Measuring the Incidence of Earnings Manipulations: A Novel Approach*

Nicole L. Cade†
Assistant Professor
University of Pittsburgh
ncade@pitt.edu

Joshua L. Gunn
Assistant Professor
University of Pittsburgh
jlgunn@pitt.edu

Alex Vandenberg
Doctoral Candidate
University of Pittsburgh
alex.vandenberg@pitt.edu

October 21, 2021

* We appreciate useful comments from Darren Bernard, Ed deHaan, Mei Feng, Shannon Garavaglia, Weili Ge, Vicky Hoffman, Christian Leuz (editor), Dawn Matsumoto, Don Moser, and an anonymous reviewer at Journal of Accounting Research.
† Corresponding author.
Measuring the Incidence of Earnings Manipulations: A Novel Approach

ABSTRACT

In this paper, we conduct a “list experiment” with top-level executives to estimate the prevalence of several forms of earnings manipulation. A list experiment is an indirect questioning method in which participants respond to questions about potentially socially undesirable behaviors under the privacy protection of plausible deniability. Because we calculate our estimates from executives’ reports of their own experiences under the highest level of privacy, we complement previous estimates calculated using archival, structural, and survey methods. To test the efficacy of the list experiment method and to provide estimates of executives’ level of dishonesty in our context, we also directly ask a separate sample of executives about their experiences with manipulating earnings. Our findings inform researchers interested in understanding and measuring managers’ strategic reporting behaviors as well as practitioners, investors, and regulators who rely on financial reporting risk assessments to allocate limited resources.
1. Introduction

Earnings manipulations are non-neutral acts in which managers, within or outside the bounds of Generally Accepted Accounting Principles (GAAP), intentionally intervene in the financial reporting process to obtain a private gain (Schipper [1989], Dechow, Sloan, and Sweeney [1996]). Earnings manipulations can take a variety of forms, ranging from less egregious interventions within GAAP to brazen fraud. Despite decades of academic research on such manipulations, there is little consensus about their prevalence (Beneish [2001], Dechow, Ge, and Schrand [2010], Ball [2013]).¹ The lack of consensus likely arises from the inherent challenges in identifying intentional acts of earnings manipulations. Archival researchers are challenged to infer manipulations from data confounded by factors like firm performance (Gerakos [2012], Leuz and Wysocki [2016]), and survey researchers are challenged by managers’ presumed desire to conceal manipulations (Dichev et al. [2013]). We circumvent these challenges by using a novel experimental method to estimate the prevalence of five forms of earnings manipulation: accrual-based earnings management, real earnings management, material omissions, disclosure obfuscation, and fraud. We then compare these estimates to estimates we calculate from executives’ responses to direct questions about the same manipulations.

Although difficult to obtain, reliable estimates of the prevalence of different forms of earnings manipulation are useful to several parties. For example, auditors can use prevalence estimates to assess pre-audit misreporting risk and to identify which actions managers are most likely to take when trying to manipulate earnings. Similarly, regulators can use prevalence

¹ For example, using a sample of 49,604 public firm-year observations, Bird, Karolyi, and Ruchti [2019] estimate that 2.61 percent of firms manipulate earnings to beat a benchmark; yet, survey evidence suggests CFOs believe 20 percent of companies manage earnings in any given period (Dichev et al. [2013]). Managers’ responses to hypothetical scenarios suggest the prevalence may be greater. For example, Graham, Harvey, and Rajgopal [2005] report that 40 percent of CFOs admit their company might shift the period of revenue recognition to meet an earnings target. Using a sample of 5,193 public firm-year observations, Zakolyukina [2018] estimates that 60 percent of CEOs manipulated earnings at least once between 2003 and 2010.
estimates to gauge past success in detecting intentional acts of misreporting and to inform
decisions of how to allocate limited resources to future enforcement activities. Prevalence
estimates also help regulators evaluate the costs and benefits of proposed regulations. For
example, it would be easier to justify a costly regulation designed to curb a more prevalent form
of earnings manipulation relative to a less prevalent form, *ceteris paribus*.

Reliable prevalence estimates are also useful to researchers and investors. For instance,
because fraud prevalence estimates are a key input into models that predict firm-specific fraud
likelihood (e.g., Perols et al. [2017]), more reliable estimates should benefit both the researchers
who develop these models and the investors who rely on them (e.g., investors who subscribe to
commercially available fraud risk measures (Price, Sharp, and Wood [2011])). Further, knowing
the prevalence of a particular form of earnings manipulation would likely inform researchers’
decisions to investigate potential determinants and deterrents (e.g., regulation) of that form of
manipulation. More reliable prevalence estimates can also improve our understanding of the
total agency costs of public ownership (Karpoff, Lee, and Martin [2008], Dyck, Morse, and
Zingales [2020]) and, therefore, researchers’ ability to model the principal-agent relationship.

In this study, we offer the first evidence on the prevalence of five different forms of earnings
manipulation based on executives’ responses to questions about their own and actual past
experiences. Study participants are top-level executives at companies in the Russell 3000 Index
whom we solicit using LinkedIn’s paid InMail (i.e., direct messaging) service. For each of the
five forms of earnings manipulation we examine, executives are asked if they are aware of a time
in the past five years where their company engaged in that form of manipulation. Executives then
provide either a direct *true or false* response or an indirect response, as explained below.

---

2 See Ball [2013] for a discussion of the costs of allocating research effort to investigating phenomena that are
uncommon or inconsequential.
To avoid ambiguity and maximize comparability with existing estimates, we word each question to be as close as possible to the relevant manipulation’s definition from prior research. For example, following Dichev et al. [2013, p. 24], our accrual-based earnings management question asks whether the executive’s company has used discretion within GAAP to report earnings that misrepresent the economic performance of the company. Similarly, our real earnings management question follows Graham et al. [2005] and Roychowdhury [2006], asking whether the executive’s company has recently changed an operational activity to meet a near-term earnings target at the expense of long-term value. We also ask about omitting material information (e.g., withholding bad news) and obfuscating disclosures (e.g., making bad news difficult to understand), as both are important behaviors that are difficult to confirm as intentional (Li [2008], Kothari, Shu, and Wysocki [2009], Dimitrov and Jain [2011], Bushee, Gow, and Taylor [2018]). Lastly, we ask about fraud—i.e., materially misrepresenting information with the intent to mislead users (Public Company Accounting Oversight Board [2010])—because fraud is a particularly consequential behavior that is difficult to detect (Karpoff et al. [2008], Amiram et al. [2018], Dyck et al. [2020]).

Although the wording of each question is identical for all participants, the questioning method varies by participant: direct or indirect. Executives randomly assigned to the direct questioning method are asked to affirm or disaffirm their awareness of each earnings manipulation occurring at their company by answering true or false, respectively, to each earnings manipulation question. The principal benefit of direct questioning is that executives’ responses are most interpretable (i.e., least noisy) when they definitively report on specific experiences they have had (Nelson and Skinner [2013]). One general drawback to directly asking executives about their experiences with earnings manipulations is that executives might lie or not
respond rather than admit to illegal or socially undesirable behavior. This response pattern, known as *social desirability bias*, would lead to downward biased prevalence estimates. To address and to subsequently measure the extent of this bias, we also use an indirect questioning method. Executives randomly assigned to the indirect questioning method respond to the same earnings manipulation questions, but, do so as part of a *list experiment*. A list experiment helps overcome social desirability bias by providing participants with plausible deniability such that not even the researcher can identify which participants admit to the self-incriminating behavior (Miller [1984], Grimm [2010]).

In a list experiment, participants view a list of statements and are asked to report only the *total number* of statements that apply to them. Participants do *not* indicate which specific statements apply. Participants in a control group receive a list of *K* statements and participants in a treatment group receive a list of *K* + 1 statements. *K* statements are innocuous statements presented to both groups (e.g., *I usually drive to work*), whereas the *K* + 1st statement concerns a sensitive behavior of interest and is only presented to the treatment group (e.g., a statement about manipulating earnings). By comparing the mean number of statements rated as applicable between groups, the researcher can estimate the prevalence of the sensitive behavior described in the *K* + 1st statement. In our study, for each form of earnings manipulation, a mean difference of 0.XX between groups would imply that approximately XX% of executives are aware of a time in the past five years where their company engaged in that form of earnings manipulation.

Our study makes several contributions. First, we extend prior research on the prevalence of earnings manipulations by providing the first estimates based on executives’ reports about their own, past experiences. To provide these estimates, we rely on a novel experimental method that complements existing research by circumventing misreporting concerns common to survey
research and identification concerns common to archival research (Miller [1984], Biemer et al. [2009]). We are also the first to provide evidence on the prevalence of disclosure obfuscations and material omissions. If our estimates suggest these forms of manipulation are more prevalent than previously understood, our study would highlight the importance of existing and future research on their determinants and consequences.

Beyond providing estimates of the prevalence of different forms of earnings manipulation, our study makes three methodological contributions. First, we introduce to the accounting literature an experimental method that, despite its widespread use in other contexts, has yet to be used in an accounting context. List experiments have been used successfully in a variety of fields to estimate the prevalence of sensitive behaviors as well as the degree of bias in estimates derived from direct questioning (for a review, see Hinsley et al. [2019]). Specifically, relative to direct questioning, list experiments have produced greater prevalence estimates on sensitive topics including deviance from COVID-19 public health guidance (Timmons et al. [2021]), bribery of foreign business leaders (Malesky, Gueorguiev, and Jensen [2015]), and religious disbelief (Gervais and Najle [2018]). Irrespective of the prevalence estimates we obtain, by testing the efficacy of the list experiment method in an accounting context, we inform researchers interested in estimating the prevalence of other sensitive accounting-related behaviors (e.g., an auditor’s choice not to report a detected financial statement error) about a potentially useful research method.

