_{x4=0}) = E(Y|X_4 = 1) - E(Y|X_4 = 0).
\]

Now assume that the DAG we have constructed for identifying the effects of wet pavement on accidents is presented in Fig. 4.

Note: Controlling for \( X_2 \) satisfies the backdoor criterion for the effect of \( X_2 \) on \( Y \).

Suppose that sprinkler systems often jut out of the ground in a manner that makes accidents happen irrespective of whether the pavement is wet and that \( U_1 \) indicates whether the sprinkler system is jutting out of the ground. Suppose we only have data on \( X_1, X_2, X_3, X_4 \) and \( Y \) but not \( U_1 \). Note that if Fig. 4 is in fact a correct representation of the causal relationships, then Fig. 3 is not a causal DAG because not all the effect of \( X_2 \) on \( Y \) is mediated by \( X_4 \). In this case, Fig. 3 could be made into a causal DAG by adding an arrow from \( X_2 \) directly to \( Y \); then both Figs 3 and 4 would be causal DAGs but Fig. 4 would simply be a more elaborate causal DAG. In any case, if Fig. 4 indicates a correct depiction of the causal relationships, we have possible confounding bias because we have two backdoor paths from \( X_4 \) to \( Y \). Namely, we have \( X_4 \leftarrow X_2 \rightarrow U_1 \rightarrow Y \) and \( X_4 \leftarrow X_3 \leftarrow X_1 \rightarrow X_2 \rightarrow U_1 \rightarrow Y \). The backdoor criterion, however, indicates that it is nonetheless possible to identify the causal effects of \( U_4 \) on \( Y \) if we adjust for \( X_2 \). Using the criteria set out above, \( X_2 \) is not a descendant of \( X_4 \) and it blocks all backdoor paths from \( X_4 \) to \( Y \). Thus, adjusting

\[
\text{FIG. 4. Causal directed acyclic graph with unobserved variable, } U_1.
\]
for $X_2$ in our model means that the expected value of the effect of wet pavement on accidents is 
\[ E(Y_{X=1} - Y_{X=0}) = \sum_{x_2} E[Y | X_4 = 1, X_2 = x_2] P(X_2 = x_2) \]
\[ - \sum_{x_2} E[Y | X_4 = 0, X_2 = x_2] P(X_2 = x_2). \]

What if $X_2$ itself was unobserved—would the analysis change? If $X_2$ were unobserved, we could not satisfy the backdoor criterion using the variables for which we had data and thus the causal effects of wet pavement on the likelihood of accidents would be confounded. This is because we would not be able to find a set of $Z$ variables that included nondescendants of $X_4$ and that would block all backdoor paths from $X_4$ to $Y$. Even if we adjusted for $X_1$ and $X_3$, we would still have the unblocked path $X_4 \leftarrow X_2 \rightarrow U_1 \rightarrow Y$ and thus, our estimate of the effect of $X_4$ on $Y$ would be confounded by the uncontrolled for effects of $X_2$ and $U_1$.

One of the important and interesting features of the backdoor criterion is that it may lead to different modelling approaches than are commonly adopted by empirical researchers. One familiar approach for addressing possible confounding, e.g. is to control for any and all pretreatment variables. This approach has the perceived advantage of assuring that the investigator will adjust for all possible confounders and in the worse case scenario will not affect the results if no confounding in fact exists. But there are also disadvantages to this approach. Adjusting for numerous and possible unnecessary variables requires far more information and thus may be costly; it may pose problems if the sample size is limited; and if the control variables are correlated only with treatment (and not the outcome), the estimates will be less efficient and thus statistical significance could be lost. Moreover, as we will see below, it can sometime even introduce bias! For all these reasons, parsimony should be preferred in modelling—and the backdoor criterion facilitates this goal.

To see how the common approach of adjusting for all pretreatment covariates differs from the approach that relies on the backdoor criterion, reconsider Figs 3 and 4. Using the former strategy, a researcher would control for $X_1$, $X_2$ and $X_3$ in order to identify the effects of $X_4$ on $Y$; the backdoor criterion, however, calls for no adjustments in Fig. 3 and for adjusting only for $X_2$ in Fig. 4. This difference is important because more complex DAGs would lead to far more adjustments under the common approach but not necessarily under the backdoor criterion.

Not only will the backdoor criterion often lead to fewer adjustments, the method enables researchers to avoid confounding when it would be induced by the rule of thumb that calls for controlling for all pretreatment variables. Consider Fig. 5, where the variables $U_1$ and $U_2$ indicate unobservable and unmeasurable variables. If the researcher controls for $X_3$, a pretreatment variable, she will unwittingly open a backdoor path from $X_4$ to $Y$ (namely, $X_4 \leftarrow U_2 \rightarrow X_3 \leftarrow U_1 \rightarrow Y$) that

![Fig. 5. Directed acyclic graph with two unobserved covariates.](https://academic.oup.com/lpr/article-abstract/10/4/329/962265)
cannot be blocked due to the existence of unobservable variables. This problem emerges because is a collider variable on this path (i.e. an effect of two different causes). As noted above, when we adjust for a collider variable, we create an association with two otherwise independent variables, which in this context creates confounding. In this example there is no confounding for the effect of $X_4$ on $Y$ without adjusting for $X_3$, but there is confounding when adjustment is made for $X_3$. Thus, if the researcher constructed a DAG similar in substance to Fig. 5 with a collider variable present and adopted the common approach of controlling for all pretreatment variables—she would create confounding when it could be avoided if she relied on the backdoor criterion.

