Received: 13 April 2023 | Revised: 13 October 2023 | Accepted: 11 January 2024

DOI: 10.1002/smj.3588

RESEARCH ARTICLE

SMS | Strategic Management Journal

WILEY

Collider bias in strategy and management research: An illustration using women CEO's effect on other women's career outcomes

Justin Frake[1] | Andreas Hagemann[1] | Jose Uribe[2]

[1]Stephen M. Ross School of Business, University of Michigan, Ann Arbor, Michigan, USA

[2]Kelley School of Business, Indiana University, Bloomington, Indiana, USA

Correspondence
Jose Uribe, Kelley School of Business, Indiana University, Bloomington, IN 47405, USA.
Email: uribej@iu.edu

Abstract

Research Summary: Collider bias can cause spurious correlations when researchers condition on a variable that is caused by—or shares a common cause with—both the outcome and the exposure variable. Despite its threat to inference, empirical research in strategy and management has largely overlooked the issue of collider bias. We distinguish colliders from other threats to identification and estimation and illustrate its importance with replications of published work suggesting that having a woman CEO reduces the career outcomes (compensation and representation) of other women executives. After accounting for collider bias, we find no evidence that women CEOs damage the career outcomes of other women in their organizations. We close by providing generalizable approaches to identify and mitigate the risk of collider bias in applied research.

Managerial Summary: Collider bias is a type of statistical problem that can generate misleading results in empirical research. Although research in strategy and management has given substantial attention to other types of statistical problems, the issue of collider bias has not received sufficient scrutiny. We illustrate this point with replications of published work suggesting that having a woman CEO reduces the career outcomes of other women executives. After accounting for collider bias, we find no evidence that women CEOs damage the career outcomes of other women in their organizations. We provide advice for detecting and addressing collider bias in empirical research.

KEYWORDS

CEO gender, colliders, identification, replication

1 | INTRODUCTION

Strategy and management scholars are increasingly concerned with potential sources of bias in their empirical work. In this regard, significant strides have been made to address the threats to inference posed by omitted variables and non-classical measurement error (Bascle, 2008; Ge et al., 2016; Hamilton & Nickerson, 2003; Shaver, 1998; Wolfolds & Siegel, 2019). Yet, we contend that there is insufficient recognition of collider bias, which poses a serious threat the validity of empirical findings in strategy and management. Colliders are variables that are caused—or share a common cause with—both the outcome and an independent variable. Collider bias occurs when the relationship between two variables is distorted by conditioning (e.g., controlling or selecting) on a collider variable (Cinelli et al., 2022; Elwert & Winship, 2014; Griffith et al., 2020; Schneider, 2020). Our review shows that papers published in leading strategy and management journals discuss collider bias at roughly half the rate of top economics papers, and four times less than those in leading sociology journals.

We formally define collider bias and explain differences from biases induced by confounders, mediators, and other identification and estimation issues. Collider bias is distinct from sample selection bias (Certo et al., 2016; Shaver, 1998) caused by a confounder, or unobserved variable affecting both the treatment and the outcome. By contrast, collider bias occurs when selecting (or otherwise conditioning) on a variable that is itself influenced by—or shares a common cause with—both the outcome and the treatment. We discuss how collider bias intersects with other identification challenges such as bad controls, selection on the dependent variable, nonresponse bias, and attrition bias.

We also illustrate how collider bias can induce spurious relationships that can have potentially negative practical effects on managerial decisions. An expansive literature examines the effects of having a woman or minority CEO on firm outcomes (e.g., Cook & Glass, 2014; Jeong et al., 2021; Jeong & Harrison, 2017). In this work, an important human capital outcome for firms is the level of diversity of top management teams (TMTs) and the career advancement of women and minorities (e.g., Chang et al., 2019; Corwin et al., 2022; Derks et al., 2016; Dezső et al., 2016; McDonald et al., 2018). We carry out an empirical demonstration of collider bias using the well-known ExecuComp and Compusat datasets to examine the relationship between women CEOs on both the prevalence and compensation of women in TMTs from 1992 to 2021. In line with published work, our analysis shows a large and precisely estimated negative effect of having a woman CEO on the compensation (cf. Dezső et al., 2022) and prevalence (cf. Corwin et al., 2022) of other women in the TMT.

We argue that the effects we find are the result of conditioning on a collider, namely the propensity of women to exit the sample when promoted from non-CEO to CEO positions.

Rather than gender dynamics, the data-generating process includes unexpected panel attrition: promotion into the CEO position implies exit from the sample. To support this claim, we show large and precisely estimated negative relationships between having a CEO with a variety of placebo characteristics (such as having the name "John") on non-CEOs who share the same characteristic. Moreover, we answer Shaver's (2020) request to go beyond calls for replication, showing that replication efforts that do not attend to the specific source of collider bias will yield robust coefficients of misleading results.

Finally, we advocate the use of directed acyclic graphs (DAGs) for identifying potential sources of collider bias. We also assess various corrective methods, including the use of fixed effects, inverse probability weighting (IPW), multiple imputation, and placebo analysis.

Our study draws attention to the importance of considering collider bias in empirical research in management and strategy. Even though there is no silver bullet for solving collider bias when present, it is critically important to acknowledge its presence. In our example, failure to detect collider bias results in findings that some could use to support policies that reduce the promotion of women into the CEO position. Broader awareness can therefore increase the validity and usefulness of our findings for informing theory and practice.

Perhaps, the most fundamental aim is a renewed call for empiricists to attend to identification and the nature of the data-generating process (see also Shaver, 2020). While other types of empirical problems such as classical measurement error often merely attenuate effects, collider bias can reverse the sign on estimated coefficients. The widely acknowledged publication pressures on faculty, particularly pre-tenured faculty, and the field's penchant for surprising and counterintuitive findings (e.g., Davis, 2015) makes attention to collider bias particularly timely.

2 | BACKGROUND

The primary aim of most empirical studies in strategy and management is to accurately estimate unbiased causal effects (Hill et al., 2021; Shaver, 1998, 2020, 2021). A stream of work has underscored the necessity of accounting for sources of bias (Hamilton & Nickerson, 2003; Shaver, 1998; Stern et al., 2021) such as omitted variable bias (Busenbark et al., 2022; Wolfolds & Siegel, 2019), selection bias (Certo et al., 2016), measurement error (Boyd et al., 2005; Ge et al., 2016), and multicollinearity (Kalnins, 2018). Despite the growing attention to estimation and identification in the field, the issue of collider bias remains a relatively unexplored threat to causal inference. To illustrate this point, we compare attention to confounders and attention to colliders in articles published between 2010 and 2023 in key journals across the fields of strategy and management, economics, and sociology. In each of the selected top journals, we conducted a search for mentions of potential bias from confounders and from colliders. The results of this exercise are presented in Table 1.

There is a relative underrepresentation of terms associated with collider bias in prominent management and strategy journals, compared to prominent journals in economics and sociology. Only 0.68% of articles in management and strategy journals mention collider-related terms, a rate that is about 50% lower than economics journals and 80% lower than sociology journals. When it comes to confounders, 19% of strategy and management articles mention the term, which is only 40% lower than sociology and 50% higher than economics. These comparisons are noteworthy for two reasons. First, the types of econometric models and data used in strategy and management are just as likely to suffer from collider bias as those used in economics and sociology. Second, threats to causal inference from collider bias are as important as the

TABLE 1 Frequency of mentions of "collider" and "bad controls."

Field	Journal (2010–2023)	# Published	% Discussing colliders	% Discussing confounders
Strategy/management	Strategic Management Journal	1504	1.33%	29.65%
	Management Science	1390	0.29%	13.60%
	Organization Science	741	0.54%	14.17%
	Administrative Science Quarterly	623	0.48%	13.32%
	Academy of Management Journal	617	0.32%	16.69%
	TOTAL Strategy/Management	4875	0.68%	18.99%
Economics	American Economic Review	2426	1.07%	11.87%
	Quarterly Journal of Economics	348	2.01%	19.54%
	Journal of Political Economy	328	1.22%	10.06%
	TOTAL ECONOMICS	3102	1.19%	12.54%
Sociology	American Sociological Review	690	3.62%	35.80%
	Annual Review of Sociology	197	2.54%	13.20%
	TOTAL SOCIOLOGY	887	3.38%	30.78%

Note: Collider-related papers are any that mention "collider," "bad control," or "endogenous selection." Papers discussing confounders were identified by searching their text for "confounder." Because collider is a commonly used word outside of the econometric context, we read the relevant passages from each article that used this term to verify that it was used in the context of a collider variable. We also selected a random sample of articles mentioning "bad control" and "endogenous selection" and confirmed over 90% were using the terms in the context of their impact on estimation or causal inference.

well-known biases that arise from omitting confounders. We next illustrate the distinction between confounders and colliders using DAGs, an approach originally developed by Pearl (2000).