Second, by comparing the prevalence estimates obtained from our list experiment to those obtained from direct questioning, we inform researchers on the tradeoffs between using indirect and direct questioning methods to investigate strategic reporting behaviors such as earnings manipulations. List experiment estimates not statistically different from direct questioning
estimates would provide evidence that previous researchers’ methodology may be based on an incorrect assumption that executives’ dishonesty would systematically bias direct questioning estimates (Galasso et al. [2020]). In contrast, list experiment estimates greater than direct questioning estimates would support the assumption that direct questioning estimates suffer from systematic downward bias. If the latter occurs, we offer evidence that supports the use of a list experiment for estimating such behaviors.

Finally, we provide evidence about the efficacy of a novel tool for recruiting difficult to reach participants such as executives, audit partners, and financial analysts—LinkedIn’s paid direct messaging service. Prior survey research on executives’ behavior has largely relied on mass emailing campaigns to recruit participants (Graham et al. [2005], Dichev et al. [2013], Bernard, Juliani, and Lawrence [2021]), which yielded response rates ranging from 2-9%. We hope to improve on this response rate by sending direct messages to potential participants’ LinkedIn profiles, which allows us to confirm executives’ receipt of our request and to eliminate issues of invalid email addresses and spam filters. Although prior studies have used LinkedIn to gather observational data (e.g., employment history) from individuals’ profiles (e.g., Merkley, Michaely, and Pacelli [2020], Lee, Naiker, and Stewart [2021]) and to recruit survey participants via solicitation messages on group pages (e.g., Durney and Rennekamp [2019]), to our knowledge, we are the first to use LinkedIn’s paid direct messaging service as a recruitment tool.

2. Complementarities between the List Experiment Approach and Prior Approaches to Estimating the Prevalence of Earnings Manipulations

Our estimates of the prevalence of earnings manipulations complement prior estimates derived from archival methods (e.g., Burgstahler and Dichev [1997], Karpoff et al. [2008], Dyck et al. [2020]), structural models (e.g., Wang [2013], Zakolyukina [2018], Bird et al. [2019]), and survey evidence (e.g., Graham et al. [2005], Dichev et al. [2013]). Because there are advantages
and disadvantages to any one approach, forming estimates from multiple approaches (i.e., triangulating research methods) can be useful for developing a comprehensive understanding of phenomena of interest. A discussion of these complementarities follows.

A relative benefit of the archival approach is the ability to use large datasets of real-world time series accounting data, potentially improving generalizability and providing opportunities for cross-sectional analyses. Structural approaches also use real-world data to model earnings manipulations and to produce estimates of unobserved parameters, such as the costs and benefits to managers of manipulating earnings (e.g., Zakolyukina [2018]). One limitation of these approaches is that they often rely on empirical proxies of manipulations that are potentially confounded by factors like firm performance and managers’ appropriate use of discretion (Dechow and Skinner [2000]). To avoid these confounds, studies have relied on external indicators like misstatements or Accounting and Auditing Enforcement Releases (AAERs) (e.g., Dechow et al. [2011]). However, these studies are challenged to differentiate intentional manipulations from unintentional errors, and the inferences of these studies are necessarily based on manipulations that have been discovered, which neglects any undetected (i.e., unobservable) manipulations. While structural models attempt to address this “tip of the iceberg” problem, they require significant modeling assumptions—for example, assumptions about CEO wealth and motivations (Zakolyukina [2018]) or about the distribution of earnings (Bird et al. [2019]).

In contrast to estimates based on archival or structural approaches, estimates based on executives’ responses to carefully crafted survey questions do not hinge on proxies for manipulations or assumptions about executives’ unobservable intentions. The primary concern about survey approaches, though, is that prevalence estimates of earnings manipulations or other sensitive behaviors may be systematically biased downwards because of social desirability bias.
Prior survey researchers attempt to address social desirability bias by asking executives about the actions their company might take in hypothetical scenarios (Graham et al. [2005]) or by asking executives to estimate the prevalence of earnings manipulations in the population at large (Dichev et al. [2013]). In contrast to prior research, we use a list experiment to address social desirability bias, which offers executives the opportunity to respond to precise questions about their own and actual past experiences under the privacy protection of plausible deniability.

The benefit of asking executives about their own and actual past experiences is nontrivial, as survey responses are most interpretable when participants are asked directly about their own experiences (Nelson and Skinner [2013]). For example, relying on responses to a hypothetical scenario entails the risk that responses do not reflect or predict real decisions (Johansson-Stenman and Svedsäter [2012], Bostyn, Sevenhant, and Roets [2018]). Prior research suggests this risk may be particularly high when the hypothetical scenario involves moral or ethical decisions or when the scenario is not highly contextualized (FeldmanHall et al. [2012], Johansson-Stenman and Svedsäter [2012], Bostyn et al. [2018]). Although asking executives to estimate the prevalence of earnings manipulations in the population at large does not rely on executives reporting honestly about their own experiences, it “relies on [executives] being familiar with other companies’ opportunistic earnings management practices” (Dichev et al. [2013, p. 24]). Naturally, there is a risk that executives are not well informed about the deceitful reporting activities of other companies (Nelson and Skinner [2013]). In addition to this setting-specific risk, there are general risks associated with asking individuals to estimate the prevalence of a specific behavior in a population. As one example, research suggests that individuals disproportionately weigh salient information when making a prevalence estimate (Tversky and
Kahneman [1973]), which can add either noise or systematic bias to resulting estimates depending on the information sets available.³

In summary, there are tradeoffs to each approach. Our list experiment approach maintains the benefit of asking executives about their own, past experiences while reducing the threat of social desirability bias. However, because additional questions would threaten both executives’ privacy and the likelihood they complete our study, we are limited in our abilities to manipulate theoretical factors of interest and to include questions that allow for robust cross-sectional analyses. Whereas prior archival, structural, and survey research offer both prevalence estimates of and explanations for earnings manipulations, we focus our efforts exclusively on offering reliable prevalence estimates. For the reasons discussed in Section 1, we believe establishing more reliable estimates of the prevalence of earnings manipulations is useful in its own right.

3. Experimental Procedures

In our study, we randomly assign executives to one of three experimental conditions: List Experiment Group 1, List Experiment Group 2, or Direct Questioning.⁴ In each condition, executives respond to questions about five different forms of earnings manipulation. Prior to beginning the study, we inform executives about the purpose and sensitive nature of the study as well as about their privacy protections. In doing so, we hope to increase the perceived importance of the study, and, therefore, executives’ likelihood of participating (Cook, Heath, and Thompson [2000]). We also hope that executives will reciprocate our transparency about the forthcoming questions by responding honestly.

³ Research also finds that individuals who engage in a socially undesirable behavior over-estimate the percentage of their peers who behave similarly (false consensus effect) while individuals who do not engage in the undesirable behavior under-estimate the percentage of their peers who behave similarly (false uniqueness effect) (Suls, Wan, and Sanders [1988], Uvacsek et al. [2011]). Both the false consensus and false uniqueness effects introduce noise to the resulting prevalence estimate when asking about the actions of others.

⁴ The institutional review board at the authors’ institution has approved the use of human subjects for this study.
3.1 PARTICIPANT RECRUITMENT

3.1.1 Target Sample Size. Our target sample size is 690 executives—275 for each list experiment condition and 140 for our direct questioning condition.\(^5\)\(^6\) The imbalanced target sample size reflects the bias-variance tradeoff between direct questioning estimates and list experiment estimates; direct questioning estimates are more precise but potentially more precise around a *biased* value, whereas list experiment estimates are less precise, but potentially less precise around a *correct* value. All else equal, list experiment estimates are relatively less precise because of the statistical noise introduced by the innocuous statements included in each list. To offset differences in precision between our two approaches, we set the target sample size for each approach such that both provide estimates within ±7 percentage points of the “true” prevalence in the population (not correcting for bias) 90 percent of the time.\(^7\)

3.1.2 LinkedIn as a Recruitment Tool. To recruit participants, we send solicitation messages to executives via LinkedIn, which is a professional networking platform used by approximately 72 to 88 percent of C-Suite executives (Quartz [2016], JP Morgan [2018]). For each Russell 3000 firm, we collect the name of the Chief Financial Officer (CFO) from ExecuComp or SEC filings and manually search LinkedIn to identify the respective CFO’s LinkedIn profile. If we cannot match a CFO to a LinkedIn profile, we first try to identify the LinkedIn profile of the Chief Executive Officer (CEO) from the respective firm and then try to identify the LinkedIn

---

\(^5\) We follow the method in Glynn [2013] to calculate our target sample size for our list experiment conditions. To ensure that our final sample size is suitable for all five of our sensitive behaviors of interest, we rely on highly conservative assumptions. Appendix A provides these assumptions and the supportive calculations.

\(^6\) We calculate the sample size for our direct questioning condition using the formula \(n = p(1 - p)\frac{Z^2}{E^2}\), where \(p\) is the unknown population parameter (0.50), \(Z\) is the Z score for the chosen confidence level (90%), and \(E\) is the margin of error (±7 percentage points). We set \(p\) to 0.50 because it is the most conservative assumption.

\(^7\) Naturally, a larger sample size would allow for even greater precision. However, because of the elite nature of our population of interest (top-level executives), we must also consider the high cost of securing each additional participant when considering our “optimal” sample size.
profile of the next highest paid executive, per SEC filings. We target one executive from each firm to avoid double counting the same action(s), as doing so would add noise to our estimates.