We are now in a position to set out a more conceptually detailed description of the backdoor criterion that will assist researchers in the decision of when to control and when not to control for pretreatment variables. For purposes of this discussion, we label a directed path such as $A \rightarrow B \rightarrow C$ as a chain, a nondirected path such as $A \leftarrow B \rightarrow C$ as a fork, and a collider variable as one that is the effect of two separate causes, such as $B$ in the path $A \rightarrow B \leftarrow C$. Recall we seek a set of variables, $Z$, that satisfies the backdoor criterion relative to the ordered pair $X$ and $Y$ in a DAG. According to Pearl’s test, we can say that the backdoor criterion is satisfied if $Z$ contains nondescendents of $X$ and if every backdoor path from $X$ to $Y$ contains

(i) a chain, $A \rightarrow B \rightarrow C$, where the middle node $B$ is in $Z$, or
(ii) a fork, $A \leftarrow B \rightarrow C$, where the middle node $B$ is in $Z$, or
(iii) a collider variable $B$, $A \rightarrow B \leftarrow C$, such that the middle node $B$ is not in $Z$ and such that no descendant of the collider is in $Z$.

With these criteria, investigators are able to conceptualize the problem of confounding in a clear and unambiguous manner. Moreover, as we show in our discussion below, the criterion allows for a systematic procedure that is applicable to diagrams of any shape, size or complexity. Finally, the backdoor criterion enables the analyst to search for the optimal and ‘minimal’ set of covariates (see Pearl, 2009, p. 80 for further discussion).

5. Controlled directed effects

One further issue concerning the identification of causal effects may be of interest. Consider again the causal DAG in Fig. 4. Suppose we are interested in the effect of sprinkler systems ($X_2$) on accidents ($Y$). Part of this effect may be mediated because the sprinkler systems make the pavement wet; this part of the effect is said to be mediated through $X_4$. Part of the effect of the sprinkler system on accidents may occur directly because the systems jut out of the ground in a manner that make accidents more likely. We may be interested principally in the direct effect of the sprinkler system on accidents controlling for wet pavement.

More generally, if some variable $X$ is a cause of some outcome $Y$ and if $M$ is a variable on a directed path from $X$ to $Y$, we may be interested in the direct effect of $X$ on $Y$ controlling for $M$. We let $Y_{x,m}$ denote the value of $Y$, which would be obtained, if possibly contrary to fact, there were interventions to set $X$ to the value $x$ and $M$ to the value $m$. The controlled direct effect is then defined as $Y_{x=1,m} - Y_{x=0,m}$. Note that this controlled effect may be different for different values of $m$. Thus, in the context of the sprinkler example, we might be interested in the direct effect of the sprinkler systems on accidents intervening to make the pavement wet; this would be denoted by $Y_{x=1,x=1} - Y_{x=0,x=1}$ and this quantity captures the ‘jutting out’ effects of sprinkler systems being on when the pavement is wet. Note that here the mediator $M$ that we are controlling for is the variable...
indicating whether the pavement is wet. Alternatively, we might be interested in the direct effect of the sprinkler systems on accidents intervening to make the pavement dry; this would be denoted by $Y_{x_2=1, x_4=0} - Y_{x_2=0, x_4=0}$ and would capture the 'jutting out' effects of sprinkler systems being on when the pavement is dry.

In general, controlling for a posttreatment variable can introduce bias (Rosenbaum, 1984; Pearl, 2009) in the analysis of total effects. However, under certain circumstances, the results of such analyses can be interpreted as direct effects (Pearl, 2009). Causal diagrams clarify the assumptions required for such a causal interpretation. We can draw conclusions about the identification of controlled direct effects using a generalization of the backdoor path criterion. Specifically, if there is some set of variables $Z$, which are not descendents of either $X$ or $M$ and if (i) all backdoor paths from $X$ to $Y$ that do not pass through $M$ are blocked by $Z$ and (ii) all backdoor paths from $M$ to $Y$ are blocked by $Z$ and $X$ then inferences can be drawn about the probability distribution of counterfactuals of the form $Y_{x,m}$. Basically, these conditions amount to $Z$ sufficing to block all the backdoor paths from $(X, M)$ jointly to $Y$. Specifically, if the conditions above hold then

$$P(Y_{x,m}) = \sum_z P(y|x, m, z)P(z).$$

More general conditions for identification are also available (Pearl, 2009), but the conditions given above will suffice for the purposes of our discussion. In the context of the sprinkler example, we see that if we want the direct effect of sprinkler systems ($X_2$) on accidents ($Y$) controlling for wet pavement ($X_4$), we do not need to adjust for anything. There are no backdoor paths from $X_2$ to $Y$ that do not pass through $M (=X_4)$ and thus, the first condition above is satisfied. Furthermore, $X_2$ blocks all backdoor paths from $X_4$ to $Y$. We could then simply regress $Y$ on $X_2$ and $X_4$ and the coefficient for $X_2$ would represent the direct effect. Estimating the total effects of the sprinkler system on accidents would, in contrast, require the investigator to control only for $X_1$ or $X_3$ to avoid confounding.