2.1 | Using DAGs to understand collider bias

DAGs are a visual tool for mapping the causal connections among variables. DAGs rely on simple graphical rules: each node symbolizes a random variable, while arrows indicate direct causal connections. An absence of arrows between two nodes signifies that no causal link exists between the variables they represent. The term "path" denotes an ordered sequence of arrows that connect two variables.[1]

When all arrows on a path point from the treatment variable to the outcome variable, it is referred to as a "causal path." In general, causal identification aims to identify the total effects of the treatment across all causal paths. For example, in Figure 1a the total causal effect of the treatment (X) on the outcome (Y) would include the following causal paths: X → Y and X → M → Y. Paths that are not causal are known as "backdoor paths," which can induce bias

[1]Online Appendix A contains a more detailed discussion of DAGs.

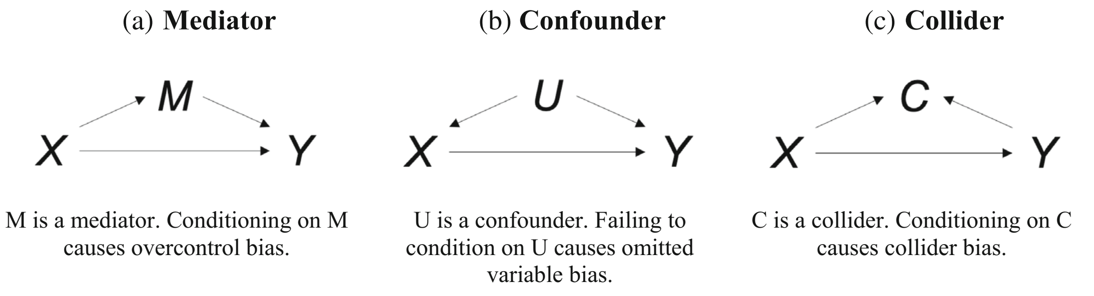


FIGURE 1 Illustrating confounders and colliders using directed acyclic graphs (DAGs).

when estimating the relationship between the treatment and the outcome. For example, X ← U → Y (in Figure 1b) and X → C ← Y (in Figure 1c), are backdoor paths.

Besides the treatment and outcome variables, three basic types of variables exist in a DAG: mediators, confounders, and colliders.[2] Mediators are variables found in a causal path. In Figure 1a, M is a mediator because it is influenced by X and subsequently affects Y. Confounders exist on backdoor paths and cause two or more other variables. In Figure 1b, U is a confounder because it causes X and Y. Colliders also exist on backdoor paths but are influenced by two or more other variables. In Figure 1c, C is a collider because it is caused by X and Y.

In the context of a DAG, the objective of causal identification is to ensure that all causal paths remain open, allowing for unobstructed analysis of direct relationships between variables, while closing all backdoor paths that could introduce bias. Conditioning—controlling, selecting, or stratifying on a variable—is the primary method used for opening and closing paths. The effect of conditioning varies depending on the type of variable involved. Conditioning on mediators and confounders closes the path. For instance, conditioning on M in Figure 1a would lead to overcontrol bias by closing the causal path X → M → Y. Further, failing to condition on U would induce omitted variable bias. Conditioning on colliders opens the path. In Figure 1c, the path X → C ← Y is already closed when C is not conditioned upon because C absorbs the variation from X and Y. However, conditioning on C opens this backdoor path and allows for a spurious correlation between X and Y, leading to collider bias, which is the focus of this article.

2.2 | Checking typical hiding spots for collider bias

Collider bias manifests in various forms and intersects with a range of empirical issues, including but not limited to bad controls, selection bias, selection on the dependent variable, nonresponse bias, and attrition bias. Table 2 summarizes the most likely forms of collider bias in strategy and management research.

2.2.1 | Controls, matching, and fixed effects

When conditioning on a collider through statistical controls, including fixed effects, researchers introduce what is known as a "bad control" (Angrist & Pischke, 2009; Cinelli et al., 2022). For

[2]DAGs do not readily accommodate moderators, which represent the drivers of effect heterogeneity but do not alter fundamental causal pathways.

TABLE 2 Main forms of collider bias.

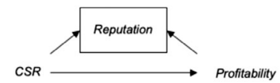
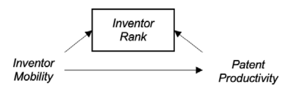
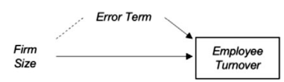
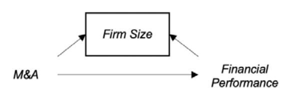
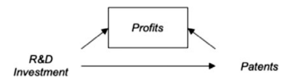
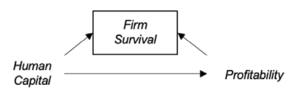
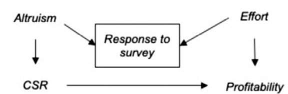
Category	Description	Example research question	Simplified DAG	Potential cause of collider bias	Explanation
Controls and matching	Do not control for or match on variables that may be colliders	What is the effect of CSR on profitability?		Controlling for or matching on firm reputation	Researchers should not control for or match on a firm's concomitant reputation when studying the impact of CSR on firm profits because both CSR and profits likely affect a firm's reputation.
Fixed effects	Do not include fixed effects that may be colliders	What is the effect of inventor mobility on patent productivity?		Inventor's rank fixed effects	Both inventor mobility and patent productivity may affect an inventor's rank in the firm (engineer vs. senior engineer), making it a collider variable when used as a fixed effect.
Selection on the-dependent variable	Ensure that the criteria for being included in a dataset is not based on the value of the dependent variable	What is the effect of firm size on employee turnover?		Selection into making LinkedIn profile public	Employees make their LinkedIn profiles public when they would like to change employers (turnover). By using public LinkedIn profiles to measure turnover, researchers are implicitly selecting on the dependent variable.
Selection into archival dataset	Ensure that the criteria for being included in a dataset is not based on a collider	What is the effect of M&A on a firm's financial performance?		Selection of Fortune 500 firms	Both the M&A activity and firm performance could influence whether a firm is on the Fortune 500 list. Thus, researchers should not use a dataset of Fortune 500 firms to study the research question because being on the Fortune 500 list is a collider.

TABLE 2 (Continued)

Category	Description	Example research question	Simplified DAG	Potential cause of collider bias	Explanation
Sample exclusion criteria	Do not exclude observations from a sample based on a collider	What is the effect of R&D investment on patent productivity?		Excluding observations from sample based on missing measure of profit	If R&D investments and patenting both affect a firm's profitability, excluding observations where profit measures are missing would result in collider bias.
Attrition	Be aware of units that leave sample based on a collider	What is the effect of entrepreneurial human capital on startup's profitability		Exit from the sample based on firm failure	Entrepreneurs with high human capital are more likely to close their firm due to better outside options. Firms with low profits are also more likely to fail. Thus, attrition due to failure is likely a collider.
Nonresponse	Ensure that subjects do not decline responding to a survey based on a collider	Does CSR affect profitability?		Nonresponse to a survey of CEOs	If altruistic CEOS are more likely to respond to a survey and to engage in CSR, and if CEOs of high performing firms are less likely to respond to a survey because they are busy, then nonresponse is a collider.

instance, when examining the effect of corporate social responsibility (CSR) on performance, one should avoid controlling for a firm's inclusion in Forbes' Most Admired Companies list, as both CSR and performance likely influence a firm's likelihood of being featured on the list. Similarly, techniques like propensity score matching or coarsened exact matching can induce collider bias if the researcher matches on a collider. Our general recommendation is to only use controls, fixed effects, or matching variables that occur and are measured prior to treatment. Any control variables that occur after treatment are likely to be mediators or colliders, both of which researchers should generally avoid conditioning on.

2.2.2 | Selection into datasets

Collider bias can be particularly difficult to detect when it arises through the sample selection process. The most egregious example is selection on the dependent variable. While most researchers acknowledge that selecting on the dependent variable can introduce bias in estimates, fewer recognize this as a special instance of collider bias. In such cases, the dependent variable serves as a collider between the treatment variable and the error term. By conditioning on the dependent variable, researchers open the backdoor path from the independent variable to the error term, thereby inducing bias. For example, when investigating the impact of firm size on employee turnover, researchers should refrain from using data sourced from public LinkedIn profiles (e.g., Revelio) because people are more likely to make their profiles public, and thus select into the dataset, when they intend to leave their employer.

Collider bias can also result from selecting on other variables that are caused by the dependent variable. For example, selection into the MSCI (KLD) database, which is based on firms' voluntary disclosure of their CSR activities, acts as a collider because a firm decides whether to disclose its CRS based on its CSR activity and its financial performance (we will elaborate on this example in Section 2.3). As another example, Miller et al. (2010) use data from the largest 1000 firms in the United States (Fortune 1000) to find that family ownership is negatively related to acquisitions. Unless firm size is causally independent from acquisition intentions or from family ownership, selecting on large firms may induce collider bias. For example, if family firms are likely to be smaller and if acquisition make firms larger, then conditioning on the largest firms acts as a collider. This can result in a negative correlation between family ownership and acquisitions even if the actual correlation in the universe of all firms is positive or nonexistent. To identify collider bias from selection into a dataset, researchers should be explicit about the dataset's inclusion criteria. After articulating a complete DAG, researchers can carefully consider whether the inclusion criteria might act as a collider in the data-generating process.