After identifying a profile for one executive from each firm, we use LinkedIn’s paid InMail service to manually send a solicitation message to each profile. In each message, we identify ourselves as academic researchers, describe the topic of study, emphasize the privacy protections afforded to participants (which vary between direct questioning and list experiment conditions), and invite the executive to enter our study through an enclosed hyperlink. In explaining participants’ privacy protections, we describe our inability to identify participants or their responses, and we note that a one-way nondisclosure agreement is attached to our message (as in Bernard et al. [2021]). To incentivize participation, we state that the research team will donate $5 to the American Cancer Society for each submitted response. Appendix B separately presents the solicitation message for the list experiment conditions and the direct questioning condition. Figure 1 outlines our messaging procedures.

We collect responses over several months because LinkedIn limits the number of direct messages we can send each month. In the first month, we send 450 messages. When an executive opens our message, LinkedIn presents the full text of the message and two separate one-click response options labeled Yes, interested… and No thanks.... Any messages either accepted or rejected result in an InMail “credit,” which allows us to increase the number of messages we send the following month. That is, we send $450 + c_{t-1}$ messages in month $t$, where

---

8 We send messages in the middle of each month as opposed to the beginning in an effort to contact executives when they are least busy.
$c_{t-1}$ represents the number of credits we roll forward from the prior month. If an executive does not reply, we do not receive a credit for that message.\textsuperscript{9}

An executive can complete the study through the link provided regardless of whether they accept, reject, or do not respond to our message. Although we cannot know with certainty whether a specific executive completes our study, we assume that executives who select *No thanks...* are not willing to participate. We record the firms from which executives respond with *No thanks...* and we identify the LinkedIn profile of the CEO or the next highest-ranking executive of the respective firm. We begin sending messages to these newly identified profiles after all messages have been sent to our initial list of profiles.

We recruit participants through LinkedIn instead of email for several reasons. First, LinkedIn reports that the “InMail response rate is three times higher than regular email” (Petrone [2015]). Although we use InMail for a new purpose, we expect that our InMail response rate will exceed the response rates reported in prior survey studies that recruit executive-level participants via email, which range from 2% to 9% (Graham et al. [2005], Dichev et al. [2013], Bernard et al. [2021]). Second, it is difficult to know with certainty whether an email is delivered to a recipient because emails might be sent to inactive accounts or blocked by spam filters. In contrast, we know with certainty whether InMail is delivered to an executive’s inbox and it is presumably less likely that InMail is delivered to an inactive account given that executives frequently use LinkedIn (JP Morgan [2018]). Third, we expect that executives will view InMail as more credible than email because executives can simply click the sender’s avatar to view his or her profile and InMail is a paid service monitored by LinkedIn. Finally, InMail makes it easier for

\textsuperscript{9} Each month’s messages are sent to a random selection of the remaining names in our list of profiles. Before we achieve our minimum sample size for the direct questioning condition, 33 percent of the messages include the direct questioning solicitation and 67 percent include the list experiment solicitation. After we achieve our minimum sample size for the direct questioning condition, all messages include the list experiment solicitation.
executives to express their unwillingness to participate by providing a one-click *No thanks...* option. In the absence of a signal that an executive is unwilling to participate, sending messages to more than one executive at a firm entails the risk of overweighting the same action(s) because we do not track which executives complete our study. By reducing the burden to express unwillingness, InMail likely increases our ability to contact other executives in a firm when the initial recipient is unwilling to participate, thereby increasing our firm-specific response rate.\(^\text{10}\)

3.1.3 Supplemental Recruitment. Despite using a potentially advantageous recruitment tool and contacting multiple executives from the same firm when necessary, we may require supplemental data collection efforts. If this occurs, we will reach out to conference organizers at multiple conferences to request that we may attend for the purpose of soliciting participation in our study. Target conferences include the MIT Sloan CFO Summit in November of 2022, the AICPA & CIMA CFO Conference in April of 2023, and the Argyle CFO Summit in June of 2023. Respectively, these conferences are regularly attended by over 500, 375, and 400 executives.\(^\text{11}\) To avoid collecting two responses from the same individual (e.g., via our LinkedIn and in-person methods), we ask all participants to indicate whether they have previously participated in research with the same purpose and method. By targeting conferences on both coasts of the U.S., we hope to minimize our chances of receiving duplicate observations and maximize the generalizability of our sample to the population of interest. To better understand the consequences of relying on two recruitment methods, we plan to investigate whether the results are sensitive to recruitment method (LinkedIn versus conference) and to report the

\(^{10}\) Although discussions with a representative at LinkedIn indicate that sending 3,000+ messages using InMail could cost upwards of $25,000, we estimate a similar cost for obtaining access to executives’ individual email addresses through other services. Further, depending on a LinkedIn user’s settings, LinkedIn automatically contacts the user via email when an InMail is received. Thus, we potentially reach executives via InMail *and* email.

\(^{11}\) See https://www.mitcfo.com/, https://future.aicpa.org/cpe-learning/conference/aicpa-cfo-conference, and https://argyleforum.com/events/cfo-virtual-summit-june-22-23/. To offer in-person participants the greatest possible privacy protections, we plan to only assign these participants to our list experiment conditions.
potential overlap between samples—i.e., the percentage of conference attendees at each conference who are executives of Russell 3000 firms.

3.2 LIST EXPERIMENT CONDITIONS

3.2.1 List Experiment Design. A list experiment is an indirect questioning method designed to elicit honest disclosures from participants about self-incriminating or sensitive behaviors. In a list experiment, participants respond to questions about their engagement in sensitive behaviors under the privacy protection of plausible deniability. To illustrate this protection, consider the following hypothetical example of a standard list experiment:

*Please report the total number of statements that are true for you. Do not identify the specific statements that are true for you.*

<table>
<thead>
<tr>
<th>List A (Control)</th>
<th>List A (Treatment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I feel like I work all the time.</td>
<td>1. I feel like I work all the time.</td>
</tr>
<tr>
<td>2. I have a hobby that I enjoy regularly.</td>
<td>2. I have a hobby that I enjoy regularly.</td>
</tr>
<tr>
<td>3. I get my news via social media.</td>
<td>3. I get my news via social media.</td>
</tr>
<tr>
<td>4. I would rather read articles on printed paper than on a screen.</td>
<td>4. I would rather read articles on printed paper than on a screen.</td>
</tr>
<tr>
<td></td>
<td>5. I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.</td>
</tr>
</tbody>
</table>

Number of true statements: $C_A$  Number of true statements: $T_A$

In this example, participants receive either the list of statements on the left (the control list) or the list of statements on the right (the treatment list). Both lists contain identical innocuous statements (statements 1 – 4), and the treatment list contains an additional statement about misrepresenting the company’s economic performance (statement 5). After reviewing the list of statements, participants report the total number of statements that are true for them. Because participants do not identify the specific statements that are true, treatment list participants can
admit to misrepresenting their company’s economic performance under the protection of plausible deniability. That is, absent design flaws in the list experiment (e.g., innocuous statements that are true for every participant), no one other than the participant can know whether a statement is rated as true or false. If random assignment is successful, participants should respond similarly to the innocuous statements irrespective of which list they view. Thus, the prevalence of within GAAP misrepresentations described in statement 5 can be estimated by $\text{Mean}(T_A) - \text{Mean}(C_A)$.

List experiments generally produce more accurate prevalence estimates of highly sensitive behaviors relative to asking individuals directly about such behaviors (Hinsley et al. [2019]). For example, compared to direct questioning, list experiments have yielded greater prevalence estimates on a variety of sensitive topics including risky sexual behaviors (LaBrie and Earleywine [2000]), anti-gay sentiments (Coffman, Coffman, and Ericson [2016]), and fraudulent voting (Carkoglu and Aytaç [2015]). If executives are unwilling to respond truthfully to direct questions about their company’s involvement in earnings manipulations because these behaviors are too sensitive, we expect a list experiment will yield more accurate (and higher) prevalence estimates than direct questioning.¹²

To increase the power of our tests, we adapt the standard list experiment to a double list experiment design (Glynn [2013]). In a double list experiment, participants respond to two lists

¹² One well-known alternative to a list experiment is the randomized response technique. Under this technique, participants are given a randomizing device (e.g., coin, die) and are instructed to automatically answer yes to the sensitive question or to respond truthfully to the sensitive question based on the output of the randomizing device, which the experimenter does not observe. For example, participants could be instructed to respond yes to “Have you engaged in fraud?” if their coin lands on heads and to respond truthfully (yes or no) if their coin lands on tails. We use a list experiment because prior research suggests that, relative to the randomized response technique, a list experiment increases participants’ sense of anonymity, reduces nonresponse bias, and ultimately produces a more reliable prevalence estimate (Coutts and Jann [2011]). Moreover, a list experiment is less cumbersome and requires less time, as it does not involve the use of a randomizing device.
(rather than one) such that participants function simultaneously as control and treatment observations. To illustrate, consider a double list version of the previous example.

