Journal of Political Institutions and Political Economy, 2023, 4: 231–258

Where Are the Missing Dead? How Metrics Management Mitigates Official Data Misreporting in China

Guoer Liu*

Department of Political Science, University of Michigan, Ann Arbor, Michigan, USA; guoerliu@umich.edu

ABSTRACT

A perennial problem in executive politics is that agents charged with carrying out tasks have private information about their performance and corresponding incentives to manipulate this information. Conventional wisdom emphasizes external stakeholders as a source of reliable information. I instead focus on agents with competing interests and argue that when officials are subject to a relative performance evaluation metric, governing outcomes that are easily observable by their peers are more likely to be truthfully reported, while other outcomes are likely to be misreported. Using a difference-in-differences design and the varying reporting cost across sectors, combined with an original dataset on workplace accidents in China, I find that incorporating a "death cap" into the workplace safety metric management system has heterogenous effects on reported accidents. While the counts of casualties and accidents dropped by 33.8% and 29.4% respectively, the decline was entirely driven by sectors in which peer monitoring was not feasible. The results highlight the conditions under which agents are restrained by their peers. In contrast to theories that truthful signals occur when interests are aligned or independent oversight is

*I thank Iza Ding, Christian Fong, Mary Gallagher, Benjamin Goehring, Mai Hassan, Nahomi Ichino, Kenneth Lowande, Kevin Quinn, Fiona Shen-Bayh, Chuck Shipan, Yuki Shiraito, participants in the 2018 APSA annual conference presentation and comparative politics workshop at the University of Michigan, the members of the Ichino Research Lab, and two anonymous reviewers for the helpful comments and discussions on earlier drafts.

Online Appendix available from:
http://dx.doi.org/10.1561/113.00000077_app
ISSN 2689-4823; DOI 10.1561/113.00000077

present, intra-agent monitoring can get lower-level agents to reveal more truthful information when alternative mechanisms are weak.

Keywords: Bureaucracy; principal-agent theory; information revelation; causal inference; China

Introduction

How do principals get lower-level agents to reveal private information? Classic "sender-receiver" models show that truthful signals only occur when the principal and agents are wholly aligned (Crawford and Sobel, 1982; Gailmard and Patty, 2012). Principals may use incentives and penalties to align agents' interests with their own (Kaufman, 1960; Wilson, 1989; Pepinsky *et al.*, 2017), but often, self-reported data are used to determine performance. Perversely, incentives contingent upon achieving certain quantified targets risk encouraging agents to play the numbers game rather than increasing their effort. They hit the target on paper but miss the point in reality. In democracies, principals can turn to outside stakeholders such as the free media and local elected representatives to expose agent failures. For instance, US healthcare workers at a veterans hospital falsified wait-time data to circumvent quotas imposed by Congress (Slack, 2016). It was uncovered by investigative reports, whistleblowers, and lawmakers who received direct complaints. In non-democratic societies, alternative sources of information like these are limited or non-existent. Under what conditions, if any, do agents report their failures and report them honestly?

This paper argue that governing outcomes directly observable by those with competing interests will be truthfully reported—while other outcomes will be manipulated. I focus on how the disclosure of information occurs in a principal-agent relationship through the involvement of *peers*, who are fellow agents accountable to the same principal. Many existing studies implicitly assume a unitary agent: a principal devises incentives to circumscribe agents' behavior, and identical agents respond to incentives and maximize their payoff. But agents may have different preferences that cannot be reconciled (Carpenter, 2014; Potter, 2019). Incentives can engender a conflict of interest among otherwise identical agents and pit them against one another (Li and Zhou, 2005; Landry *et al.*, 2018; Chen and Hong, 2021). Strategic interactions among multiple agents can lead to a different dynamic in information revelation. Specifically, a decline in manipulation can be observed if agents are incentivized through a relative performance metric and if reporting on other agents' actual performances requires little effort.

Exploiting a phased adoption of a career incentive in the Chinese bureaucracy and the varying reporting cost across sectors, I use a difference-in-differences design to investigate the effect of peer monitoring on reported workplace accidents. The "death cap" assigns an upper limit on workplace accidents to executive bureaucrats—heads of government at the lower rungs of the administrative hierarchy (prefectures, counties, and townships). The reported accidents in each locale are ranked and evaluated annually by those bureaucrats' supervisors at one level above. Those who exceeded the cap or occupy the bottom rank are fined, barred from promotion, and prosecuted with criminal charges in some extreme instances. Furthermore, the varied protocols across sectors make reporting on fellow agents' performance in road traffic accidents comparatively more straightforward than in other sectors. The introduction of a death cap along with the relatively lower costs of reporting in the transportation sector presents an opportunity to examine the empirical implications of peer monitoring on information disclosure. That is, the reduction in reported accidents is smaller for traffic accidents as compared to other sectors.