2.2.3 | Sample exclusion criteria

Scholars should also be careful to avoid selecting on colliders when excluding observations from their sample. For instance, strategy and management scholars often drop observations for which they do not have control variables. If the reason for not having control variables is related to a collider, then the decision to exclude observations could result in collider bias. For instance, several studies explore the link between an inventor's patenting history and their likelihood to change employers (Hoisl, 2007; Melero et al., 2020; Palomeras & Melero, 2010). These studies often limit their samples to inventors who have patented at least twice. This constraint is due to the need for a minimum of two patents to measure a single mobility event—one to identify the originating firm and another to identify the destination firm. However, if the first patent increases the likelihood of subsequent patenting, and if changing employers also influences the propensity to patent (Kaiser et al., 2015), then this sample restriction based on the total number of career patents can introduce collider bias.

Attrition and nonresponse bias can also operate as a form of collider bias when leaving the sample or participating in a survey is influenced by both the independent and dependent variables. Imagine a study where a researcher surveys CEOs to examine whether CSR activities lead to improved financial performance. If CEOs who are more altruistic are both more likely to engage in CSR and more inclined to respond to the survey, and if hard working/busy CEOs are more profitable and are less likely to respond to surveys, then the study could be compromised by collider bias arising from nonresponse.

2.3 | Illustrative example

To make the concept of collider bias more concrete, consider the example of studying the effect of CSR on a firm's financial performance. Despite the existence of over 2200 empirical papers on this topic (Friede et al., 2015), scholarly opinion remains divided (Awaysheh et al., 2020). While some propose a negative relationship, contending that CSR benefits stakeholders at the expense of shareholders (Friedman, 1970), others posit a positive effect due to the reputational advantages of CSR Hornstein & Zhao, 2018). Scholars like McWilliams and Siegel (2000) attribute this lack of consensus to confounders and omitted variable bias, whereas Awaysheh et al. (2020) point to measurement issues. However, collider bias may be just as severe a threat in many studies.

To illustrate this, imagine a researcher using the MSCI (KLD) database to obtain CSR metrics to construct their key independent variable.[3] This researcher also acquires a complete and reliable dataset to accurately measure revenues for all firms. After carefully controlling for all confounding variables such as industry and R&D (McWilliams & Siegel, 2000), the researcher runs a regression and estimates a significant, positive effect of CSR on revenues. Because not all firms have CSR metrics in the MSCI (KLD) database, authors will typically acknowledge the sample's limitations with a disclaimer such as "Among firms that voluntarily disclose their CSR activities, CSR increases firm performance." In other words, after acknowledging possible threats to generalizability, the estimate is deemed valid for the sample at hand. However, this conclusion may be wrong due to collider bias.

Collider bias in this case stems from sample selection based on firms' CSR disclosures. This is not a minor concern affecting merely the generalizability of the results; rather, it renders the observed CSR effect on financial performance spurious *even within the selected sample*. For further clarity, let us assume that the true causal relationship between CSR and financial performance is negative (as shown in Figure 2a). The MSCI (KLD) database relies on voluntary CSR disclosures via various channels, including company websites and annual reports. With that in mind, suppose that firms with strong CSR records are more likely to disclose their CSR activities, and thus be overrepresented in the MSCI (KLD) database. Suppose also that firms with weak financial performance are more likely to emphasize their CSR activities as a

[3]Other similar datasets include Refinitiv ASSET4, Dow Jones Sustainability Index, and Sustainalytics.

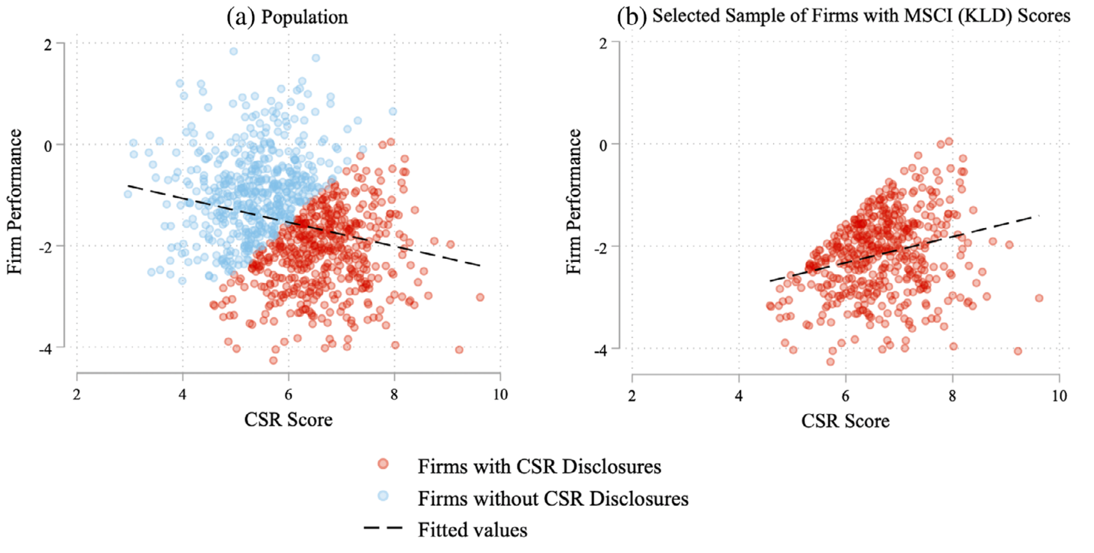


FIGURE 2 Illustration of collider bias in corporate social responsibility (CSR) score on firm performance. Panel (a) is a scatterplot of 1000 simulated observations where the true effect of corporate social responsibility (CSR) score on firm performance is −0.25 (firm performance = −0.25 × CSR + e ~ (0,1)). Panel (b) is the same dataset with only firms that have CSR disclosures.

compensatory strategy. Given these assumptions, the estimated positive effect of CSR on financial performance among the disclosing firms (as seen in Figure 2b) is spurious. It is not a genuine causal effect but rather a consequence of conditioning on a collider—CSR disclosure. This relationship can be graphically illustrated in the DAG in Figure 1c by letting X, Y, and C, respectively, represent CSR activities, financial performance, and CSR disclosure.

2.4 | Collider bias in published work

We next illustrate the problem concretely in the context of strategic human capital. We explore the question of the effect of having a woman CEO on the compensation of other women managers in the TMT of the firm (cf. Dezső et al., 2022). We will return to this application throughout the remainder of this article. In this context, the treatment variable is having a woman CEO, and the outcome is a change in the compensation of other women in the TMT.

This is a challenging question to answer empirically. A naïve researcher might regress all women TMT members' (including the CEO) compensation on whether there was a woman CEO in the focal or previous year. However, if the sample includes the CEO, then the researcher would wrongly attribute the increased compensation caused by a woman TMT member becoming a CEO to the fact that there is a woman CEO. For example, in Figure 3a, the woman CFO in period 1 was promoted to CEO in period 2 and received an increase in compensation with her promotion. However, if we include the compensation of the woman CEO in the analysis, we will wrongly conclude that having a woman CEO increased the compensation of women TMT members by $25,000.

To address this problem, the authors excluded CEOs from their sample and condition the analysis on non-CEO women only, as seen in Figure 3b. This approach reasonably addresses

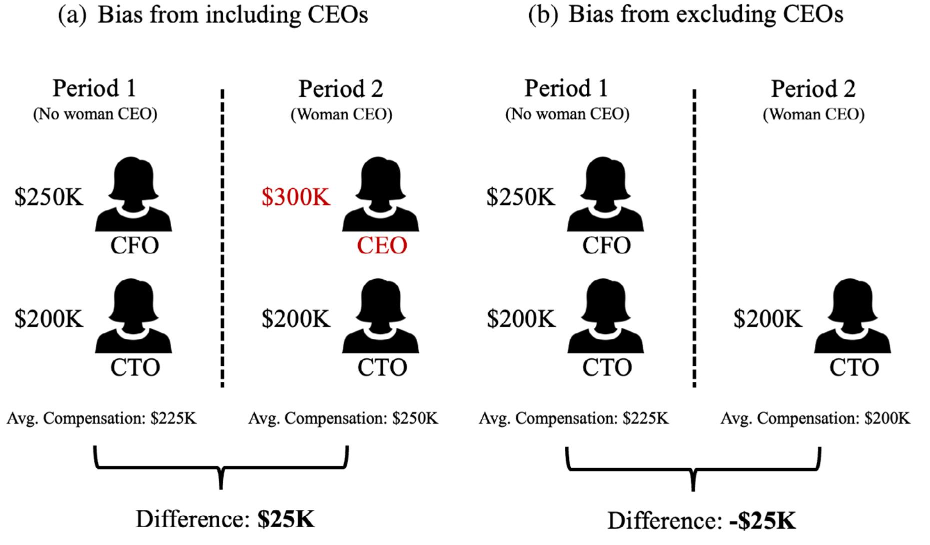


FIGURE 3 Illustrative example of the challenge of estimating the effect of a woman CEO on top management team (TMT) women's compensation.