<table>
<thead>
<tr>
<th>List Experiment Group 1</th>
<th>List Experiment Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>List A (Control)</strong></td>
<td><strong>List A (Treatment)</strong></td>
</tr>
<tr>
<td>1. I feel like I work all the time.</td>
<td>1. I feel like I work all the time.</td>
</tr>
<tr>
<td>2. I have a hobby that I enjoy regularly.</td>
<td>2. I have a hobby that I enjoy regularly.</td>
</tr>
<tr>
<td>3. I get my news via social media.</td>
<td>3. I get my news via social media.</td>
</tr>
<tr>
<td>4. I would rather read articles on printed paper than on a screen.</td>
<td>4. I would rather read articles on printed paper than on a screen.</td>
</tr>
<tr>
<td>5. I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.</td>
<td>5. I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>List B (Treatment)</th>
<th>List B (Control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I usually get home from work by 5:30 pm</td>
<td>1. I usually get home from work by 5:30 pm</td>
</tr>
<tr>
<td>2. My commute to work is 45 minutes or longer.</td>
<td>2. My commute to work is 45 minutes or longer.</td>
</tr>
<tr>
<td>3. I intend to retire while working for my current employer.</td>
<td>3. I intend to retire while working for my current employer.</td>
</tr>
<tr>
<td>4. In the past year, at least one other employer has expressed interest in hiring me.</td>
<td>4. In the past year, at least one other employer has expressed interest in hiring me.</td>
</tr>
<tr>
<td>5. I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.</td>
<td>5. I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.</td>
</tr>
</tbody>
</table>

Number of true statements: $G_{1A}$  
Number of true statements: $G_{2A}$  
Number of true statements: $G_{1B}$  
Number of true statements: $G_{2B}$

In this example, participants in list experiment group 1 serve as control observations for list A and as treatment observations for list B. The opposite is true for participants in list experiment group 2. While both treatment list A and treatment list B include the same sensitive statement (statement 5), they contain different innocuous statements. A single estimate of the prevalence of a sensitive behavior is calculated after pooling observations across treatment lists A and B and across control lists A and B. Given we are interested in five sensitive behaviors, participants in
either list experiment condition view a total of ten lists, five that contain a sensitive statement and five that do not.

Although using a double rather than standard list experiment design necessarily adds length to the experimental instrument, this additional length is trivial in light of two significant benefits. First, a double list experiment doubles the sample size of a standard list experiment, which increases the precision of the resulting estimates and the statistical power of the relevant tests (Glynn [2013]). Second, by employing a double list experiment, we ensure that each executive views all five sensitive statements during the study. If we were to instead employ a standard list experiment, executives assigned to the control condition would only view innocuous statements, which could cause executives to be confused or angry with us for seemingly wasting their time.

3.2.2 Sensitive Statements. We estimate the prevalence of five forms of earnings manipulation: accrual-based earnings management, real earnings management, material omissions, disclosure obfuscation, and fraud. To estimate these forms of manipulation, we present executives with five true/false statements, all of which are worded as *I am aware of a time in the past five years where my company ... [description of the relevant form of earnings manipulation]*. Our statements concern experiences from the past five years because we want to capture manipulations over a time frame that is long enough for executives to have had the opportunity to develop first-hand knowledge of these manipulations while keeping the time frame recent enough to provide evidence about current reporting practices.

The wording of our statements is inspired by prior research on how individuals’ desire to preserve a positive self-concept affects how they categorize or describe their own prior dishonest acts. First, we refer to the *company’s* actions rather than the *executive’s* actions because doing so increases the degree to which executives can disassociate themselves with earnings
manipulations and attribute responsibility to others (Ayal and Gino [2011]). In our context, this disassociation is beneficial because executives can accurately select *true* without having to admit to themselves that they have personally manipulated earnings. Second, we precisely describe each form of manipulation because it decreases the degree to which executives can form self-serving interpretations of the statements to justify a *false* response (Shalvi et al. [2015]). For example, relative to precise descriptions, ambiguous terms like “accrual-based earnings management” make it easier for executives to interpret the meaning of this manipulation in a manner that justifies a *false* response.

To precisely describe the forms of earnings manipulation, we draw on prior, related research. First, our accrual-based earnings management statement closely follows the wording from Dichev et al. [2013, p. 24]. In their survey, they ask managers “From your impressions of companies in general, in any given year, what percentage of companies use discretion within GAAP to report earnings which misrepresent the economic performance of the business?” Thus, our first statement is:

1. **Accrual-based Earnings Management**
   
   *I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.*

Our second statement, about real earnings management, also follows directly from prior literature. For example, Graham et al. [2005] report that CFOs are willing to consider sacrificing long-term value to smooth earnings, and Roychowdhury [2006] finds that firms manipulate real activities to avoid reporting annual losses. Our second statement is:

2. **Real Earnings Management**
   
   *I am aware of a time in the past five years where my company has changed an operational activity to meet a near-term earnings target at the expense of long-term value.*
The third statement concerns withholding bad news, another important disclosure topic in the literature. For example, Kothari et al. [2009] conclude that managers accumulate and withhold bad news up to a threshold, but quickly leak good news to investors. In addition, Dimitrov and Jain [2011] find that managers prefer to report good news before annual shareholder meetings and delay negative news until after shareholder meetings. We hope to capture both types of material omission with the following statement:

3. Material Omissions

I am aware of a time in the past five years where my company deliberately withheld information that, if known by investors, would discourage investment in the company.

The fourth statement, about obfuscating disclosures, is related to research investigating managers’ opportunistic use of complex language to obfuscate negative information (Li [2008], Bushee et al. [2018]). Importantly, when managers convey complex information to outsiders, they may unintentionally increase linguistic complexity, implying some uncertainty around the extent to which managers intentionally obfuscate negative information. Thus, we obtain an estimate of intentional disclosure obfuscation using the following statement:

4. Disclosure Obfuscation

I am aware of a time in the past five years where my company altered a disclosure to make unfavorable information more difficult to understand.

Finally, the fifth statement uses a common definition of fraud as materially misstating firm performance with the intent to mislead (Public Company Accounting Oversight Board [2010]).

5. Fraud

I am aware of a time in the past five years where my company materially misrepresented information in the financial statements with the intent to mislead users.

3.2.3 Innocuous Statements. Given our double list design, each sensitive statement requires eight unique innocuous statements (40 total). These innocuous statements are meant to be benign and elicit responses largely based on subjective and non-public information (e.g., I feel like I
work all the time). We select innocuous statements that are often unrelated to financial reporting because we expect these statements can be answered quickly and because there is no need to hide which statements are of research interest due to participating executives already being informed of the purpose of the study. To avoid participant confusion, we make it clear that while some statements may appear random, we include these statements to protect participants’ privacy. Importantly, in a meta-analysis of list experiments, Li and Van den Noortgate [2019] report that the topical similarity between innocuous statements and a sensitive statement does not significantly influence the effectiveness of a list experiment. We limit the use of innocuous statements that are objective and based on public information because these types of statements threaten participants’ privacy. That is, if we could objectively infer which innocuous statements an executive should respond affirmatively to, then we could also infer that executive’s response to a sensitive statement. Thus, we take care to match any objectively knowable statements with subjective statements such that each list confers plausible deniability. For example, a statement about a participant’s educational pedigree could be matched with a statement about a participant’s opinion.

Although list experiments can be run with as few as two innocuous statements for each sensitive statement, including four innocuous statements reduces the likelihood that an executive would honestly rate all innocuous statements in a list as true or all innocuous statements in a list as false. If such a ceiling or floor effect were to occur, executives may rightfully worry that their plausible deniability is threatened and neglect to affirm their engagement in the sensitive behavior accordingly. To further reduce the likelihood of ceiling or floor effects, we construct each set of four innocuous statements such that a typical participant will likely rate two statements as true and two statements as false. For example, we expect negatively correlated
responses to *I feel like I work all the time* and *I have a hobby that I enjoy regularly*, so we match this pair of statements with another pair of statements that we also expect will produce negatively correlated responses. Including negatively correlated pairs of innocuous statements in each list has the added benefits of minimizing the variances of responses and increasing the power of our statistical tests (Blair and Imai [2012], Glynn [2013]).

3.2.4 Planned Pilot Study. We recruit sixty participants with the job title of “accountant” through YouGov Direct (https://business.yougov.com/product/custom-research) to complete a pilot study. We use YouGov Direct instead of other participant-sourcing platforms like MTurk/Cloud Research because YouGov Direct allows us to specifically target participants in accounting roles. Pilot study participants respond directly to 80 innocuous statements (see Appendix C) and our five sensitive statements.

Our pilot study serves two purposes: (i) to identify 40 innocuous statements to use in our list experiment conditions, and (ii) to verify that all innocuous and sensitive statements are unambiguous. As described above, our list experiment is best served by identifying pairs of innocuous statements that likely result in negatively correlated responses. Thus, our list of 80 innocuous statements comprises 40 pairs of statements that we expect will produce negatively correlated responses. We use our pilot study to verify our predictions and to identify other negatively correlated pairs that we might not have previously considered. Because we include twice as many statements in our pilot study than are needed for our final instrument, we do not foresee issues in identifying enough statements.\(^{13}\)

---

\(^{13}\) Although we have a total of 80 innocuous statements to test, YouGov Direct limits the number of questions we can ask in a single survey to 50. Thus, we run two pilots simultaneously, each containing 40 innocuous statements and one containing the five sensitive statements. We only include the sensitive statements in one pilot to avoid the collection of duplicate responses to any single question.
To verify that all innocuous and sensitive statements are unambiguous, we ask pilot study participants to flag any statements they find confusing or ambiguous as they work through the study. This verification is especially important for our sensitive statements because individuals may differ in their beliefs about what constitutes a manipulation of earnings. Because our pilot participants work as accountants, we expect they can offer meaningful feedback on our sensitive statements even if they are not in a position to manipulate earnings.

3.2.5 List Experiment Procedures. Figure 2 illustrates the design of our list experiment conditions. First, we create ten lists of four innocuous statements (lists A-J) by selecting 20 pairs of innocuous statements that produce negatively correlated responses in our pilot study. We then match two unique lists of innocuous statements with each sensitive statement of interest. Having two lists of innocuous statements for every one sensitive statement allows each list experiment participant to simultaneously function as both a treatment and control observation for each sensitive behavior of interest.