It is important to note that to identify controlled direct effects, there are two backdoor path conditions given above that must be satisfied, not just one. In addition to the condition that the backdoor paths from $X$ to $Y$ (not through $M$) be blocked by $Z$, the backdoor paths from $M$ to $Y$ must be blocked by $(X, Z)$. Both conditions are needed if we want to estimate controlled direct effects. That is to say, we must not just adjust for variables that confound the relationship between intervention $X$ and outcome $Y$ but also for those variables that confound the relationship between the mediator $M$ and the outcome $Y$. If we do not control for variables that confound the relationship between the mediator $M$ and the outcome $Y$ (i.e. if the second condition is not satisfied), then our estimate of controlled direct effects will be biased (Judd and Kenny, 1981; Robins and Greenland, 1992; Cole and Hernández, 2002). Other more subtle definitions of direct and indirect effects, which allow for the partitioning of a total effect into a direct and indirect effect are also available (Pearl, 2009).

6. Improving empirical research with DAG methodology

We now turn from abstract rules and hypothetical examples to Cox and Miles’ study, ‘Judging the Voting Rights Act’, with which we began our discussion. Before we begin our investigation, it is useful to note that Section 2 of the Voting Rights Act of 1965 prohibits the use of discriminatory voting practices or procedures. Over the years, this provision of the law has become an important
litigation tool for individuals and groups alleging a denial or abridgement of their voting rights. In terms of numbers, federal courts issued more than 750 decisions between 1982 and 2004 in cases raising Section 2 claims (Cox and Miles, 2008). As noted above, Cox and Miles sought to identify whether personal and background characteristics play a role in a judge’s decision to hold a state official liable (or not) for violating the mandates found in Section 2 of Voting Rights Act.

In this section, we use the causal DAG methodology to illustrate the assumptions that authors often unknowingly make in their empirical work and, at the same time, demonstrate just how causal DAGs can aid researchers in basic empirical tasks. More specifically, in Section 6.1, we note that that Cox and Mile’s regression analyses depend on the accuracy of Fig. 1 (presented in Section 1); the authors’ estimates, however, are confounded if Fig. 2 is a better depiction of the data. In Section 6.2, we refit the data to the models assuming the accuracy of Fig. 2 and highlight the qualitative and quantitative changes that emerge between our findings and those presented by Cox and Miles in their study.

6.1 Cox and miles’ causal assumptions and the potential for confounding

In Section 1, we presented, in Figs 1 and 2, DAGs depicting possible cause and effect relationships between and among the variables of interest. If Fig. 1 is accurate, the authors need not worry about possible confounding. Nor is there any distinction to be made between total and direct effects because these two effects are identical in Fig. 1 for all variables. One could regress judicial decisions on each of the variables one by one to obtain total effects. Alternatively, if Fig. 1 is correct, then one could also regress judicial decisions on all seven variables simultaneously and use this regression to also obtain the causal effects. The difference between the two is that the latter approach will produce more precise estimates.

If Fig. 2 is accurate, confounding is a potential problem for purposes of estimating the effects of ideology, education and employment. This is because open backdoor paths exist from these variables to the outcome of interest. For ideology, there are backdoor paths to judicial decisions including ‘ideology ← gender → judicial decisions’ and ‘ideology ← race → judicial decisions’ and ‘ideology ← age → judicial decisions’; similar backdoor paths exist for employment and education. In Fig. 2, gender, race and age are confounded only by case characteristics. Thus, if control is made only for the characteristics of the case, the total effect of gender, race or age could be estimated. Moreover, in Fig. 2, ideology and employment are the only variables with causal effects that are not mediated by intervening variables and thus, the total and direct effects are identical for these nodes, but not for any other node in the figure.

Cox and Miles present their multivariate regression analysis for the effect of ideology on individual judicial decisions and we reproduce these regression estimates in Table 1 below. In each of the regression models reported in this table, Cox and Miles regress judicial decisions on ideology and a number of variables recording characteristics of the case—in different variations as depicted in columns 1, 2 and 3 of the table. They do not, however, control for race, gender, age or education and yet as noted above, if the structural relationships given in Fig. 2 are correct, this would indicate that all the estimates of the effect of ideology are confounded. To obtain unconfounded estimates of the effect of ideology under Fig. 2, one would have to control for race, gender, age and education as well as the characteristics of the case.

Cox and Miles do not completely ignore the effects of race, gender, age, education and employment on judicial decisions. Table 2 reproduces the authors’ regression estimates in models
that account for these characteristics. Consider first the regression model presented in column 1 with ideology, race, gender and the characteristics of the case as covariates. If our Fig. 2 is a correct representation of the causal relationships among the variables, then the estimate for race reported by Cox and Miles cannot be interpreted as the direct effect of race, controlling for ideology, because ideology is confounded by education and age, which are not controlled for, and thus, the set of controls does not satisfy the second backdoor path criterion for direct effects as discussed in the previous section. Furthermore, the estimate for race cannot be interpreted as the total effect of race because the analysis controls for judge’s ideology, which is a descendent of race in Fig. 2. Similarly, the estimate for gender cannot be interpreted as the direct effect of gender on judicial decisions controlling for ideology because ideology is confounded by education and age, which are not in the model; nor can their estimate for gender be interpreted as a total effect because control is being made for ideology, which is a descendent of gender.
### TABLE 2 Likelihood of judges voting for Section 2 liability in Cox and Miles (2008) study, Table 6, columns 1–3 (probit regressions)