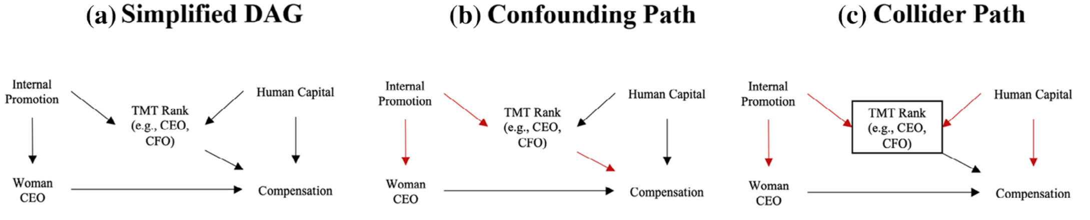


FIGURE 4 Simplified directed acyclic graph (DAG) for the effect of a woman CEO on women's compensation.

the problem of omitted variable bias identified above. However, the sample exclusion criteria (excluding women CEOs) also inadvertently introduces collider bias. Here, we would wrongly estimate that women CEOs cause TMT women to experience a $25,000 *reduction* in their compensation.

We represent a simplified DAG of the assumed causal structure in Figure 4a (online Appendix A contains the full DAG including all steps in its construction). *Internal promotion* of a woman in the TMT causes the TMT member's rank to change (e.g., from CFO to CEO) and also causes there to be a woman CEO. There is also a direct effect of *TMT rank* on *compensation*, since compensation is a function of the person's rank. The variable *internal promotion* refers to the unobserved opportunity structure at a firm by which a focal woman may come to occupy the CEO position, such as the previous CEOs retirement or firing, and the availability of other suitable candidates. Further, a TMT member's *human capital* (which may include experience, social connections, and other factors) affects their *rank* (TMT members with greater human capital are more likely to be CEO) and their *compensation*. The DAG makes it clear that the TMT member's rank exists along two backdoor paths. The first is the path highlighted in Figure 4b, where *internal promotion* is a confounder and *TMT Rank* is a mediator. By conditioning on the mediator *TMT rank*, the authors effectively closed the backdoor path highlighted in Figure 4b. However, in doing so, they open another path, highlighted in Figure 4c. Specifically, because *TMT rank* is a collider between internal promotion and human capital, and because conditioning on a collider opens a path, conditioning on *TMT rank* opens the backdoor path highlighted in Figure 4b. Thus, by conditioning on *TMT rank*, the researchers have closed one biasing path, but inadvertently opened another. Note that in this particular causal model, *TMT rank* is both a collider and a confounder. This highlights the importance for researchers to explicitly state their assumed causal model.

Next, we generalize and formalize the intuition in this example by deriving the bias term using a simple mathematical model. As in Figure 3, suppose there are two time periods $t=1,2$ and two women $i=1,2$ with compensation $Y_{i,t}$. In period 1, both women are on the TMT in non-CEO positions but in period 2, the individual with greater human capital of the two women gets promoted to CEO. Suppose higher human capital manifests itself in higher compensation in period 1 so that $Y_{2,1}=Y_{1,1}+W$, where the compensation premium W is a random variable with $\delta=E(W)>0$.

The researcher uses a stylized regression model with firm fixed effects to account for unobserved firm characteristics (e.g., a women-friendly environment). We capture this in the model for average women's TMT compensation:

$$\overline{Y}_t=\beta_0+\beta_1 1\{\text{woman CEO at time } t\}+U_t+V. \tag{1}$$

The term U_t is a mean-zero idiosyncratic error uncorrelated with all other variables and V is a firm-specific unobserved variable potentially correlated with whether the firm has a woman CEO. To remove V from the equation, the researcher uses fixed effects estimation (which is equivalent to first differences here) to estimate:

$$\Delta\overline{Y}=\beta_1+\Delta U, \tag{2}$$

where $\Delta\overline{Y}=\overline{Y}_2-\overline{Y}_1$, β_0 and V have dropped out because they are time-invariant, and β_1 remains because a woman transitioned into the CEO role. Taking expectations yields

$$\begin{aligned}\beta_1 &= E\left(\Delta\overline{Y}\right)=E\left(Y_{1,2}-\frac{Y_{1,1}+Y_{2,1}}{2}\right)=E\left(Y_{1,2}-\frac{2Y_{1,1}+W}{2}\right)\\ &= \underbrace{E(Y_{1,2}-Y_{1,1})}_{\text{average change in compensation}} - \underbrace{\delta/2}_{\text{collider bias}}.\end{aligned} \tag{3}$$

where we used $W=Y_{2,1}-Y_{1,1}$ and $\delta=E(W)>0$ as defined above (1). In words, the regression coefficient β_1 measures the average change in the compensation of the remaining woman employee minus a bias term that measures the compensation differential between the promoted woman and the remaining woman. If this differential is large, then the bias term will overwhelm even a large *increase* in women's compensation. A regression of average non-CEO women's compensation on indicators for having a woman CEO will have a non-ignorable and systematic downward bias. This bias will not be present in regressions of the effect of having a woman CEO on men's compensation. The bias persists even though the fixed effects estimator successfully eliminated the confounder V. Large negative coefficients are not informative, and at best, lower bounds.

The 2 in the bias term ($\delta/2$) represents the number of employees, so, all else equal, bias can be expected to be smaller if the number of non-CEO women on the TMT is large. Formalizing this statement makes the derivation more complex but the same idea applies. Suppose there are n women at time $t=1$. Compensation packages $Y_{1,t},\ldots,Y_{n-1,t}$ are identically distributed copies of a random variable Y_t. The compensation $Y_{n,t}$ of the n-th person (the future CEO) is larger on average, such that $E(Y_{n,t})=E(Y_t)+\delta$. Let $\overline{Y}_1$ be the average with all observations at time $t=1$, $\overline{Y}_1^*$ be the average with observation n removed at time $t=1$, and $\overline{Y}_2^*$ be the average with observation n removed at time $t=2$. If observation n becomes CEO, the fixed effects regression coefficient now identifies

$$\begin{aligned}\beta_1 &= E\left(\Delta\overline{Y}\right)=E\left(\overline{Y}_2^*-\overline{Y}_1\right)=E\left(\overline{Y}_2^*-\overline{Y}_1^*+\frac{\overline{Y}_1^*-Y_{n,1}}{n}\right)\\ &= \underbrace{E\left(\overline{Y}_2^*-\overline{Y}_1^*\right)}_{\text{average change in compensation for non}-\text{CEOs}}\underbrace{-\delta/n}_{\text{bias}}.\end{aligned} \tag{4}$$

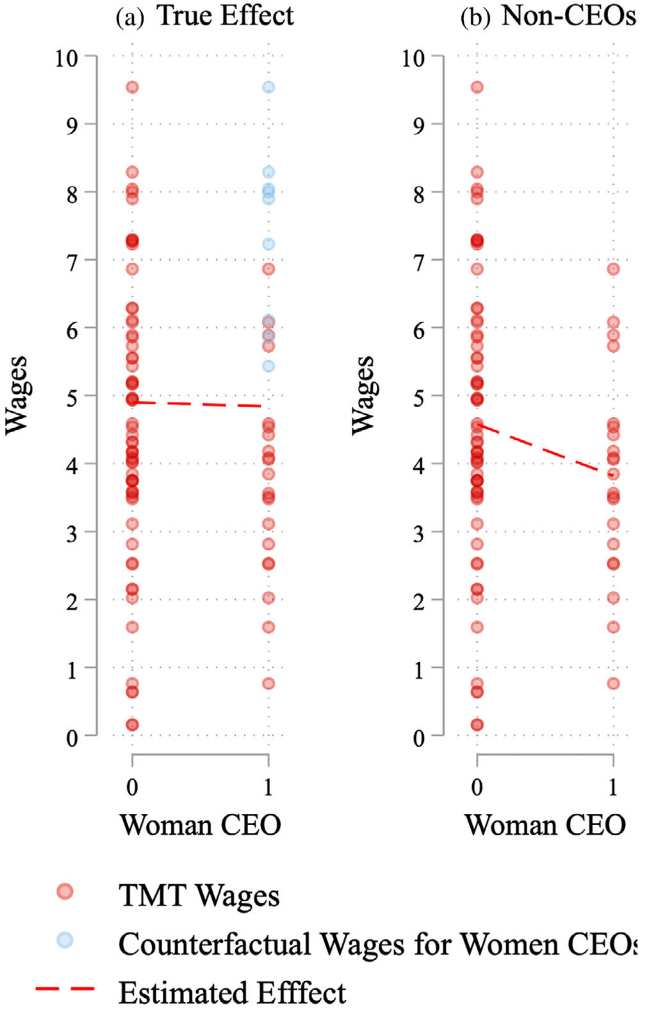


FIGURE 5 Illustration of collider bias using gender of CEO on the compensation of non-CEO women.