We use Qualtrics survey software to develop our online instrument. The initial welcome page provides participants with information about the purpose and sensitive nature of the study as well as participants’ privacy protections. After the initial welcome page, participants view each list on its own screen, with the four or five statements presented in random order. Upon viewing each list, participants provide a single response indicating the number of statements in the list that are true for them. Given the sensitive nature of our research, participants are not required to provide a response to each list; however, if a participant attempts to skip a question, the software asks the participant to confirm this action is intentional.

Irrespective of list experiment group, we begin with a treatment list rather than a control list because we suspect that some executives would promptly exit if they initially viewed a set of
seemingly random questions unrelated to financial reporting. We then alternate between
treatment and control lists to maintain participants’ interest. We keep pairs of lists together to
achieve as close to an equal number of responses to treatment and control lists as possible,
despite any attrition we may observe. Specifically, following each treatment list, we present the
paired control list (e.g., list B follows list A for list experiment group 1, and vice versa for list
experiment group 2).

We present treatment lists in ascending order of sensitivity as presented in Figure 2. Ordering
the lists in ascending order of sensitivity likely increases the overall completion rate, as it is less
jarring for participants (Bradburn, Sudman, and Wansink [2004]). Specifically, we present the
treatment lists in the following order: real earnings management, disclosure obfuscation, accrual-
earnings management, material omissions, and fraud. Although we do not know executives’
relative sensitivity rating of each form of manipulation, we assume that executives will perceive
within GAAP manipulations (e.g., real earnings management) as less sensitive than outside
GAAP manipulations (e.g., fraud). Further, following Graham et al. [2005]’s finding that
executives are more likely to admit to considering engaging in real earnings management than
accrual-earnings management, we assume that real earnings management is perceived as
relatively less sensitive. The study ends after participants view all ten lists and a final page
containing demographic questions.

3.3 DIRECT QUESTIONING CONDITION

Participants randomly assigned to the direct questioning condition view only our five
sensitive statements and demographic questions. All five sensitive statements are presented on a
single screen in the same order as that of the list experiment conditions. Next to each statement,
the participant simply selects true or false, indicating whether the statement does or does not
apply to them. Including only the sensitive statements has the benefit of keeping the instrument very short, which should help maximize the response rate to our initial participation request and help minimize attrition.

4. Planned Analyses and Results

4.1 SAMPLE

4.1.1 Observed Sample Size and Response Rates. We sent XXXX messages via LinkedIn’s InMail service, XX% of which were accepted. Ultimately, XXX executives completed some portion of our study, leaving us with an overall response rate of XX%. We also compare the response rate of executives assigned to participate in our list experiment conditions to the response rate of executives assigned to the direct questioning condition. On one hand, asking executives to report directly on their company’s involvement in manipulating earnings may deter potential participants. On the other hand, being able to tell executives in advance that there are only five questions to answer may encourage potential participants. Irrespective of the results of this comparison, examining our response rates conditional on each approach should inform researchers on how a direct questioning approach or a list experiment approach may influence target participants’ willingness to participate in a study. This information is important, as maximizing such willingness is fundamental to quality survey and experimental research (Cook, Campbell, and Shadish [2002]).

4.1.2 Descriptive Statistics. To avoid threatening participants’ privacy, we collect limited demographic information. Specifically, we ask participants to share information about their job title, job tenure, gender, and CPA designation as well as about their firm’s industry, age, and annual revenue. Table 1 presents demographic information about our sample of executives and their respective firms. Because participants indicate their consent to participate in the research by
beginning the study, we use all available responses in the subsequent analyses irrespective of
whether the study was completed in its entirety.\footnote{Although we only reach out to executives of public companies in our first round of data collection (via LinkedIn), if we subsequently collect data from executives in-person at conferences, we will ask participants to indicate whether their company is publicly traded or privately owned.}

4.2 PLANNED PRIMARY ANALYSES

4.2.1 List Experiment Prevalence Estimates. We conduct a double list experiment such that
each participant responds to five treatment lists and five control lists, with four unique innocuous
statements presented in each of these ten lists. As illustrated in Figure 2, a participant from list
experiment group 1 views each sensitive statement in conjunction with the first of two innocuous
statement lists shown in each Panel (lists A, C, E, G, and I). In contrast, a participant from list
experiment group 2 views each sensitive statement in conjunction with the second of two
innocuous statement lists shown in each Panel (lists B, D, F, H, and J). Table 2 separately reports
the findings by list (A or B, C or D, etc.) as well as pooled across lists (A and B, C and D, etc.).

Our primary dependent variable of interest for the list experiment conditions is participants’
report of the total number of statements in a list that are true for them (NumberTrue).
NumberTrue is a discrete variable that ranges from 0 to 5 for lists that contain a sensitive
statement (treatment lists) and from 0 to 4 for lists that do not contain a sensitive statement
(control lists). Columns 2 and 3 report mean NumberTrue for treatment and control lists
separately for each type of earnings manipulation (Panels A through E).

To measure the prevalence of the five forms of earnings manipulation of interest, we estimate
ten variations of Equation (1) (one for each behavior/list combination) and five variations of
Equation (2) (one for each behavior pooled across the two relevant lists):

\begin{align}
(1) \quad \text{NumberTrue} &= \lambda_0 + \lambda_1 \ast \text{Treatment} + \varepsilon \\
(2) \quad \text{NumberTrue} &= \beta_0 + \beta_1 \ast \text{Treatment} + \beta_2 \ast \text{List} + \varepsilon
\end{align}
NumberTrue is as defined above and Treatment is a binary variable equal to one if the response is from a treatment list, and zero otherwise.\textsuperscript{15} Equation (2) differs from Equation (1) in that we pool our data across the two relevant lists (A and B, C and D, etc.) and add a control variable List, which is a binary variable equal to one for the first of the two relevant lists (A, C, E, G, and I), and zero otherwise. By including List, we account for the differing innocuous statements in the two treatment lists. Because each participant provides responses to both lists, we cluster standard errors at the individual level in Equation (2). Column 4 reports $\lambda_1$ and $\beta_1$ from each variation of Equations (1) and (2), which correspond to the difference between the average number of statements agreed to in a treatment list and the average number of statements agreed to in a control list.\textsuperscript{16} If the prevalence of a form of earnings manipulation is statistically greater than zero, the respective 90\% confidence interval reported in column 6 will not include zero.

Importantly, there is little risk that our list experiment method over estimates the prevalence of earnings manipulations because executives are unlikely to admit to socially undesirable behavior that has not actually occurred. However, despite the plausible deniability afforded by the list experiment method, some executives may still not be willing to admit to a sensitive behavior. In this sense, each list experiment estimate likely represents a lower (rather than an upper) bound on the prevalence of a particular form of earnings manipulation.

\textbf{4.2.2 Direct Questioning Prevalence Estimates.} Our primary dependent variables of interest for our direct questioning condition are participants’ true or false responses to each sensitive

\textsuperscript{15} We estimate Equations (1) and (2) using OLS for easier interpretation of the coefficients. Since our dependent variable (NumberTrue) is a discrete variable, we evaluate the robustness of our results using a Poisson regression.\textsuperscript{16} Because we design each list of innocuous statements to include two statements we predict will apply to the average participant and two statements we predict will not apply to the average participant, we expect all control lists of statements to have a mean NumberTrue of approximately two. Thus, we do not expect $\beta_2$ to generally be statistically different from zero.
statement. Thus, RealEarningsManagement, DisclosureObfuscation, AccEarningsManagement, MaterialOmissions, and Fraud are each equal to one if the participant confirms the relevant behavior, and zero otherwise. Table 2 reports the mean of each dependent variable in column 8, which represents the prevalence estimate of the relevant behavior derived from executives’ responses to our direct questions.

4.2.3 Examining Executives’ Dishonesty. Although list experiments can reduce misreporting, there is a potential loss of precision relative to direct questioning. Accordingly, list experiments are only preferable to direct questioning when a large portion of participants respond dishonestly to direct questions. To estimate executives’ dishonesty when asked directly about their company’s involvement in manipulating earnings, we compare the prevalence estimates we obtain using our list experiment approach to those we calculate from our direct questioning approach (see column 9). Each comparison reported in column 9, therefore, represents the degree to which prevalence rates derived from direct questioning under(over)-estimate the prevalence of the relevant form of earnings manipulation. We use a Wald test to compare this difference to zero, expecting either that executives underreport or do not misreport when asked directly about their experiences with earnings manipulations. Rejecting the null hypothesis would support the assumption relied upon in prior research that executives’ dishonesty systematically biases direct questioning estimates. Failing to reject the null hypothesis would suggest limited upside to using a list experiment to measure executives’ strategic reporting behaviors.

We also examine whether executives’ dishonesty varies by form of earnings manipulation. To measure the rate of dishonesty, we scale each estimate reported in column 9 by the respective list experiment estimate reported in column 4. The resulting percentages indicate the probability that an executive fails to directly admit to a behavior given having engaged in the behavior.
Whereas high likelihoods of dishonesty indicate executives’ awareness of the inappropriateness of the behavior, low likelihoods (if matched with high prevalence rates) indicate executives’ indifference towards the behavior.