Using an original event-based dataset on workplace accidents across sectors, I find that the death cap reduced reported workplace casualties by 33.8% and accident counts by 29.4%—both at statistically significant levels. Yet these declines were driven by sectors in which peer monitoring is not feasible—reported fatalities dropped by 42.4%, and accident counts dropped by 41.3%. In transportation sector where peer monitoring was viable, no statistically significant decrease was observed in either measures. Examining the closely related but less politicized insurance data, I show that the number of occupational injury claimants remains unchanged during the same period, despite the decline in reported workplace accidents. Taking advantage of the fine-grained data, I further distinguish peer competition from other observationally equivalent mechanisms that may account for the reported improvement or lack thereof. Complementary empirical strategies, different model specifications, and placebo tests were employed to invalidate the hypothesis, and the results are stable nonetheless.

Understanding how various actors and institutions might mitigate the sender-receiver problem has significant consequences for scholarship on bureaucratic control. My analysis shows that interest alignment between principal and agent is not a necessary condition for information revelation. Agent peer competition can also lead to the truthful reporting of private information. The idea is similar to Ting's (2003) findings in which principals facing distal agents benefit from the redundant assignment of bureaucratic functions. Likewise, the threat of whistleblowing induces compliance in hierarchical organizations (Beim *et al.*, 2014). As a form of accountability within the state apparatus, peer scrutiny is an alternative mechanism to narrow principal-agent informational asymmetries. Especially in countries with an underdeveloped civil society and interest groups, competing officials may be able to provide reliable information on the policy failures of others.

The analysis also contributes to the growing empirical research on suspicious official statistics. By examining the *changes* in reporting following an uneven incentive rollout, I rely on a different empirical approach to data manipulation. Estimating the *level* of data manipulation is incredibly challenging because the actual benchmark is unknown. Researchers have dealt with the problem by using alternative measurements (Wallace, 2016; Chen *et al.*, 2019; Cook and Fortunato, 2023), developing theoretical models (Nakamura *et al.*, 2016), and uncovering statistical irregularities (Ghanem and Zhang, 2014; Fisman and Wang, 2017; Carlitz and McLellan, 2020). While focusing on the changes can tell us little about the actual level of manipulation, it has the virtue of shedding light on factors that give rise to manipulated data.

The remainder of this paper proceeds as follows. In the next section discusses the theoretical implication of peer monitoring on information revelation. I then explain the institutional context of the death cap incentive in China. Next, I describe the identification strategy and data. The section after that presents the findings and engages with alternative explanations. Subsequent to the section on robustness checks, I end the paper with directions for future research.

Theory

Agent noncompliance is a thorny issue in any organization. Government bureaucrats, in particular, often work out of sight of their supervisors. The asymmetric information poses a threat to bureaucratic control, because how well agents perform on duty and what they produce cannot be directly observed (Wilson, 1989). Political elites, the principal in the principal-agent relationship, seek to devise incentives that align the agents' interests with their own (Crawford and Sobel, 1982; McCubbins and Schwartz, 1984; Moe, 2012). A popular approach nowadays is to use performance indicators to quantify the output of agents, and then link reward and punishment schemes to these indicators. This *metrics management* system promises improved performance because the deliverables can be measured, monitored, and controlled. As a result, bureaucratic accountability increases and shirking decreases.

In principle, introducing the metrics management system or incorporating a performance indicator into the system works because it sets the agents' priorities straight. A measured performance indicator means that the principal is clearly signaling that this task is more critical than other unmeasured tasks and agents are expected to exert extra effort to accomplish it. When multiple tasks compete for limited time and funds, resources should be allocated to ensure these goals are completed first.

Alternatively, metrics management can boost performance records while leaving the actual performance unaffected, particularly when the agents themselves report them. Because rewards and punishments are linked to the indicators, incentives intended to improve outcomes may inadvertently encourage agents to falsify the numbers (Sandefur and Glassman, 2014; Fisman and Wang, 2017; Jia, 2017; Pan and Chen, 2018; Carlitz and McLellan, 2020). For example, Eckhouse (2022) shows that the adoption of *CompStat*, a metrics management technique by police departments in 55 of 68 largest cities in the United States, is associated with those departments redirecting the workforce to activities that yield short-term observable output and manipulated crime statistics. Wallace (2016) demonstrates that Chinese growth statistics are more likely to be overstated when a political turnover is on the horizon. Ghanem and Zhang (2014) find that more than 50% of Chinese cities have reported suspicious air pollution data. Compared to the substantial professional and financial benefits, the cost of "juking the stats" was negligible and the probability of being caught was very low. Self-reported indicators are no longer a reliable measure to assess agent efforts under these circumstances.