The δ/n term is a manifestation of collider bias. We illustrate the mechanics graphically in Figure 5, which plots compensation against whether there is a woman CEO. The red dots represent observations of non-CEO women on the TMT, and the blue dots indicate the counterfactual compensation of the women CEOs had they not been promoted, but still had a (different) woman CEO. We have constructed the data such that there is no causal effect of CEO gender on the compensation of non-CEO women. This is represented by the flat line in Figure 5a. However, Figure 5b demonstrates that the correlation is negative once the women who were promoted to CEO are removed. The fact that we systematically do not observe women who were promoted to CEO in period two acts as the collider.

3 | EMPIRICAL ILLUSTRATION

Until now, we have used thought experiments, simulated data, DAGs, and mathematical models to illustrate the problem of collider bias. In this section, we use archival data to demonstrate how collider bias can result in spurious findings when examining an important question for human capital scholars. Specifically, we produce results that are consistent with findings in Dezső et al. (2022) (hereafter DLR), who find that the presence of a woman CEO (compared to a man CEO) reduces the compensation of other women on the TMT by more than 16%. Based on these findings, the authors argue that having a woman CEO reduces the diversity benefits contributed by other women in the TMT. Relatedly, Corwin et al. (2022) (hereafter CLR) find that, compared to a man CEO, the presence of a woman CEO reduces the number of other women in the TMT by 7% in the subsequent year. Based on these results, the authors argue that women CEOs may actively exert pressure to resist the advancement of other women in the company (the so-called "queen bee" effect). These are not inconsequential findings, especially if policymakers use them as the basis for decision-making. For example, one possible implication of these findings is that if boards want to reduce the gender pay gap or the gender diversity of their TMT, they may hesitate appointing women to the CEO position.

We investigate these findings in the remainder of this section from the perspective of collider bias. To summarize, we use standard OLS regressions with separate firm and year fixed effects to estimate the effects of women CEOs on other women managers. Our results are similar in magnitude and precision to those reported by DLR and CLR.

While in line with the findings in extant work, we aim to demonstrate that the observed correlations we report are unrelated to the gender of the CEO or other managers, and can be explained by collider bias. Again, CEO rank is the collider, and by conditioning the sample on individuals who are not CEOs, these analyses are threatened by collider bias. To substantiate our claim, we use placebo regressions to show that various CEO characteristics unrelated to gender produce comparable effects on the outcomes of non-CEOs sharing the same characteristic. Later we will show that excluding observations from individuals who transition between non-CEO and CEO positions eliminates the effects across the board.

3.1 | Data and sample

We use 30 years of data (1992–2021) from the Compustat and ExecuComp datasets to examine the effect of having a woman or minority CEO on two important outcomes: (1) the compensation of other managers in the TMT (cf. DLR) and (2) the proportion of other women in the

TMT (cf. CLR). We select all firms and employees with valid compensation data in the ExecuComp dataset, which collects information on the highest paid employees in S&P 1500 firms and various other firms. We construct the TMT as consisting of the CEO and any additional employees reported, who typically hold titles such as COO, CFO, Executive VP, and so forth. Our final sample consists of 298,975 observations for 55,077 executives working at 3960 firms. The ExecuComp dataset includes a variable indicating whether an executive is a man or a woman. Observations for women employees constitute 7.40% of the total sample.[4]

3.2 | Placebo groups

We create placebo groups that are unrelated to the gender dynamics advanced by DLR and CLR. Thus, if placebo regressions exhibit similar correlations to those found for women, we can reasonably infer that the correlations are not driven by the authors' proposed mechanisms, but rather by some other factor.

We compare the results for women with identical regressions for three placebo groups of employees that appear in the dataset at comparable rates: employees whose first name is "John" (5.00% of the total sample), employees whose first name starts with the letter "M" (9.00% of the sample), and a group of randomly selected employees (7.9% of the sample).

Women in our sample are largely nonoverlapping with the placebo groups: The "John" group contains 0% women observations, the "Letter M" group contains 6.5% women observations, and the "Random" group contains 6.9% women observations. Table 3 describes the size of the various groups used in the analyses.

3.3 | Variable definitions

3.3.1 | Dependent variables: Manager compensation and TMT representation

We followed DLR in constructing a dependent variable, *Top Manager Compensation*, as the natural log transformation of a top manager's total compensation, including salary, bonus, and grants of stock and options. We followed CLR in constructing another dependent variable, *% Women in TMT*, as the percentage of women TMT members (excluding the CEO) at time $t + 1$. A value of 0 indicates that there are no women on the TMT, while a measure of 1 indicates that every member of the TMT is a woman. We construct parallel measures for *% "John" in TMT*, *% "Letter M" in TMT*, and *% Random Group in TMT*.

3.3.2 | Independent variables: CEO Types

Has CEO of Type X. We directly followed DLR and CLR in the construction of the independent variable *Woman CEO*, which takes the value 1 if, in a given year, a firm has a woman CEO and 0 otherwise. We construct parallel measures for *John CEO*, *Letter M CEO* and *Random Group*

[4]In online Appendix B, we replicate the main results from DLR and CLR across various specifications, noting differences in sampling and estimation. In online Appendix D, we provide Stata code to replicate the main DLR results.

TABLE 3 Frequency of women and placebo groups in sample.

	Observations	Individuals	Number of firms
Employee is a woman	22,166	5309	2514
	7.4%	**9.6%**	**63.5%**
Employee first name is "John"	14,813	2680	1931
	5.0%	**4.9%**	**48.8%**
Employee name starts with "M"	26,893	5003	2712
	9.0%	**9.1%**	**68.5%**
Employee in random group	23,493	4374	2600
	7.9%	**7.9%**	**65.7%**

Note: Percentage of total in bold, gray background.

CEO to indicate that the firm in that year is led by a CEO whose name is John, a CEO whose name starts with the letter M or a CEO from the randomly assigned group.

3.3.3 | Control variables

We followed DLR in the selection and measurement of all firm-level covariates and two manager characteristics as explained below.

Advertising intensity is the log transformation of the ratio of advertising expense to assets. *Firm age* is the log transformation of the firm's age in years, measured as the difference between the current year and the earlier of the firm's first year in Compustat or initial public trading date. *Leverage* is the ratio of debt to the market value of a firm's assets. *R&D intensity* is the log transformation of the ratio of R&D expense to assets. *Size from assets* is the log transformation of the lagged book value of a firm's assets. *Size from employees* is the log transformation of the lagged size of a firm's workforce. *Tobin's q* is the log transformation of the lagged ratio of the market value of a firm's assets to their replacement value. *Manager age* is the log transformation of the manager's age. *Employee is CFO* is a dummy variable indicating that the employee has the title of Chief Financial Officer, a variable that is readily available in the ExecuComp dataset. Table 4 contains the descriptive statistics and zero order correlations for the variables in our analyses.

3.4 | Results

3.4.1 | The effect of women CEOs on the compensation of other top managers

The data structure for this analysis is defined at the level of a firm, year and individual executive. We use separate fixed effects for firms and years to estimate the coefficient of having a woman CEO on the total compensation of other women on the TMT. As in DLR, we run the analysis in the unpooled sample corresponding to the group sharing the same characteristic as the CEO.[5]

[5]DLR also show a null effect for having a woman CEO on the compensation of men. These results are unproblematic with respect to collider bias. This is because in the ExecuComp dataset non-CEO men do not exit the sample to become women CEOs.

TABLE 4 Summary statistics and correlations ($N = 298{,}975$).

		Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1.	Total compensation	7.19	1.14													
2.	CEO is a woman	0.03	0.17	0.03												
3.	CEO's first name is John	0.05	0.22	0.01	−0.04											
4.	CEO's name starts with M	0.08	0.27	0.03	0.05	−0.07										
5.	CEO belongs to random group	0.07	0.26	0.02	−0.01	0.02	−0.02									
6.	Advertising Intensity	0.01	0.04	0.01	0.05	−0.02	0.04	−0.00								
7.	Firm age	3.03	0.77	0.20	0.02	0.02	0.00	0.01	−0.07							
8.	Book leverage	0.25	0.45	0.04	−0.00	−0.00	0.00	0.01	−0.01	0.05						
9.	R&D intensity	0.03	0.07	−0.04	−0.01	0.00	−0.00	−0.01	0.06	−0.18	−0.06					
10.	Size from assets	7.56	1.92	0.52	0.01	0.03	0.01	0.02	−0.11	0.43	0.08	−0.31				
11.	Size from employees	1.89	1.30	0.41	0.02	0.02	0.01	0.01	0.05	0.38	0.05	−0.21	0.59			
12.	Tobin's Q	0.99	0.37	0.08	0.00	0.01	0.01	0.00	0.18	−0.15	−0.02	0.41	−0.27	−0.05		
13.	Executive's age	3.97	0.14	0.17	0.00	0.01	−0.01	0.00	−0.05	0.20	0.01	−0.07	0.16	0.11	−0.06	
14.	Employee is CFO	0.11	0.31	0.03	0.03	−0.01	0.02	0.00	−0.00	0.04	0.01	−0.01	0.04	−0.02	−0.01	−0.05

Note: These correlations correspond to the executive-firm-year level data used in the first analysis. Correlations larger than $|0.005|$ are statistically significant at $p < .05$.