4.3 PLANNED CROSS-SECTIONAL ANALYSES

We next examine whether any of the demographic variables we collect could be useful predictors of either executives’ (mis)behavior or their responses to direct questions about such behavior. For each form of earnings manipulation, we estimate Equation (3) using OLS (list experiment conditions), clustering standard errors at the individual level, and Equation (4) using logistic regression (direct questioning condition).

\[ NumberTrue = \omega_0 + \omega_1 \times Treatment + \omega_2 \times List + \sum_{i=1}^{6} \omega_{i+2} \times Characteristic_i + \sum_{i=1}^{6} \omega_{i+8} \times Treatment \times Characteristic_i + \varepsilon \]

\[ Manipulation = \phi_0 + \sum_{i=1}^{6} \phi_i \times Characteristic_i + \varepsilon \]

NumberTrue, Treatment, and List are as defined above, Manipulation is RealEarnings Management, DisclosureObfuscation, AccEarningsManagement, MaterialOmissions, or Fraud, as defined above, and Characteristic_i corresponds to a single individual characteristic. Based on the demographic information we collect, we investigate six different characteristics: ExecMale, ExecTenure, ExecCPA, FirmAge, FirmSize, and FirmSectorIsTech. ExecMale is a binary variable equal to one if the participant identifies as male, and zero otherwise. ExecTenure is a discrete variable equal to one, two, or three, corresponding to whether the participant reports being at their current company less than 4 years, 4-9 years, or 10 years or more, respectively. ExecCPA is a binary variable equal to one if the participant reports being a licensed CPA, and zero otherwise. FirmAge is a binary variable equal to one for observations where the participant
reports a firm age greater than the median in our sample, and zero otherwise. \textit{FirmSize} is a binary variable equal to one for observations where the participant reports revenue in the most recent fiscal year greater than the median in our sample, and zero otherwise. \textit{FirmSectorIsTech} is a binary variable equal to one for observations where the participant reports working for a company in the \textit{Technology [Software/Biotech]} industry, and zero otherwise. We isolate executives from the technology sector following Graham et al. [2005], who find that CFOs in the technology industry respond differently than CFOs in other industries to some hypothetical scenarios involving earnings manipulation. Only executives that provide data on all six demographic characteristics are included in the analyses reported in Table 3.

Table 3 reports results of these analyses. Significant coefficients on \textit{Treatment x Characteristic} in Equation (3) represent the existence of cross-sectional variation in company reporting behavior. Significant coefficients on \textit{Characteristic} in Equation (4) represent the existence of cross-sectional variation in executives’ responses to direct questions about their company’s behaviors. We do not have specific predictions for the explanatory power of the demographic characteristics we measure, consistent with Graham et al. [2005]’s finding that demographic variables generally do not affect their measures of misreporting behaviors.

4.4 PLANNED SUPPLEMENTAL ANALYSES

4.4.1 Internal Consistency of our List Experiment Approach. Our double list experiment is, by construction, two versions of a standard list experiment. Thus, to provide further evidence on the efficacy of the list experiment approach in an accounting context, we examine the internal

\footnote{Due to executives’ ability to skip questions in our instrument, there is likely to be variation in the number of observations we have for each demographic variable. To ensure the most powerful test of each individual characteristic, we also conduct six additional variations of Equations (3) and (4) for each form of earnings manipulation; in each variation, we include a single characteristic variable, excluding the other five. We report any differences in inferences between these results and those reported in Table 3.}
consistency of the prevalence estimates obtained from the standard list experiment approach. Specifically, we use a Wald test to compare the coefficient on Treatment \((\lambda_1)\) across each of the five pairs of lists (A and B, C and D, etc.). Failing to reject the null hypothesis supports the method’s internal consistency.

4.4.2 Precision Gains of our Double List Experiment. Relative to a single list experimental design, we expect the double list design to allow for a significant increase in precision. We examine this increase in precision by comparing the standard error from each variation of Equation (1) to the standard error from the relevant variation of Equation (2).

4.4.3 List Experiment Assumptions. Our list experiment relies on three main assumptions: (i) successful randomization of participants to conditions, (ii) the absence of design effects, and (iii) the absence of ceiling and floor effects (Kuklinski, Cobb, and Gilens [1997], Blair and Imai [2012], Lépine, Treibich, and d’Exelle [2020]).

Successful randomization refers to the assumption that participants in each group are sufficiently similar such that, on average, they respond affirmatively to the same number of innocuous statements in each list. Ex ante, we randomly assign participants to experimental conditions. Ex post, we test whether randomization was successful by comparing participants’ demographic characteristics across list experiment conditions. Tests that fail to reject the null hypothesis of no difference would support successful randomization.

Assuming the absence of design effects means assuming the addition of a sensitive statement to a list does not change the sum of affirmative responses to the innocuous statements in the list. Following Blair and Imai [2012], we test this assumption by examining whether the proportion of participants who agree with no more than 0, 1, 2, 3, and 4 statements in a list differs based on the inclusion of a sensitive statement—i.e., we compare each proportion across treatment and
control lists with identical innocuous statements (lists A-J in Figure 2). We use two approaches to examine the existence of design effects. First, the proportion of participants that agree with no more than 0, 1, 2, 3, or 4 statements in a control list should be greater than the proportion of participants that agree with no more than 0, 1, 2, 3, or 4 statements in a treatment list, respectively. Second, the proportion of participants that agree with no more than 1, 2, 3, 4, or 5 statements in a treatment list should be greater than the proportion of participants that agree with no more than 0, 1, 2, 3, or 4 statements in a control list, respectively.

Finally, list experiments rely on the assumption that responses to innocuous statements cannot be used to infer responses to a sensitive statement. This assumption is threatened when participants would truthfully affirm or disaffirm all innocuous statements in a list. We refer to such occurrences as ceiling and floor effects, respectively. We take several steps to decrease the likelihood of experiencing ceiling or floor effects, as described in detail in Section 3.2.3.

Although we cannot test for the influence of ceiling or floor effects, for transparency, we report the proportion of participants who agree with zero statements and the proportion of participants who agree with four statements for each control lists across our two list experiment conditions.

5. Conclusion

In this paper, we provide the first estimates of the prevalence of five forms of earnings manipulation based on executives’ reports about their own, past experiences. To provide these estimates, we use both a direct and an indirect questioning method. Executives assigned to our direct questioning method affirm or disaffirm their awareness of each form of earnings manipulation occurring at their company by responding true or false to five separate manipulation questions. Executives assigned to our indirect questioning method respond to the same five questions but do so within a list experiment, which is designed to elicit honest
responses about socially undesirable behaviors under the privacy protection of plausible
deniability. By using both questioning methods, we estimate the level of dishonesty exhibited by
executives assigned to our direct questioning method and produce prevalence estimates that are
meaningful in the absence or presence of dishonesty. Our approach also addresses important
concerns about existing estimates of the prevalence of various forms of earnings manipulation.

One reason we believe our study is well-suited for the registration based editorial process is
that we do not make or test formal hypotheses. That is, our results should inform readers’
understanding of the prevalence of earnings manipulations irrespective of whether our analyses
offer estimates that are lower or higher than any individual reader might expect, *ex ante*. Even
without formal hypotheses, there remains a risk that the results of our planned analyses could be
difficult to interpret in some way. For example, although unlikely, it is possible that our list
experiment estimates turn out to be statistically lower than our direct questioning estimates.
Although we take several steps to minimize the risk of difficult to interpret results, we can never
fully eliminate this risk. Nonetheless, results that are difficult to interpret would not negate the
methodological contributions we discuss in the Introduction. For instance, we would interpret
such results as evidence that the list experiment method is perhaps not useful for examining
executives’ strategic reporting behaviors. In this scenario, we would also still learn something
about the efficacy of our novel recruitment method—LinkedIn’s direct messaging service.

Furthermore, we acknowledge that, like prior estimates, our estimates are subject to the time
and place of data collection. Considering the generality of this limitation, we believe researchers
should consider measuring the prevalence of earnings manipulations *more* rather than less
frequently. New estimates would seem especially useful following significant economic or
regulatory changes. For example, the inferences from Graham et al. [2005] are based on data
collected during the highly scrutinized post Enron/SOX-implementation period. As another example, Zakolyukina [2018] estimates the prevalence of earnings manipulations from 2003-2010, a seven-year period overlapping the financial crisis. Both examples suggest some need to periodically reexamine managers’ strategic reporting behavior to see if its prevalence changes over time.

Finally, we see the need to reexamine the prevalence of managers’ strategic reporting choices as related to our methodological contributions. If we can show that list experiments are useful in an accounting context, future research could periodically provide up-to-date estimates using the same or similar research design (e.g., conducting another list experiment about the prevalence of earnings manipulations five to ten years from now). Future research could also conduct list experiments with other populations, such as audit partners, to determine how frequently they detect certain earnings manipulations.
APPENDIX A

We follow the method in Glynn [2013] to calculate our target sample size for our double list experiment. Specifically, Glynn [2013] derives the sample size ($n$) required to obtain a prevalence estimate within a margin of error ($E$) as

$$n = \left( \frac{Z_{\alpha/2}}{E} \right)^2 \left( V \left[ y_{K+}^A \right] + V \left[ y_{K+}^B \right] + V[y_{iK+1}] - 2Cov \left[ y_{K+}^A, y_{K+}^B \right] \right),$$

where participant $i$ responds to either (i) control list $A$ containing $K$ statements and treatment list $B$ containing $K + 1$ statements, or (ii) control list $B$ containing $K$ statements and treatment list $A$ containing $K + 1$ statements. Based on the conservative inputs presented below, our target sample size for each list experiment condition is 275.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_{\alpha/2}$</td>
<td>Standardized Z score for confidence level</td>
<td>1.645 (90% confidence level)</td>
</tr>
<tr>
<td>$E$</td>
<td>Margin of error</td>
<td>±7%</td>
</tr>
<tr>
<td>$V \left[ y_{K+}^A \right]$</td>
<td>Variance of total number of statements participant $i$ rates as true on control list $A$</td>
<td>0.50</td>
</tr>
<tr>
<td>$V \left[ y_{K+}^B \right]$</td>
<td>Variance of total number of statements participant $i$ rates as true on control list $B$</td>
<td>0.50</td>
</tr>
<tr>
<td>$V[y_{iK+1}]$</td>
<td>Variance of participant $i$ response to sensitive statement</td>
<td>0.25 (maximum possible variance)</td>
</tr>
<tr>
<td>$Cov \left[ y_{K+}^A, y_{K+}^B \right]$</td>
<td>Covariance between total number of statements rated as true on control list $A$ and total number of statements rated as true on control list $B$</td>
<td>0.01</td>
</tr>
</tbody>
</table>
APPENDIX B

We recruit participants by sending solicitation messages to executives via LinkedIn’s InMail. The specific message a potential participant receives depends on which of three conditions the participant is randomly assigned – List Experiment Group 1, List Experiment Group 2, or Direct Questioning.