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) Cox and Miles Table 6, column 1</th>
<th>(2) Cox and Miles Table 6, column 4</th>
<th>(3) Cox and Miles Table 6, column 5</th>
<th>(4) Cox and Miles Table 6, column 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge was democratic appointee</td>
<td>0.125** (0.04)</td>
<td>0.140** (0.037)</td>
<td>0.166** (0.039)</td>
<td>0.155** (0.038)</td>
</tr>
<tr>
<td>Judge was African American</td>
<td>0.300** (0.09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge was female</td>
<td>-0.020 (0.056)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>0.003 (0.002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge attended ivy league college</td>
<td></td>
<td></td>
<td>0.018 (0.051)</td>
<td></td>
</tr>
<tr>
<td>Judge attended elite law school</td>
<td></td>
<td></td>
<td>-0.078* (0.041)</td>
<td></td>
</tr>
<tr>
<td>Judge previously served as law clerk</td>
<td></td>
<td></td>
<td>0.016 (0.044)</td>
<td></td>
</tr>
<tr>
<td>Judge previously served in state legal or executive branch</td>
<td></td>
<td></td>
<td></td>
<td>0.021 (0.039)</td>
</tr>
<tr>
<td>Judge previously served on state court</td>
<td></td>
<td></td>
<td></td>
<td>-0.056 (0.040)</td>
</tr>
<tr>
<td>Judge previously served in federal legal or executive branch</td>
<td></td>
<td></td>
<td></td>
<td>0.002 (0.04)</td>
</tr>
</tbody>
</table>

The regressions also include controls for whether the case occurred in a jurisdiction covered by Section 5, whether the case was an appeal, whether the plaintiffs were African American, whether the challenge was to an at-large election scheme or a reapportionment plan, whether the governing body challenged was local and fixed-effect controls for judicial circuits and years. Also see the note immediately below Table 1.
If our Fig. 2 is correct, then to obtain the direct effect of race, controlling for ideology and education, one could regress judicial decisions on race, gender, age, education, ideology and case characteristics; similarly to obtain the direct effect of gender, controlling for ideology and education, one could again simply regress judicial decisions on race, gender, age, education, ideology and case characteristics. Note, however, that these direct effects controlling for ideology and education would include the effects of race (or gender) mediated by employment. One could alternatively obtain estimates of the direct effects of race (or gender), controlling for ideology, education and employment, on judicial decisions by regressing judicial decisions on race, gender, age, ideology, education, employment and case characteristics. These direct effect estimates would then not include effects of race on judicial decisions mediated through prior employment, but it would include the effects mediated through other ‘life experiences’ due to race, a possible mechanism suggested by Cox and Miles. To obtain the total effect of race, one could regress judicial decisions on race and characteristics of the case; to obtain the total effect of gender, one could regress judicial decisions on gender and characteristics of the case. However, if Fig. 2 is correct, the regression analysis of Cox and Miles reported in column 1 (in which judicial decision is regressed on race, gender, ideology and the characteristics of the case) cannot be interpreted as any of the aforementioned effects because ideology is confounded by education and age and no control is made for education and age in this model.

We now turn to Cox and Miles’ discussion of the effects of age, education and employment. Again assuming that our Fig. 2 is correct, the age coefficient reported in Table 2, column 2 (in which judicial decision is regressed on ideology, age and the characteristics of the case) cannot be interpreted as a direct effect of age on judicial decisions controlling for ideology because ideology is confounded by education, race and gender, which are not in the model; nor can their estimate for age be interpreted as a total effect of age because control is being made for ideology, which is a descendent of age. To obtain the direct effect of age on judicial decisions controlling for ideology and education, one could regress judicial decisions on race, gender, age, education, ideology and case characteristics. To obtain the total effect of age on judicial decisions, one could regress judicial decisions on simply age and the characteristics of the case.

Cox and Miles report the results of a regression of judicial decisions on ideology, education and the characteristics of the case (Table 2, column 3) and results of a regression of judicial decisions on ideology, employment and the characteristics of the case (Table 2, column 4). However, if Fig. 2 is correct, the estimates from these regressions cannot be interpreted as the total effects—nor as the direct effects—of education and employment on judicial decisions. This is because the effects of education and employment on judicial decisions are confounded by race, gender and age: there are unblocked backdoor paths from education and employment to judicial decision through race, gender and age. To obtain the total effect of education on judicial decisions, one could regress judicial decisions on education, race, gender, age and the characteristics of the case. Note that race, gender, age and the characteristics of the case block all backdoor paths from education to judicial decisions in Fig. 2. To obtain the direct effect of education, controlling for ideology and employment, on judicial decisions, one could regress judicial decisions on education, ideology, employment, race, gender, age and the characteristics of the case. In Fig. 2, the total and direct effects of employment on

---

6 We note that one could also obtain the direct effect of race, controlling just for ideology (and not education), but this requires techniques other than standard regression analyses (cf. VanderWeele, 2009). This direct effect of race on judicial decisions, controlling only for ideology, would then include the effects mediated through education and employment.
judicial decisions coincide. To obtain the effect (total or direct) of employment on judicial decisions, one could regress judicial decisions on education, race, gender, age and the characteristics of the case.