Model 1 in Table 5 shows a negative coefficient of −0.131, which corresponds to a 12.3% decrease in the compensation of other women in the TMT from having a woman CEO compared to a man CEO. This effect is similar in size to the estimated effect reported by DLR (cf. Table 5, Model 2 in DLR). These regressions do not, however, demonstrate that collider bias is driving the effects. It is possible that women CEOs do have real negative effects on other women TMT members. To explore whether this is the case, we turn to placebo regressions in Models 2–4.

The results are quite similar for the placebo groups. Individuals named "John" (Model 2) appear to suffer an 18% decrease in their compensation when their CEO is also named John, compared to having a CEO with another name. In Models 3 and 4 of Table 5 we also find large, and precisely estimated negative effects on the compensation of executives whose name starts with the letter "M" (−10%) or who belong to a random group (−15%), from having a CEO sharing that specific trait. Unless we are to believe that there are causal mechanisms causing CEOs named John to reduce the compensation of other TMT members named John, the results cast doubt on the causal explanation that women CEOs hurt the compensation of other women executives.

TABLE 5 CEO effect on the compensation of other top managers.

Sample restricted to	**Other women in TMT Model 1**	**Other "Johns" in TMT Model 2**	**Other "starts with M" in TMT Model 3**	**Other "random" in TMT Model 4**
Advertising intensity	0.097 (.806)	0.496 (.010)	−0.352 (.398)	−0.229 (.635)
Firm age	−0.045 (.355)	−0.135 (.058)	−0.036 (.455)	−0.153 (.001)
Leverage	−0.112 (.036)	−0.223 (.004)	−0.146 (.001)	−0.148 (.017)
R&D intensity	−0.421 (.066)	−0.498 (.014)	−0.527 (.090)	−0.616 (.011)
Size from assets	0.120 (.000)	0.106 (.000)	0.124 (.000)	0.127 (.000)
Size from employees	0.133 (.000)	0.171 (.000)	0.144 (.000)	0.140 (.000)
Tobin's Q	0.319 (.000)	0.408 (.000)	0.369 (.000)	0.345 (.000)
Executive's age	0.425 (.000)	0.654 (.000)	0.417 (.000)	0.334 (.001)
Employee is CFO	0.119 (.000)	0.180 (.000)	0.104 (.000)	0.131 (.000)
CEO is a woman	−0.131 (.001)			
CEO's first name is John		−0.198 (.002)		
CEO's name starts with M			−0.110 (.002)	
CEO belongs to random group				−0.161 (.000)
Constant	2.996 (.000)	2.607 (.000)	3.271 (.000)	3.793 (.000)
Observations	20,240	11,743	22,037	19,050
Adj. R-squared	.707	.707	.657	.670

Note: All models include year and firm fixed effects. p-Values are in parentheses. Robust standard errors clustered by firm.

3.4.2 | The effect of woman CEOs on TMT gender diversity

Next, we follow CLR and collapse the data into firm-year observations to analyze the effect of having a woman CEO on the proportion of other women on the TMT in the subsequent year. Table 6 displays the coefficients for the analyses on women and on the three placebo groups.

The magnitude of the effect of a woman CEO on the proportion of women on the TMT in the subsequent period is −3.1%, about half of the negative effect reported by CLR. The effect of women CEOs on gender representation is very similar to the effect of CEOs whose name starts with the letter M on the proportion of employees whose name also starts with the letter M (−3.8%, see Model 3 in Table 6). The effect of a CEO named "John" (Model 2) and of a CEO from a randomly assigned group (Model 4) is also negative and precisely estimated on the proportion of employees named "John" (−2.0%), and on employees from the same randomly selected group (−2.0%). Across all groups, the overall pattern produced by collider bias is consistent. Unless we are to believe that there are causal mechanisms causing CEOs named John to reduce the number of other TMT members named John, the results again cast doubt on the causal explanation that women CEOs erode the gender diversity of TMTs.

TABLE 6 CEO effect on TMT representation.

Dependent variable is	% women in TMT Model 1	% "John" in TMT Model 2	% "starts with M" in TMT Model 3	% random in TMT Model 4
Advertising intensity	−0.002 (.970)	−0.007 (.831)	0.044 (.473)	−0.074 (.073)
Firm age	−0.024 (.000)	−0.000 (.926)	−0.011 (.086)	−0.003 (.586)
Leverage	0.001 (.606)	−0.001 (.170)	−0.001 (.377)	0.001 (.563)
R&D intensity	−0.005 (.871)	−0.013 (.544)	0.005 (.856)	−0.045 (.067)
Size from assets	0.000 (.817)	−0.002 (.167)	−0.001 (.514)	0.001 (.328)
Size from employees	−0.002 (.489)	0.001 (.721)	0.001 (.723)	0.002 (.540)
Tobin's Q	0.002 (.619)	−0.001 (.813)	0.004 (.376)	−0.001 (.718)
CEO is a woman	−0.031 (.007)			
CEO's first name is John		−0.020 (.001)		
CEO's name starts with M			−0.038 (.000)	
CEO belongs to random group				−0.020 (.000)
Constant	0.070 (.000)	0.078 (.000)	0.080 (.000)	0.085 (.000)
Observations	49,252	49,252	49,252	49,252
Adj. *R*-squared	.462	.393	.406	.403

Note: All models include year and firm fixed effects. *p*-Values are in parentheses. Robust standard errors clustered by firm.

3.5 | Correcting collider bias

3.5.1 | Excluding individuals who will become CEOs in future periods

The solutions to collider bias vary by context. The simplest solution is to avoid conditioning on colliders. If we return to the simplified DAG in Figure 4, it is possible to develop an identification strategy. As we explained before, conditioning on TMT rank opens a backdoor path because TMT rank is a collider. However, conditioning on internal promotions closes the backdoor path identified in Figure 4b without opening the backdoor path highlighted in Figure 4c. Thus, the DAG makes it clear that one solution is to condition only on *internal promotion*. For our empirical example, this can be done by excluding (selecting out) all women who were ever *internally promoted to CEO*. Another effective approach is to condition on TMT rank *and* include individual-level fixed effects, which effectively closes the backdoor path highlighted in Figure 4c by conditioning on the TMT member's human capital (assuming human capital is relatively stable over the sample period). As shown in online Appendix C, individual level fixed effects greatly reduce the contribution to the bias of women who are internally promoted to CEO to the estimated coefficient. In Tables 7 and 8, we compare the coefficients and *p*-values previously reported in Tables 5 and 6 with those of identical regressions run on samples that exclude individuals who will become CEOs in future periods.

Table 7 illustrates the dramatic changes in non-CEO compensation resulting from removing a small percentage of observations that transition from the non-CEO sample to the (excluded) CEO sample. The large and precisely estimated negative coefficients for women and the three placebo groups all become very small and statistically indistinguishable from zero.

TABLE 7 CEO effect on manager compensation with and without collider bias.

Dependent variable includes	Women	First name is "John"	Name starts with "M"	In random group
All TMT members (*p*-value)	−0.131 (.001)	−0.198 (.002)	−0.110 (.002)	−0.161 (.000)
TMT members who never become CEO (*p*-value)	0.012 (.744)	0.030 (.608)	0.012 (.721)	0.040 (.307)
Percent obs. excluded within category (No. obs. excluded)	4.9% (982)	12.0% (1414)	11.1% (2444)	10.9% (2068)

Note: All models include year and firm fixed effects as well as the controls as those in Table 5. *p*-Values are in parentheses. Robust standard errors clustered by firm.

TABLE 8 CEO effect on TMT representation with and without collider bias.

Dependent variable includes	Women	First name is "John"	Name starts with "M"	In random group
All TMT members (*p*-value)	−0.031 (.007)	−0.020 (.001)	−0.038 (.000)	−0.020 (.000)
TMT members who never become CEO (*p*-value)	0.012 (.256)	0.007 (.146)	−0.001 (.877)	0.004 (.416)

Note: All models include year and firm fixed effects as well as the controls as those in Table 6. *p*-Values are in parentheses. Robust standard errors clustered by firm.

Similarly, Table 8 illustrates the changes in TMT representation resulting from removing a small percentage of observations that transition from the TMT sample to the (excluded) CEO sample. The coefficients become smaller and statistically indistinguishable from zero across all groups.

Including individual fixed effects or removing "transitioning" observations both eliminate collider bias in this application. However, this does not mean that the resulting estimates are unbiased for the entire population of TMT members. By excluding women who ever become CEO from the sample, the estimates are only valid for executives who never transition from the non-CEO to the CEO pool, rather than all executives.

This empirical example demonstrates how conditioning on colliders can severely bias empirical analysis, resulting in potentially spurious findings. We can reasonably assume that the true effect in the three placebo groups ("John", "First Letter M" and "Random") is zero. A natural question then is what can explain the wide range in our spurious findings, from −0.198 to −0.110 in Table 7 and from −0.038 to −0.020 in Table 8. In online Appendix E, we use Monte Carlo analysis to explore the conditions under which the bias is more or less severe.