List Experiment Conditions

![Example of a solicitation message sent via LinkedIn's InMail](image-url)
Direct Questioning Condition

Hello [Exec Name],

My name is [Faculty Name] and I am an accounting professor at [University Name]. I am writing because your perspective on accounting practices is invaluable, and I want to invite you to a 2-minute study about your financial reporting experiences. For each completed study, my research team is donating $5 to the American Cancer Society.

Because the study involves a few sensitive questions about misleading financial reports, we use an anonymized link that makes it impossible for the research team or anyone else to know who completes the study and to match individuals with specific responses. Despite this impossibility, I have attached a one-way Non-Disclosure Agreement to offer additional comfort that your involvement is confidential. The NDA automatically binds me and my colleagues and you do not need to sign anything.

To complete the quick study and earn money for charity, click here: [insert link]

If you have any questions or concerns, please let me know.
Sincerely,
[Faculty Name]
APPENDIX C

We recruit sixty participants with the job title of “accountant” through YouGov Direct to complete a pilot study. Within the pilot, participants respond to 80 innocuous statements (listed below). These statements were chosen with two criteria in mind: (1) they are generally opinion (not factual) based, and (2) each has at least one other statement on this list for which we expect to produce negatively correlated responses.

1. My commute to work is 45 minutes or longer.
2. I usually get home from work by 5:30 pm.
3. I typically travel out of state for work more than three times each year.
4. I am happy with my work-life balance.
5. I completed an econometrics course in college.
6. I have an accounting degree.
7. The company I work for employs less than 250 people.
8. My company was founded more than 25 years ago.
9. In my current job, I work closely with external auditors.
10. A large part of my job is capital budgeting and planning.
11. I consider myself to be an avid golfer.
12. I take less than 2 weeks of vacation each year.
13. I feel like I work all the time.
14. I have a hobby that I enjoy regularly.
15. I think the employees in the accounting department have a poor work-life balance.
16. My company has an employee-friendly vacation policy.
17. My company provides excellent health care benefits.
18. I am not satisfied with the matching contributions my company makes to employee retirement accounts.
19. My perfect vacation involves spending all day on a beach.
20. When I travel, I spend considerable time visiting tourist attractions.
21. My family has multiple cars.
22. I take public transportation regularly.
23. I believe student-athletes make poor employees.
24. I played sports at the collegiate level.
25. I have recently changed my diet.
26. I find it difficult to find time to cook at home.
27. I get my groceries delivered from a service.
28. I do not like to shop for clothes online.
29. I grew up in a place with a warm climate.
30. I enjoy skiing.
31. I get my news via social media.
32. I would rather read articles on printed paper than on a screen.
33. I order lunch more than three times a week.
34. I currently work from home more days than I work from the office.
35. I grew up in a small town.
36. I believe my company is not doing enough in terms of its corporate social responsibility.
37. Part of my compensation includes financial support for my vehicle.
38. I pay out-of-pocket for continuing professional education (CPE).
39. I think the corporate tax rate is too high.
40. I supported the Tax Cut and Jobs Act of 2019.
41. The federal government has done an adequate job in response to the COVID-19 pandemic.
42. The COVID-19 pandemic has highlighted and exacerbated a number of societal issues.
43. I believe that I am under-compensated relative to my peers.
44. I am satisfied with the relationship I have with my superiors.
45. I believe I am well-liked by my colleagues.
46. There has been significant turnover in my department in the past five years.
47. My colleagues really came together as a team during the COVID-19 pandemic.
48. I do not care about whether my colleagues like me.
49. I intend to retire while working for my current employer.
50. In the past year, at least one other employer has expressed interest in hiring me.
51. I have an MBA degree.
52. I attended a trade school.
53. I was recently a guest lecturer at a local university.
54. I moved to a different state in the past 2 years.
55. I provide significant input into operational decisions at my company.
56. I make decisions that shape the long-term plan for my company.
57. A significant proportion of employees at my company interact directly with customers.
58. Most of my company’s customers are other businesses.
59. My company gains new customers primarily through word of mouth.
60. My company spends a lot of money on advertising.
61. A pressing concern for my company is generating revenue growth.
62. My company has healthy cash flow.
63. A pressing concern for my company is expanding in international markets.
64. My company outsources most of the repetitive and highly transactional processes.
65. I spend most of my time helping to craft and execute corporate strategy.
66. A significant part of my job involves control and compliance processes, including financial reporting.
67. My company successfully uses data analysis to aid decision making.
68. A key priority for my company in the coming year is to automate routine processes via technology.
69. My company’s financial performance does not depend much on which political party is in control of Congress and the White House.
70. My company engages in considerable political lobbying.
71. Excessive environmental regulation threatens my company’s financial performance.
72. My company has a low carbon footprint.
73. The majority of my company’s employees are paid hourly.
74. My company plans to increase work-from-home opportunities for most of our employees.
75. I am content with the relationship I have with my coworkers.
76. There is high turnover among the executives in my firm.
77. In a typical year, I travel to three or more countries for business purposes.
78. My firm is primarily focused on serving the U.S. market.
79. I generally support the economic policies of the Biden Administration.
80. My company is under significant pressure from stakeholders to consider climate change.
REFERENCES


PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD. *Consideration of Fraud in a Financial Statement Audit, AS 2401.05. 2010.*


Timeline of sending solicitation messages via LinkedIn

<table>
<thead>
<tr>
<th>Send messages to 1st executive from each firm in the Russell 3,000 Index</th>
<th>Begin Retargeting Missing Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month 1</td>
<td>Month 2</td>
</tr>
<tr>
<td>Send 450</td>
<td>Send 450 + c₁</td>
</tr>
</tbody>
</table>

Where cᵣ is the number of LinkedIn messages in month 𝑡 in which an executive accepts or rejects the message via the one-click response options labeled “Yes, interested…” or “No thanks…”. LinkedIn limits the number of messages that can be sent in month 𝑡 to 450 + cᵣ ᵉᵣ.

Figure 1 illustrates the timeline for sending solicitation messages to executives via LinkedIn. Prior to month 1, we identify a LinkedIn profile for one executive (primarily CFOs) from each firm in the Russell 3,000 Index. In each month, we randomly assign 450 + cᵣ ᵉᵣ executives to one of three conditions: List Experiment Group 1, List Experiment Group 2, and Direct Questioning. We then send the relevant solicitation message shown in Appendix B to each executive. We continue this procedure until we have sent a message to one executive from each firm in the Russell 3,000 Index. If we achieve the target sample size for any condition prior to messaging all 3,000 executives, we assign remaining executives to conditions that are below the target sample size. After sending all 3,000 messages, for each message that receives a No thanks… response, we send a message to a different executive from the respective firm using a similar procedure.
<table>
<thead>
<tr>
<th>List Experiment Group 1</th>
<th>List Experiment Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Real Earnings Management</strong></td>
<td><strong>List A (Control)</strong></td>
</tr>
<tr>
<td>List A (Treatment)</td>
<td>1. Innocuous statement 1</td>
</tr>
<tr>
<td>1. Innocuous statement 1</td>
<td>2. Innocuous statement 2</td>
</tr>
<tr>
<td>2. Innocuous statement 2</td>
<td>3. Innocuous statement 3</td>
</tr>
<tr>
<td>3. Innocuous statement 3</td>
<td>4. Innocuous statement 4</td>
</tr>
<tr>
<td>4. Innocuous statement 4</td>
<td>5. I am aware of a time in the past five years where my company has changed an operational activity to meet a near-term earnings target at the expense of long-term value.</td>
</tr>
<tr>
<td><strong>List B (Control)</strong></td>
<td><strong>List B (Treatment)</strong></td>
</tr>
<tr>
<td>1. Innocuous statement 5</td>
<td>1. Innocuous statement 5</td>
</tr>
<tr>
<td>2. Innocuous statement 6</td>
<td>2. Innocuous statement 6</td>
</tr>
<tr>
<td>3. Innocuous statement 7</td>
<td>3. Innocuous statement 7</td>
</tr>
<tr>
<td>4. Innocuous statement 8</td>
<td>4. Innocuous statement 8</td>
</tr>
<tr>
<td>5. I am aware of a time in the past five years where my company has changed an operational activity to meet a near-term earnings target at the expense of long-term value.</td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: Disclosure Obfuscation</strong></td>
<td><strong>List C (Control)</strong></td>
</tr>
<tr>
<td>List C (Treatment)</td>
<td>1. Innocuous statement 9</td>
</tr>
<tr>
<td>1. Innocuous statement 9</td>
<td>2. Innocuous statement 10</td>
</tr>
<tr>
<td>2. Innocuous statement 10</td>
<td>3. Innocuous statement 11</td>
</tr>
<tr>
<td>3. Innocuous statement 11</td>
<td>4. Innocuous statement 12</td>
</tr>
<tr>
<td>4. Innocuous statement 12</td>
<td>5. I am aware of a time in the past five years where my company altered a disclosure to make unfavorable information more difficult to understand.</td>
</tr>
<tr>
<td><strong>List D (Control)</strong></td>
<td><strong>List D (Treatment)</strong></td>
</tr>
<tr>
<td>1. Innocuous statement 13</td>
<td>1. Innocuous statement 13</td>
</tr>
<tr>
<td>2. Innocuous statement 14</td>
<td>2. Innocuous statement 14</td>
</tr>
<tr>
<td>3. Innocuous statement 15</td>
<td>3. Innocuous statement 15</td>
</tr>
<tr>
<td>4. Innocuous statement 16</td>
<td>4. Innocuous statement 16</td>
</tr>
<tr>
<td>5. I am aware of a time in the past five years where my company altered a disclosure to make unfavorable information more difficult to understand.</td>
<td></td>
</tr>
</tbody>
</table>
Panel C: Accrual-based Earnings Management

**List E (Treatment)**
1. Innocuous statement 17
2. Innocuous statement 18
3. Innocuous statement 19
4. Innocuous statement 20
5. I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.