In summary, if the regression analyses in Tables 1 and 2 are to all be interpreted as Cox and Miles suggest, the causal diagram depicted in our Fig. 1 must be correct, but we believe that Fig. 2 is a more realistic description of the data. Cox and Miles would have more closely approximated the effects of interest, therefore, had they carried out the regression analyses that we suggested above. In the next section, we put our methodology to work: we explore whether and how Cox and Miles’ empirical results would actually change if they had relied upon the DAGs.

6.2 Re-analysis of the Cox and Miles data

While many of the estimates in the Cox and Miles’ study are confounded, it is nonetheless possible that these effects remain statistically and substantively significant even when the proper controls are included in the model. In fact, we find that many (but by no means all) of Cox and Miles’ ‘qualitative’ conclusions survive our re-analysis (i.e. the sign of the coefficients are accurate)—but their ‘quantitative’ conclusions tend to be consistently over- and understated given the problems of bias and confounding in their choice of variables to include in their statistical models.

To enable our re-analysis of the data, Cox and Miles generously agreed to share their data and for this reason, we are able to compare precisely how the differing estimation strategies can and will affect the parameters of interest. To begin our investigation, we first fit the data to Cox and Miles’ statistical model, assuming with the authors that Fig. 1 above accurately reflects the underlying relationships between and among the variables. As expected, we were able to replicate their findings with only minor differences;7 the results of this replication process are presented in Table 3, columns 1–3. We then refit the data, assuming the variables as depicted in Fig. 2. Our results are juxtaposed to those found by Cox and Miles’ in Table 3, columns 1(a), 2(a) and 3(a).

The first thing to note about Table 3 is that, at least in the voting rights context, the qualitative effects of ideology are robust to various sets of controls. Specifically, the role of ideology is both positive and is statistically significant (with p values of approximately 0.05 or less). This finding confirms Cox and Miles’ claim that ideology is causally related to proplaintiff outcomes and Section 2 liability in particular. The authors’ modelling strategy, however, has ‘inflated’ the size of this effect in every context. After including the proper controls, we obtained estimates that were 2.4–5 percentage points lower than those obtained by Cox and Miles; given the relatively small size of the coefficients in all the models, this means that due to confounding Cox and Miles have over-estimated the effects of ideology by 19, 45, and 50% in column 1, 2, and 3, respectively.8 Although accounting for this inadvertent exaggeration does not change the authors’ underlying claim with respect to the positive effects of ideology in this particular context, in other situations such changes can have a substantive effect even on both the qualitative and the quantitative conclusions given the possibility that the sign of the coefficient will change—a problem that we show in fact emerges in the context of gender.

---

7 Cox and Miles reported a marginal effect of 0.158 for ideology in Table 1, column 3, but when we re-estimated their model, we found a marginal effect of 0.148. We believe that this is simply a typographical error.

8 We calculated these percentages by dividing Cox and Miles’ estimates for ideology by our estimates. For example, 0.145/0.121 = 1.19 indicating that Cox and Miles’ estimate is a 119% of ours, or 19% larger.
<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) Replication of Cox and Miles Table 5, column 1</th>
<th>(1a) VanderWeele and Staudt model</th>
<th>(2) Replication of Cox and Miles Table 5, column 2</th>
<th>(2a) VanderWeele and Staudt model</th>
<th>(3) Replication of Cox and Miles Table 5, column 3</th>
<th>(3a) VanderWeele and Staudt model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge was democratic</td>
<td>0.145*</td>
<td>0.121**</td>
<td>0.151**</td>
<td>0.104**</td>
<td>0.149**</td>
<td>0.099**</td>
</tr>
<tr>
<td>appointee</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.039)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Judge was democratic</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>appointee × year after 1994</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Year was after 1994</td>
<td>−0.12**</td>
<td>−0.14**</td>
<td>−0.102**</td>
<td>−0.11**</td>
<td>−0.101**</td>
<td>−0.11**</td>
</tr>
<tr>
<td>Case occurred in south</td>
<td>0.01</td>
<td>0.01</td>
<td>−0.018</td>
<td>−0.02</td>
<td>−0.018</td>
<td>−0.02</td>
</tr>
<tr>
<td>Appellate case</td>
<td>−0.08**</td>
<td>−0.09**</td>
<td>−0.11**</td>
<td>−0.11**</td>
<td>−0.101**</td>
<td>−0.11**</td>
</tr>
<tr>
<td>Challenge to at-large election</td>
<td>0.10**</td>
<td>0.08*</td>
<td>0.078</td>
<td>0.06</td>
<td>0.077</td>
<td>0.06</td>
</tr>
<tr>
<td>Challenge to reappointment plan</td>
<td>0.05</td>
<td>0.03</td>
<td>0.034</td>
<td>0.008</td>
<td>0.034</td>
<td>0.008</td>
</tr>
<tr>
<td>Challenge to local election practice</td>
<td>0.004</td>
<td>0.005</td>
<td>−0.018</td>
<td>−0.02</td>
<td>−0.018</td>
<td>−0.02</td>
</tr>
<tr>
<td>Plaintiffs were</td>
<td>0.02</td>
<td>0.01</td>
<td>0.114**</td>
<td>0.11**</td>
<td>0.113**</td>
<td>0.11**</td>
</tr>
<tr>
<td>African American</td>
<td>0.04</td>
<td>0.04</td>
<td>0.045</td>
<td>0.03</td>
<td>0.044</td>
<td>0.03</td>
</tr>
<tr>
<td>Case occurred in jurisdiction covered by §5</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Judge’s age</td>
<td>—</td>
<td>—</td>
<td>0.003**</td>
<td>—</td>
<td>0.005**</td>
<td>—</td>
</tr>
<tr>
<td>Judges’ race</td>
<td>—</td>
<td>0.27**</td>
<td>—</td>
<td>0.36**</td>
<td>—</td>
<td>0.36**</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(0.08)</td>
<td>—</td>
<td>(0.09)</td>
<td>—</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Variable</td>
<td>(1) Replication of Cox and Miles Table 5, column 1</td>
<td>(1a) Replication of VanderWeele and Staudt model</td>
<td>(2) Replication of Cox and Miles Table 5, column 2</td>
<td>(2a) Replication of VanderWeele and Staudt model</td>
<td>(3) Replication of Cox and Miles Table 5, column 3</td>
<td>(3a) Replication of VanderWeele and Staudt model</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Judge’s gender</td>
<td>—</td>
<td>0.003</td>
<td>—</td>
<td>0.02</td>
<td>—</td>
<td>0.02</td>
</tr>
<tr>
<td>Judge attended ivy league college</td>
<td>—</td>
<td>−0.003</td>
<td>—</td>
<td>0.02</td>
<td>—</td>
<td>−0.02</td>
</tr>
<tr>
<td>Judge attended elite law school</td>
<td>—</td>
<td>−0.07*</td>
<td>—</td>
<td>−0.07*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Judge previously served as law clerk</td>
<td>—</td>
<td>0.09*</td>
<td>—</td>
<td>8</td>
<td>—</td>
<td>0.08</td>
</tr>
<tr>
<td>Circuit fixed effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>679</td>
<td>677</td>
<td>652</td>
<td>650</td>
<td>652</td>
<td>650</td>
</tr>
</tbody>
</table>