The problem of data manipulation can be mitigated. In democracies, outside stakeholders such as the free press, interest groups, and local elected officials have been empowered and incentivized to pull "fire alarms" (McCubbins and Schwartz, 1984) and report agent failures. In autocracies, however, these alternative mechanisms are weak or absent. With limited means to monitor agents' behavior, the principal is caught in a dilemma. If the incentives are not powerful enough, the agents are not motivated to work, but too strong an incentive might result in data manipulation, defeating the purpose of metrics management. When will agents provide private information honestly if their self-reported data is used to determine their performance?

Peers—other agents who are accountable to the same principal—can play a crucial role in monitoring because, under some circumstances, they have both the *incentive* and *ability* to correct falsely reported performance information by other agents. Incentives hinge on the potential benefits that truthful reporting will deliver, and the ability to report limits the potential costs. Both factors affect agents' calculation and, consequentially, their reporting behavior. In particular, the incentive to report on others is high if agents are rewarded based on relative performance. Suppose performance is evaluated independently by the principal and the amount of reward is determined only by the output level. Peers' performance should not affect how much effort an agent exerts and how much she expects to receive in return. But when relative instead of marginal differences in output differentiates agent rewards (Lazear, 1985; Holmstrom, 1982), peer competition can lead to obstructive behavior. Agents may actively sabotage or withhold help when cooperation is crucial for others to succeed. When it comes to agent self-reported performance indicators, peer sabotage can, in fact, be positive: disclosing competitors' shortcomings or falsified records can provide an advantage to the reporting agent in the relative performance measurement. One agent's loss becomes another agent's gain.

The ability of peers to observe and report on one another is equally important. Imagine if all accident data are self-reported. Relative performance evaluation alone is not enough to make agents truthfully reveal information. If anything, everyone is likely to report numbers in their favor to outperform their competitors. This is the equilibrium identified in the existing literature on falsified official statistics (Wallace, 2022). Ability may arise from individual-level resources, experience, and even innate aptitude to identify irregularities in their competitors' records. But a systematic impact of ability is likely the result of structural opportunities that make monitoring others' output effortless and cost-effective.

When relative performance increases agents' incentive to report on others and structural opportunities are available for them to do so, peer monitoring can reduce data manipulation. Peers can be a compelling substitute for oversight by top-down forces or independent actors. The risk of being reported by her competitors could compel the reported agent to reveal truthful information that could otherwise be manipulated, thus forestalling others from exploiting her wrongdoing. Note that introducing peers leads to more information disclosure, not just from the competing agent but also from the culpable agent herself. Therefore, more truthful information will be revealed when agents have the incentives and ability to monitor their peers.

The Death Cap

Testing the causal effect of peer monitoring on information revelation requires variation in relative performance metrics and variation in agents' ability to observe and report on others. China's workplace death cap meets both criteria.

Incorporating the Death Cap into Metrics Management

Executive bureaucrats in China are reviewed annually and promoted by their supervisors one level above through the "Cadre Evaluation System" (CES) (Landry *et al.*, 2018).[1] CES is a metrics management system because bureaucrat performance is measured by quantified policy output in terms of how well their respective administrative jurisdiction[2] performs in different policy areas

[1]Different scholars use different terms, such as the Cadre Responsibility System, Target Responsibility System, or Responsibility System. I use this phrase because it is more intuitive.

[2]Although a large number of individuals—street-level bureaucrats, bureau chiefs, department heads—contribute to policy outcomes, CES refers to government heads as actors. Multiple actors within a unit are assumed to be aligned with the government head being evaluated.

during a given time. The performance is evaluated against that of others under the same supervisor, meaning that it is *relative* performance that matters (Zuo, 2015).