**List E (Control)**
1. Innocuous statement 17
2. Innocuous statement 18
3. Innocuous statement 19
4. Innocuous statement 20

Panel D: Material Omissions

**List F (Control)**
1. Innocuous statement 21
2. Innocuous statement 22
3. Innocuous statement 23
4. Innocuous statement 24
5. I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.

**List F (Treatment)**
1. Innocuous statement 21
2. Innocuous statement 22
3. Innocuous statement 23
4. Innocuous statement 24
5. I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.

**List G (Treatment)**
1. Innocuous statement 25
2. Innocuous statement 26
3. Innocuous statement 27
4. Innocuous statement 28
5. I am aware of a time in the past five years where my company deliberately withheld information that, if known by investors, would discourage investment in the company.

**List G (Control)**
1. Innocuous statement 25
2. Innocuous statement 26
3. Innocuous statement 27
4. Innocuous statement 28

**List H (Control)**
1. Innocuous statement 29
2. Innocuous statement 30
3. Innocuous statement 31
4. Innocuous statement 32

**List H (Treatment)**
1. Innocuous statement 29
2. Innocuous statement 30
3. Innocuous statement 31
4. Innocuous statement 32
5. I am aware of a time in the past five years where my company deliberately withheld information that, if known by investors, would discourage investment in the company.
Panel E: Fraud

<table>
<thead>
<tr>
<th>List I (Treatment)</th>
<th>List I (Control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Innocuous statement 33</td>
<td>1. Innocuous statement 33</td>
</tr>
<tr>
<td>2. Innocuous statement 34</td>
<td>2. Innocuous statement 34</td>
</tr>
<tr>
<td>3. Innocuous statement 35</td>
<td>3. Innocuous statement 35</td>
</tr>
<tr>
<td>4. Innocuous statement 36</td>
<td>4. Innocuous statement 36</td>
</tr>
<tr>
<td>5. I am aware of a time in the past five years where my company materially misrepresented information in the financial statements with the intent to mislead users.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>List J (Control)</th>
<th>List J (Treatment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Innocuous statement 37</td>
<td>1. Innocuous statement 37</td>
</tr>
<tr>
<td>2. Innocuous statement 38</td>
<td>2. Innocuous statement 38</td>
</tr>
<tr>
<td>3. Innocuous statement 39</td>
<td>3. Innocuous statement 39</td>
</tr>
<tr>
<td>4. Innocuous statement 40</td>
<td>4. Innocuous statement 40</td>
</tr>
<tr>
<td>5. I am aware of a time in the past five years where my company materially misrepresented information in the financial statements with the intent to mislead users.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 illustrates our double list experimental design. First, we create ten lists of four innocuous statements (lists A-J). We then match two unique lists of innocuous statements with each sensitive statement of interest. Having two lists of innocuous statements for every one sensitive statement ensures each list experiment participant simultaneously functions as both a treatment and control observation for each sensitive behavior or interest.

We use Qualtrics survey software to develop our online instrument. We present each list on its own screen, and participants provide a single response indicating the number of statements in the list that are true for them. Irrespective of list experiment group, we first present the treatment list and then present the control list from each successive Panel. For example, list B follows list A for list experiment group 1, or vice versa for list experiment group 2.
Table 1 presents the demographic characteristics of executive participants and their respective firms. Frequencies are based on non-missing responses to the demographic questions that we ask participants.
### TABLE 2

Estimated prevalence and misreporting using list experiments versus direct questioning.

<table>
<thead>
<tr>
<th>Column</th>
<th>(1) Projected N</th>
<th>(2) Number True Treatment</th>
<th>(3) Number True Control</th>
<th>(4) Estimated Prevalence</th>
<th>(5) Standard Error</th>
<th>(6) 90% Confidence Interval</th>
<th>(7) Projected N Direct Questioning</th>
<th>(8) Prevalence</th>
<th>(9) Under(Over)-Reporting</th>
<th>(10) P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Real Earnings Management - I am aware of a time in the past five years where my company has changed an operational activity to meet a near-term earnings target at the expense of long-term value.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List A(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List B(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lists A &amp; B(^b)</td>
<td>550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>140</td>
</tr>
<tr>
<td>List C(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List D(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lists C &amp; D(^b)</td>
<td>550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>140</td>
</tr>
<tr>
<td>Panel B: Disclosure Obfuscation - I am aware of a time in the past five years where my company altered a disclosure to make unfavorable information more difficult to understand.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List E(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List F(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lists E &amp; F(^b)</td>
<td>550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>140</td>
</tr>
<tr>
<td>Panel C: Accrual-based Earnings Management - I am aware of a time in the past five years where my company has used discretion within GAAP to report earnings that misrepresent the economic performance of the business.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List G(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List H(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lists G &amp; H(^b)</td>
<td>550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>140</td>
</tr>
<tr>
<td>Panel D: Material Omissions - I am aware of a time in the past five years where my company deliberately withheld information that, if known by investors, would discourage investment in the company.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List G(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List H(^a)</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lists G &amp; H(^b)</td>
<td>550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>140</td>
</tr>
</tbody>
</table>
Panel E: Fraud - *I am aware of a time in the past five years where my company materially misrepresented information in the financial statements with the intent to mislead users.*

<table>
<thead>
<tr>
<th>List</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>275</td>
</tr>
<tr>
<td>J</td>
<td>275</td>
</tr>
<tr>
<td>I &amp; J</td>
<td>550</td>
</tr>
</tbody>
</table>

Fraud

Table 2 reports the results of our planned primary analyses. Our primary dependent variable of interest for the list experiment conditions is participants’ report of the total number of statements in a list that are true for them (NumberTrue). NumberTrue is a discrete variable that ranges from 0 to 5 for lists that contain a sensitive statement (treatment lists) and from 0 to 4 for lists that do not contain a sensitive statement (control lists). Columns 2 and 3 report mean NumberTrue for treatment and control lists separately for each type of earnings manipulation (Panels A through E).

Our primary dependent variables of interest for our direct questioning condition are participants’ “true” or “false” responses to each sensitive statement. Thus, RealEarningsManagement, DisclosureObfuscation, AccEarningsManagement, MaterialOmissions, and Fraud are each equal to one if the participant confirms the relevant behavior, and zero otherwise. Column 8 reports mean of each dependent variable.

*a* Estimated prevalences correspond to the $\lambda_1$ in equation (1) $\text{NumberTrue} = \lambda_0 + \lambda_1 \times \text{Treatment} + \epsilon$

*b* Estimated prevalences correspond to the $\beta_1$ in equation (2) $\text{NumberTrue} = \beta_0 + \beta_1 \times \text{Treatment} + \beta_2 \times \text{List} + \epsilon$, SE are clustered at the individual level. Under(over)-reporting is computed by comparing list experiment and direct questioning prevalence estimates, and p-values are from a Wald test used to test whether the estimated prevalence differs between the indirect and direct estimation methods.
Table 3 presents the results of our planned cross-sectional analyses. Columns 1, 3, 5, 7, and 9 correspond to OLS estimates of equation (3), which relates to our list experiment approach, and columns 2, 4, 6, 8, and 10 correspond to logistic regression estimates of equation (4), which relates to our direct questioning approach. Standard errors are clustered at the individual level in estimates of equation (3). NumberTrue is a discrete variable that ranges from 0 to 5 for lists that contain a sensitive statement (treatment lists) and from 0 to 4 for lists that do not contain a sensitive statement (control lists), corresponding to participants’ report of the total number of statements in a list that are true for them. Manipulation is equal to RealEarningsManagement, DisclosureObfuscation, AccEarningsManagement, MaterialOmissions, and Fraud, in columns 2, 4, 6, 8, and 10, respectively. Each of these dependent
variables is equal to one if the participant confirms the relevant behavior in our direct questioning approach, and zero otherwise. *Treatment* is a binary variable equal to one if the response is from a treatment list, and zero otherwise. *List* is a binary variable equal to one for the first of the two relevant lists (A, C, E, G, and I), and zero otherwise. *ExecMale* is a binary variable equal to one if the participant identifies as male, and zero otherwise. *ExecTenure* is a discrete variable equal to one, two, or three, corresponding to whether the participant reports being at their current company “less than 4 years,” “4-9 years,” or “10 years or more,” respectively. *ExecCPA* is a binary variable equal to one if the participant reports being as a licensed CPA, and zero otherwise. *FirmAge* is a binary variable equal to one for observations where the participant reports a firm age greater than the median in our sample, and zero otherwise. *FirmSize* is a binary variable equal to one for observations where the participant reports revenue in the most recent fiscal year greater than the median in our sample, and zero otherwise. *FirmSectorIsTech* is a binary variable equal to one for observations where the participant reports working for a company in the “Technology[Software/Biotech]” industry, and zero otherwise.