See note immediately below Table 1.
### TABLE 4 Replication of Cox and Miles’ Table 6, columns (1), (5) and (6) and re-estimation of results to account for bias and confounding

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) Replication Cox and Miles Table 6, column 1</th>
<th>(1a) Replication VanderWeele and Staudt model</th>
<th>(2) Replication Cox and Miles Table 6, column 5</th>
<th>(2a) Replication VanderWeele and Staudt model</th>
<th>(3) Replication Cox and Miles Table 6, column 6</th>
<th>(3a) Replication VanderWeele and Staudt model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge was democratic</td>
<td>0.124**</td>
<td>0.104**</td>
<td>0.166**</td>
<td>0.104**</td>
<td>0.154**</td>
<td>0.113*</td>
</tr>
<tr>
<td>appointee</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Judge was African American</td>
<td>0.300**</td>
<td>0.36**</td>
<td>—</td>
<td>0.362**</td>
<td>—</td>
<td>0.377**</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Judge was female</td>
<td>−0.020</td>
<td>0.02</td>
<td>—</td>
<td>0.02</td>
<td>—</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Age</td>
<td>—</td>
<td>0.005**</td>
<td>—</td>
<td>0.005**</td>
<td>—</td>
<td>0.006**</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Judge attended ivy league college</td>
<td>—</td>
<td>0.02</td>
<td>0.018</td>
<td>0.02(0.05)</td>
<td>—</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Judge attended elite law school</td>
<td>—</td>
<td>−0.11*</td>
<td>−0.078*</td>
<td>−0.07*</td>
<td>—</td>
<td>−0.08</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Judge previously served as law clerk</td>
<td>—</td>
<td>0.08</td>
<td>0.016</td>
<td>0.08</td>
<td>—</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Judge previously served in state legal or executive branch</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.021</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Judge previously served on state court</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>−0.056</td>
<td>−0.08*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Judge previously served in federal legal or executive branch</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.002</td>
<td>−0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>N</td>
<td>652</td>
<td>650</td>
<td>652</td>
<td>650</td>
<td>652</td>
<td>650</td>
</tr>
</tbody>
</table>

The regressions include controls for whether the case occurred in a jurisdiction covered by Section 5, whether the case was an appeal, whether the plaintiffs were African American, whether the challenge was to an at-large election scheme or a reapportionment plan, whether the governing body challenged was local, circuit fixed effects and year fixed effects. Also see the note immediately below Table 1.
After identifying the effects of ideology, Cox and Miles turn to the judges’ other personal characteristics: race, gender, education and past employment. With minor exceptions, we were again able to replicate the authors’ findings and we present these results in columns 1–3 of Table 4 below.\(^9\) We then adopted the modelling strategy that we believe better accounts for the true underlying causal relationships and present our findings in columns 1(a), 2(a) and 3(a) below.