4 | PRACTICAL ADVICE FOR AVOIDING COLLIDER BIAS

4.1 | Using directed DAGs

Because of the fundamental problem of causal inference—that we can never observe what would have happened to the treated unit had they not been treated—researchers can never know whether their specifications are biased. However, a well-articulated causal model can lead to a specification that delivers unbiased estimates, assuming the model itself is accurate. To clearly represent and articulate this assumed causal model, we advocate that researcher map their assumed causal model using DAGs. DAGs are particularly beneficial in observational studies, where the lack of experimental control can complicate causal inference. DAGs also help researchers identify the minimal set of variables that must be controlled for to obtain an unbiased estimate of a causal relationship between a particular independent and dependent variable (assuming the DAG is accurate). Moreover, DAGs offer a systematic method for sensitivity analysis, allowing researchers to explore alternative causal pathways and evaluate the robustness of their findings. For instance, if an omitted variable is a concern, adding it to the DAG helps researchers assess its potential impact on the causal estimates. Finally, the explicit articulation of the causal model that the researcher has in mind enhances the transparency, rigor, and credibility of empirical research.

While a comprehensive discussion of DAGs is outside the scope of this article, we refer interested readers to Chap 6–9 of Huntington-Klein (2022) and Huenermund et al. (2022). Briefly, researchers should start by identifying the treatment, outcome, and other relevant variables that cause the treatment and outcome variables. Then, they should connect these nodes with directed arrows to signify the causal links, ensuring that the arrows flow from cause to effect and do not form feedback loops. After constructing the DAG, they should examine it to identify colliders, which are variables affected by two or more other variables, identifiable by incoming arrows from multiple sources. See online Appendix A for a detailed primer and a demonstration of how we created the simplified DAGs in Figure 4.

4.2 | Approaches to addressing collider bias

In the following section, we present several tools that may assist researchers in ensuring the integrity of their empirical results. While there are many potential strategies to address collider bias in observational studies, no single solution is universally applicable. Rather, the right approach depends on the study design and data available. In what follows, we present a series of practical approaches and solutions to collider bias. Several of these solutions are illustrated via Monte Carlo simulations in Figure E4 in the online Appendix.

4.2.1 | Avoiding "bad controls"

After understanding the causal relationships and potential colliders using the DAG, researchers should avoid conditioning on potential colliders. In regression analysis, this means excluding controls that might be colliders (Cinelli et al., 2022). As a general heuristic, we recommend avoiding controlling for any potential intermediate outcomes, or variables that were measured after treatment.

4.2.2 | Fixed effects and subsampling

While the solution to "bad controls" is to not control for them, the solutions to selection on colliders are less straightforward. This is because, in many cases, a sample that is not selected on a collider is not available. In this case, there are several potential solutions. One potential approach is to subsample on observations that do not suffer from collider bias. In our case, this would involve removing all observations for individuals who ever became CEO or including individual fixed effects, which are roughly equivalent in our setting. In our setting, either of these approaches would help the researcher recover an unbiased estimate of the true causal effect. While this is a reasonable solution in our case, it may not generalize to all cases of collider bias. This is because collider bias often does not only affect a clearly defined set of units (e.g., women who ever become CEO), but may affect all or most units in the sample. This approach can also cause selection bias if the resulting sample is no longer representative of the population.

4.2.3 | Inverse probability weighting

An alternative approach for addressing collider bias is through IPW. This method involves weighting observations according to their likelihood of being included in the sample. The objective of this weighting is to balance the representation of units that may be overrepresented or underrepresented as a result of conditioning on the collider. In practice, these weights signify the probability of different units being selected into the sample based on their observable characteristics. For instance, in an empirical example involving CEOs, their selection into the sample might depend on their human capital (as illustrated by the DAG in Figure 4). If a researcher has a proxy measure for human capital, they could use IPW to estimate and adjust for the likelihood of a TMT member being promoted to CEO. Accordingly, we weighted each observation according to the individual's human capital to correct for their probability of being promoted.

As demonstrated in Figure E4 in the online appendix, this approach also recovers the true causal effect. Breen and Ermisch (2014) demonstrate that IPW can recover unbiased estimates in the case where selection is a function of the outcome variable only. In other cases, IPW can reduce the bias, sometimes to negligible levels, if certain conditions are met (Griffith et al., 2020).

4.2.4 | Multiple imputation

Multiple imputation is another statistical technique that can be employed to address collider bias when the collider is related to missing data. The method involves generating multiple complete datasets by imputing missing values using a suitable model that accounts for the relationships between variables. Each of these completed datasets is then analyzed independently, and the results are combined to produce a single, pooled estimate. By accounting for the uncertainty associated with the imputed values, multiple imputation mitigates the bias introduced by conditioning on the collider while preserving the relationships between the exposure, outcome, and any confounders. In the case of our empirical example, we imputed the compensation of women in the years that they were CEOs based on their human capital. As seen in Figure E4 of the Appendix, this approach recovers unbiased estimates of the causal effect. It is crucial to note that the effectiveness of multiple imputation in addressing collider bias hinges on the proper specification of the imputation model, which again can be clarified by using a DAG.

4.2.5 | Placebo analysis

In the case where the above approaches are not feasible, researchers may use placebo analysis to explore the likelihood of collider bias in their setting. While this approach cannot necessarily rule out collider bias, it can be useful in identifying cases where collider bias is likely present. To implement a placebo test, the researcher should choose a variable that is unrelated to both the treatment and outcome. Then they include the placebo in the regression alongside the original treatment variable. If the regression estimates a significant effect of the placebo on the outcome when conditioned on the collider, collider bias may be present. This is because any observed relationship between the placebo variable and the outcome is likely due to the bias introduced by conditioning on the collider. To further evaluate the presence of collider bias, the researcher could compare the results of the analysis with and without conditioning on the collider (if possible). A significant difference in the estimates for the exposure or the placebo variable may support the presence of collider bias.

5 | DISCUSSION AND CONCLUSION

Collider bias is a pervasive problem in social science research. Unlike confounders, however, very little scholarship in management and strategy discusses the threats that collider bias presents to the validity of empirical findings. We describe the problem in general terms, provide relevant examples where collider bias may be present, and demonstrate that important research questions examined in recently published research may suffer from collider bias.

Our study on the relationship between the outcomes of women non-CEOs from having a CEO who is a woman provides a clear example of how collider bias can result in spurious findings. Specifically, we obtain large and precisely estimated negative coefficients on the effect of having a woman CEO on the compensation and representation of other women in the TMTs of firms in the ExecuComp dataset. Our results for women CEOs are comparable in terms of magnitude and statistical significance to recent work. We provide evidence that collider bias—rather than the gender of individuals—drives our results. Without awareness of the collider bias problem, we would wrongly conclude that a woman CEO reduces the compensation and prevalence of other women in the TMT.

Recent replication efforts in strategic management have tested and redefined the robustness and scope conditions of previously published claims (cf. Bettis et al., 2016). Replication efforts in strategic management typically end up restricting a claim's scope conditions by extending the sampling and analytical strategy in various directions, including a longer sampling period (e.g., Howard et al., 2016), a broader population (e.g., Kalnins, 2016), or more up-to-date model specifications (e.g., Park et al., 2016). Attention to collider bias can complement these efforts by uncovering potentially misleading findings that are surprisingly robust to statistical replication. In online Appendix B, we replicate the main result for women executives from DLR and CLR after extending the sample to include additional years of data, a larger pool of executives, and after removing all time-varying controls. In the main paper, we also replicate the main pattern of results even after changing the focal population from women executives to executives named John, executives whose last name starts with the letter "M," and a group of randomly selected executives.

We offer several potential remediating approaches. First, scholars can use common statistical adjustments, like fixed effects at the appropriate level, IPW, or multiple imputation to reduce collider bias. These solutions may not be a silver bullet. For instance, having a future CEO in the TMT may have influenced the composition of the TMT or the trajectory of the company in ways that cannot be controlled for by removing CEOs from the sample. Removing CEO observations for individuals who will become CEOs in future periods also changes the composition of the sample such that, at best, the effects recovered from having a woman CEO apply only to TMT members who never become CEOs. While removing collider bias may be difficult, detecting its presence is much simpler. At a minimum, scholars can use DAGs, placebo regressions, and simple Monte Carlo simulations to understand the potential threat of collider bias. For example, researchers studying the presence of women in TMTs may compare the observed distribution across firms with simulated distributions (e.g., Dezső et al., 2016). An examination of the effect of a woman CEO on broader within-firm gender equality may require obtaining data on gender disparities among employees below the TMT. Deepening the sampling pool would reduce the effect of collider bias caused by attrition due to promotion to CEO.