Beginning first with the effects of race. We again find support for Cox and Miles’ claim that a judge’s race affects the likelihood of voting in favour of Section 2 liability, but in this context, we believe that the authors reported ‘deflated’ coefficients. They find that controlling for ideology, African American judges are 30% more likely to vote in favour of liability than white judges,\(^10\) but we find the actual direct effect is 6–8 percentage points higher.\(^11\) This means that the authors underestimated causal effects of race by 17–20%.\(^12\) These findings along with those discussed immediately above with respect to ideology suggest that race has quite a bit stronger effect on outcomes than the authors believe, while ideology has less of an effect than originally estimated. We deem these twin findings of our modelling process to be important for at least two reasons. First, identifying unbiased effects of the variables allows researchers to have better confidence in their empirical claims and in inferences about causation. Without addressing the problem of confounding, causal claims are completely unwarranted. Second, our findings make Cox and Miles’ study all the more important to the literature. As they note, quite a few scholars have investigated the effects of ideology on judicial decision making but few have explored the effects of race and those that have done so have produced null findings. Cox and Miles’ data—after addressing the problems of bias and confounding by appropriate covariate controls—not only suggest that race is an important factor to consider but also that the direct effect of race is roughly 250% greater than that of ideology in the decision-making context when it comes to voting rights claims.\(^13\) This is a finding that is important for understanding and predicting judicial outcomes and also for the judicial appointment process.

With respect to the direct effects of gender, our models produce contrasting qualitative results: Cox and Miles suggest that gender has a negative effect on the likelihood of voting for Section 2 liability, while our revised model suggests that the effects are positive. By underestimating the effects of gender on judicial outcomes, in short, the authors misidentified the sign of the coefficient. In neither approach, however, do the results achieve statistical significance as shown in columns 1 and 1(a) of Table 4.\(^14\)

With respect to the direct effects of education, presented in columns 2 and 2(a) of Table 4, our qualitative results are very similar: we, like Cox and Models, find positive but statistically insignificant effects associated with college and clerking, and negative, slightly significant effects caused by a judge’s decision to attend an elite law school.\(^15\) A comparison of the two sets of findings, however, suggests that Cox and Miles’ education coefficients are slightly overstated.

\(^9\) Compare our replicated findings in Table 3 to Cox and Miles’ original finding reproduced in Table 2.
\(^10\) Cox and Miles note that the data set comprised primarily black and white judges (Cox and Miles, 2008 p. 30).
\(^11\) The total effects of race can be estimated by regressing the judicial decision on race with controls only for case characteristics. We estimated these effects and obtained a marginal effect of 0.286, which is statistically significant at the \(p \leq 0.01\) level.
\(^12\) We calculated these percentages by dividing Cox and Miles estimate (0.3) by our own estimates (0.36, 0.362, 0.377).
\(^13\) Cox and Miles estimation process suggests this difference is 140%. See results presented in Table 2.
\(^14\) The total effects of gender on the likelihood of finding liability can be estimated by regressing the judicial decision on gender and case characteristics; this effect was also not statistically significant.
\(^15\) The total effects of education, obtained by regressing the judicial decision on education, race, gender, age and characteristics, indicates that an ivy league college and elite law school education have a negative (but not statistically significant)
Columns 3 and 3(a) depict the effects of past employment experience: our models again produce similar qualitative results ‘except’ with respect to a judge’s past experience on a state court. We estimated the total and direct effects (they coincide as noted above) of a judge’s state court experience and identified a 8% decrease in the likelihood of a judge voting for the plaintiff and this finding is statistically significant at the $p \leq 0.10$ level. Cox and Miles also estimated the effects of state court experience to be negative but not at a statistically significant level—an imprecise estimate most likely due to the decision to exclude the confounding variables, race, gender, age and education.

Finally, Cox and Miles’ findings suggest that age plays no direct role in a judge’s propensity to vote for or against the plaintiff (see Table 2, column 2 above), but we consistently find a positive and statistically significant effect for age in every model we estimate. For every year that a judge is older (or for every year earlier the judge was born), our models suggest that the likelihood of voting for Section 2 liability increases by 0.3–0.6%.\footnote{The total effects of age can be estimated by regressing judicial decisions on age and case characteristics. This model produces a marginal effect of 0.003 and statistical significance at $p \leq 0.10$.} This means that the oldest judge in the database (90 years of age) is 18–34% more likely to render a proplaintiff vote that the youngest judge (31 years of age). Without the proper controls for race, gender and education in the estimation of this direct effect, this finding was not observed.

We summarize the differences between our findings and Cox and Miles’ findings in Table 5 below. As the table indicates, our methodology slightly weakens the authors’ conclusions with respect to ideology but strengthens the conclusions with respect to race. Moreover, we identify a statistically significant role for both age and past service on a state court—findings hidden in Cox and Miles’ study due to bias and confounding. In short, while we admire Cox and the Miles’ work, we believe the comparisons presented below confirm our claim that empirical researchers should spend more time and energy considering the underlying causal relationships of the variables of interest prior to specifying and fitting statistical models. This will help assure scholars’ claims about causality are justifiable, will avoid problems of under- and overestimation of target parameters and will potentially enable more precise estimates.

7. Conclusions

Recent methodological advances associated with graphical modelling of causal relationships have made it possible to address the barriers to causal inference in a remarkably simple and straightforward, yet rigorous, manner. Specifically, directed acyclic graphing methods or causal DAGs, developed primarily in statistics, epidemiology and computer science, enable empirical researchers to construct diagrams that not only make modelling assumptions explicit but also determine when these assumptions are sufficient for obtaining consistent estimates, and how to specify a closed-form model for determining the quantity of interest when identification is possible (Pearl, 2009; Greenland et al., 1999). We discussed these methods in the context of empirical legal research and show how investigators can use DAGs to address the most basic and fundamental tasks of empirical research.