Strategy and management research has made great strides in increasing empirical rigor within the field (Ethiraj et al., 2016; Ethiraj et al., 2017; Hamilton & Nickerson, 2003; Quigley et al., 2023; Shaver, 1998; Wolfolds & Siegel, 2019). We hope to contribute to this agenda by highlighting the threat of collider bias to the validity of empirical results. While some other empirical problems, like classical measurement error on the independent variable, merely attenuate effects, collider bias can result in correlations with the wrong sign that can support claims of counterintuitive, and thereby publishable findings. This makes awareness of collider bias particularly important. We hope that by bringing a broader awareness of collider bias, and by providing approaches to mitigate its effects, we can increase the validity and usefulness of our findings for informing theory and practice.

ACKNOWLEDGMENTS

The authors thank Evan Starr and Jordan Siegel for their helpful comments on earlier drafts, and Jaewoo Lee for excellent research assistance. The authors would also like to thank Co-Editor Mary Benner and three anonymous reviewers for their constructive suggestions.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://osf.io/qvgtz/.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

REFERENCES

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Awaysheh, A., Heron, R. A., Perry, T., & Wilson, J. I. (2020). On the relation between corporate social responsibility and financial performance. *Strategic Management Journal*, *41*(6), 965–987.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, *6*(3), 285–327.

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication. *Strategic Management Journal*, *37*(11), 2193–2203.

Boyd, B. K., Gove, S., & Hitt, M. A. (2005). Consequences of measurement problems in strategic management research: The case of Amihud and lev. *Strategic Management Journal*, *26*(4), 367–375.

Breen, R., & Ermisch, J. (2014). Using inverse probability weighting to address post-outcome collider bias. *Sociological Methods & Research*, *53*(1), 5–27.

Busenbark, J. R., Yoon, H., Gamache, D. L., & Withers, M. C. (2022). Omitted variable bias: Examining management research with the impact threshold of a confounding variable (ITCV). *Journal of Management*, *48*(1), 17–48.

Certo, S. T., Busenbark, J. R., Woo, H. S., & Semadeni, M. (2016). Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal*, *37*(13), 2639–2657.

Chang, E. H., Milkman, K. L., Chugh, D., & Akinola, M. (2019). Diversity thresholds: How social norms, visibility, and scrutiny relate to group composition. *Academy of Management Journal*, *62*, 144–171.

Cinelli, C., Forney, A., & Pearl, J. (2022). A crash course in good and bad controls. *Sociological Methods & Research*, 00491241221099552.

Cook, A., & Glass, C. (2014). Above the glass ceiling: When are women and racial/ethnic minorities promoted to CEO? *Strategic Management Journal*, *35*, 1080–1089.

Corwin, E. S., Loncarich, H., & Ridge, J. W. (2022). What's it like inside the hive? Managerial discretion drives TMT gender diversity of women-led firms. *Journal of Management*, *48*(4), 1003–1034.

Davis, G. F. (2015). Editorial essay: What is organizational research for? *Administrative Science Quarterly*, *60*(2), 179–188.

Derks, B., Van Laar, C., & Ellemers, N. (2016). The queen bee phenomenon: Why women leaders distance themselves from junior women. *Leadership Quarterly*, *27*, 456–469.

Dezső, C. L., Li, Y., & Ross, D. G. (2022). Female CEOs and the compensation of other top managers. *Journal of Applied Psychology*, *107*(12), 2306–2318.

Dezső, C. L., Ross, D. G., & Uribe, J. (2016). Is there an implicit quota on women in top management? A large-sample statistical analysis. *Strategic Management Journal*, *371*, 98–115.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*(1), 31–53.

Ethiraj, S. K., Gambardella, A., & Helfat, C. E. (2016). Replication in strategic management. *Strategic Management Journal*, *37*(11), 2191–2192.

Ethiraj, S. K., Gambardella, A., & Helfat, C. E. (2017). Improving data availability: A new SMJ initiative. *Strategic Management Journal*, *38*(11), 2145–2146.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*(4), 210–233.

Friedman, M. (1970). The corporate social responsibility is to increase its profits. *New York Times Magazine*.

Ge, C., Huang, K. W., & Png, I. P. (2016). Engineer/scientist careers: Patents, online profiles, and misclassification bias. *Strategic Management Journal*, *37*(1), 232–253.

Griffith, G. J., Morris, T. T., Tudball, M. J., Herbert, A., Mancano, G., Pike, L., Sharp, G. C., Sterne, J., Palmer, T. M., Davey Smith, G., & Tilling, K. (2020). Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nature Communications*, *11*(1), 5749.

Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic Organization*, *1*(1), 51–78.

Hill, A. D., Johnson, S. G., Greco, L. M., O'Boyle, E. H., & Walter, S. L. (2021). Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management*, *47*(1), 105–143.

Hoisl, K. (2007). Tracing mobile inventors—The causality between inventor mobility and inventor productivity. *Research Policy*, *36*(5), 619–636.

Hornstein, A. S., & Zhao, M. (2018). Reaching through the fog: Institutional environment and crossborder giving of corporate foundations. *Strategic Management Journal*, *39*(10), 2666–2690.

Howard, M. D., Withers, M. C., Carnes, C. M., & Hillman, A. J. (2016). Friends or strangers? It all depends on context: A replication and extension of Beckman, Haunschild, and Phillips (2004). *Strategic Management Journal*, *37*(11), 2222–2234.

Huenermund, P., Louw, B., & Rönkkö, M. (2022). The choice of control variables: How causal graphs can inform the decision. In Academy of Management Proceedings (Vol. 2022, No. 1, p. 15534). Briarcliff Manor, NY 10510: Academy of Management.

Huntington-Klein, N. (2022). *The effect: An introduction to research design and causality*. Chapman and Hall/CRC.

Jeong, S. H., & Harrison, D. A. (2017). Glass breaking, strategy making, and value creating: Meta-analytic outcomes of women as CEOs and TMT members. *Academy of Management Journal*, *60*(4), 1219–1252.

Jeong, S. H., Mooney, A., Zhang, Y., & Quigley, T. J. (2021). How do investors really react to the appointment of Black CEOs? A comment on Gligor et al. *Strategic Management Journal*, *44*(7), 1733–1752.

Kaiser, U., Kongsted, H. C., & Rønde, T. (2015). Does the mobility of R&D labor increase innovation? *Journal of Economic Behavior & Organization*, *110*, 91–105.

Kalnins, A. (2016). Beyond Manhattan: Localized competition and organizational failure in urban hotel markets throughout the United States, 2000–2014. *Strategic Management Journal*, *37*(11), 2235–2253.

Kalnins, A. (2018). Multicollinearity: How common factors cause type 1 errors in multivariate regression. *Strategic Management Journal*, *39*(8), 2362–2385.

McDonald, M. L., Keeves, G. D., & Westphal, J. D. (2018). One step forward, one step back: White male top manager organizational identification and helping behavior toward other executives following the appointment of a female or racial minority CEO. *Academy of Management Journal*, *61*(2), 405–439.

McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification?. *Strategic Management Journal*, *21*(5), 603–609.

Melero, E., Palomeras, N., & Wehrheim, D. (2020). The effect of patent protection on inventor mobility. *Management Science*, *66*(12), 5485–5504.

Miller, D., Le Breton-Miller, I., & Lester, R. H. (2010). Family ownership and acquisition behavior in publicly-traded companies. *Strategic Management Journal*, *31*(2), 201–223.

Palomeras, N., & Melero, E. (2010). Markets for inventors: Learning-by-hiring as a driver of mobility. *Management Science*, *56*(5), 881–895.

Park, U. D., Borah, A., & Kotha, S. (2016). Signaling revisited: The use of signals in the market for IPO s. *Strategic Management Journal*, *37*(11), 2362–2377.
Pearl, J. (2000). *Causality*. Cambridge University Press.
Quigley, T. J., Hill, A. D., Blake, A., & Petrenko, O. (2023). Improving our field through code and data sharing. *Journal of Management*, *49*(3), 875–880.
Schneider, E. B. (2020). Collider bias in economic history research. *Explorations in Economic History*, *78*(1), 101356.
Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science*, *44*(4), 571–585.
Shaver, J. M. (2020). Causal identification through a cumulative body of research in the study of strategy and organizations. *Journal of Management*, *46*(7), 1244–1256.
Shaver, J. M. (2021). Evolution of quantitative research methods in strategic management. In I. M. Duhaime, M. A. Hitt, & M. A. Lyles (Eds.), *Strategic management: State of the field and its future* (pp. 83–97). Oxford University Press.
Stern, I., Deng, X., Chen, G., & Gao, H. (2021). The "butterfly effect" in strategic human capital: Mitigating the endogeneity concern about the relationship between turnover and performance. *Strategic Management Journal*, *42*(13), 2493–2510.
Wolfolds, S. E., & Siegel, J. (2019). Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument. *Strategic Management Journal*, *40*(3), 432–462.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Frake, J., Hagemann, A., & Uribe, J. (2024). Collider bias in strategy and management research: An illustration using women CEO's effect on other women's career outcomes. *Strategic Management Journal*, *45*(7), 1393–1419. https://doi.org/10.1002/smj.3588