We have described the formal rules governing inferences about confounding that can be drawn from the causal DAG. The graphs themselves encode structural assumptions. Legal researchers will
### Table 5: Summary of conclusions with respect to Cox and Miles’ study on the voting rights act

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cox and Miles’ marginal effects</th>
<th>Change in statistical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge was democratic appointee</td>
<td>Overestimated by 19–50%</td>
<td>Yes; while still statistically significant not always at $p \leq 0.05$ as authors suggest</td>
</tr>
<tr>
<td>Judge was African American</td>
<td>Underestimated by 17–20%</td>
<td>No</td>
</tr>
<tr>
<td>Judge was female</td>
<td>Sign change: estimated negative when in fact positive</td>
<td>No</td>
</tr>
<tr>
<td>Age</td>
<td>Underestimated by 0–50%</td>
<td>Yes; finding is statistically significant but authors argued that it was not</td>
</tr>
<tr>
<td>Judge Attended ivy league college</td>
<td>Underestimated by 9%</td>
<td>No</td>
</tr>
<tr>
<td>Judge Attended elite law school</td>
<td>Overestimated by 11%</td>
<td>No</td>
</tr>
<tr>
<td>Judge previously served as law clerk</td>
<td>Underestimated by 80%</td>
<td>No</td>
</tr>
<tr>
<td>Judge previously served in state legal or executive branch</td>
<td>Overestimated by 210%</td>
<td>No</td>
</tr>
<tr>
<td>Judge previously served on state court</td>
<td>Underestimated by 30%</td>
<td>Yes; finding is statistically significant but authors argued that it was not</td>
</tr>
<tr>
<td>Judge previously served in federal legal or executive branch</td>
<td>Sign change: estimated positive when in fact negative</td>
<td>No</td>
</tr>
</tbody>
</table>
use substantive knowledge and prior studies (if available) to draw these graphs. In cases in which
the causal structure of the graph is not clear, it is possible to draw several graphs and consider how
conclusions about confounding vary with each graph and how the results of empirical analyses vary
when control is made for different variables based on these conclusions.

Several broad intuitive conclusions also emerge from the use of these causal DAGs in reasoning
about confounding. First, if the total effect of a particular variable is of interest, then control should
generally not be made for intermediates on the pathway from the variable of interest to the outcome
but control should be made for variables that affect both the exposure variable of interest and the
outcome. If controlled direct effects are of interest, and control for one or more intermediates along
the pathway is made, then it is also necessary control for variables that confound the relationship
between the intermediates and the outcome. An important implication of these guidelines is that
separate regressions will often be required for different effects of interest. Variables that confound
one effect of interest may not confound another; variables that are on the pathway for a certain effect
of interest may not be on the pathway for some other effect. For each effect of interest, a researcher
should use the guidelines given above to determine for which variables control should be made. In
short, if the effect of interest changes, then the variables for which control is to be made will often
change as well.

It should be noted that to obtain valid estimates of causal effects, empirical researchers must
identify a set of variables that effectively control for confounding. In situations in which causal anal-
ysis is impossible due to the existence of important but unmeasurable variables, sensitivity analysis
techniques are useful. Techniques have been developed that have made it possible to assess the ex-
tent to which an unmeasured confounding variable would have to be related to both the treatment or
exposure of interest and the outcome in order to invalidate the qualitative conclusions drawn from the
analysis (Rosenbaum and Rubin, 1983; Imbens, 2003; VanderWeele and Arah, 2011).

The intuitive rules described above can help legal researchers in their decisions about which
variables to include in a model when seeking to identify the particular effect of interest. In cases
in which it is not clear whether control should be made for a variable the precise rules de-
scribed in Section 4 concerning blocked paths can be used to guide the researcher’s decision
making. We demonstrated how these rules could be applied to an empirical study of judicial
decision making in the voting rights context. Specifically, with the help of the DAG frame-
work, we demonstrated how even widely admired studies can be plagued with problems of over-
and underestimation of coefficients and imprecise results when researchers fail to rigor-
ously specify the underlying causal structure of the variables. Importantly, the methodology can
be applied to a wide range of empirical legal studies not only to those investigating judicial deci-
sion making: indeed, virtually every legal empirical researcher who estimates a statistical model
would benefit by making qualitative assumptions about their data transparent with the help of a
DAG.

This paper is merely an introduction to the topic; many other extensions to the causal DAG
framework are possible (Hernán et al. 2004; VanderWeele and Robins, 2007; VanderWeele et al.
2008; Pearl, 2009; White and Chalak, 2009; Shpitser et al., 2010) and we encourage empirical legal
researchers to consult the literature further.

Supplementary Material

Online appendix is available at http://www.lpr.oxfordjournals.
Acknowledgements

We would like to thank Frank Easterbrook, Lee Epstein, Bill Landis, Jim Lindgren, Richard Posner, and the participants in the Northwestern University Law School and University of Chicago Law School Judicial Behavior Workshop, the editor and two anonymous reviewers for their helpful insights.

REFERENCES