To combat the rising rate of workplace accidents, the CES introduced a death cap indicator to assess officials' performance in occupational safety in 2004 (The State Council, 2007). Examples of workplace accidents include coal mine accidents, factory fires, and road traffic accidents involving commercial transportation companies. Prior to the reform, workplace safety was far from a priority, and there was no standard measure for evaluating bureaucratic performance. Instead, economic development, investment and tax revenue growth, stability maintenance, and birth control were the critical measures. In the early 2000s, several large-scale coal mining accidents led to a public outcry over lax government regulations. In response, the national government released a series of decrees aimed at improving workplace safety. In January 2004, the State Council required provincial governments to provide quarterly reports on accidental fatalities to the State Administration of Work Safety (SAWS) at the national level (The State Council, 2004a). On February 3, the Work Safety Commission of the State Council issued a decree that links work safety to the annual CES evaluation of executive bureaucrats at each level (The State Council, 2004b).

In practice, the death cap assigns an upper limit for the number of workplace accidents and fatalities in each locale. If an administrative locale exceeds its annual cap, the local government head is subject to penalties that could jeopardize her political career. For exceptional cases that result in numerous casualties, the directive allows superiors to remove executive bureaucrats from their positions and file criminal charges (China Legal Publishing House, 2005).

Varied Monitoring Costs across Sectors

An important but often overlooked feature of the death cap is that the accident reporting protocols imply different costs for officials in reporting an accident, depending on whether they are "witness officials" or "liable officials." Articles 1.4 and 2.2 of the "National Work Safety Accidents Emergency Response Plan" distinguish two types of officials: *Witness officials* refer to those who serve in the location where an accident occurs (*shifadi*). Not only are they expected to report observed accidents to their supervisors, but they are authorized to participate in the accident investigation (The State Council, 2006).[3] *Liable officials* are those who serve in the location where the involved companies are registered (*shudi*). Casualties of an accident count toward the liable officials' death cap allowance in the CES. Once the allowance is exceeded, all

[3]The regulation was promulgated in 2006, and the period of the study was between 2000 to 2006. Official sources suggest that the national regulation's precedents were implemented at the local level (*People's Daily*, 2006) for earlier periods.

the negative consequences of violation ensue. Hence, liable officials are held accountable for an accident under the death cap, while witness officials are not.

The distinction in responsibility attribution means that cadres face different incentives for reporting different types of accidents. Especially for road traffic accidents, because vehicles move, accident reporting is likely to be *off-site*—accidents are observed and reported by witness officials rather than liable officials. If a traffic accident results in casualties, the number of deaths would count toward a liable official's cap. Witness officials, besides not having their performance record affected, have the authority to investigate accidents that happen in their jurisdiction. In other sectors, accidents usually occur where the involved companies are registered. The reporting is *on-site*: the officials who report an accident and those held accountable are the same individuals.

For officials, there are different calculations for on- and off-site reporting. Officials are incentivized to under-report or not report on-site accidents, knowing that honest reporting will fill their allowance and disadvantage them relative to their peers under the CES. However, witness officials are incentivized to truthfully report the off-site type because the regulation means that others will bear the cost. This rationale implies that death-cap-induced peer monitoring would have heterogeneous effects on reporting traffic accidents versus others. Of course, traffic accidents do occur in places where the vehicles are registered. In this case, witness officials and liable officials can be the same. Officials can also exchange favors—"I won't report you this time if you don't report me next time." Despite the complexities, the expected sectoral difference between traffic and other types of accidents should remain because on-site reporting has effectively been the only reporting system for other sectors.

In summary, the new incentive scheme and varied reporting cost allow me to causally estimate the effect of peer monitoring on reported accidents. Specifically, because peer monitoring is more prevalent in the transportation sector as compared to other sectors, it follows that the reduction in reported traffic accidents will be comparatively lower than that in other accidents. In the next section, I discuss the design and data in detail.

Research Design

Hypothesis and Identification Strategy

The phased implementation of the death cap provides rare opportunity to empirically test the *peer monitoring hypothesis*, peer monitoring leads to a smaller decline in the *reported* traffic accidents than other types of accidents. Under the assumption that outcome variables followed parallel trends after the introduction of death cap, the difference between the post-treatment and the pre-treatment outcome across sectors is attributable to the incentive, controlling for other contemporaneous changes.

Despite the introduction of the regulation by the national government in 2004, it was not implemented nationwide until 2006.[4] However, four provinces (Shandong, Jiangxi, Yunnan, and Guizhou) took rigorous measures to incorporate the State Council initiative into the provincial regulations in 2004, two years before everyone else. They disaggregated the provincial allowances and assigned them to prefectures and counties in local CES evaluation metrics (Wang, 2004c). Prefecture and county heads in the four provinces were subject to punishments if their jurisdictions exceeded their assigned death cap. Utilizing the temporal and spatial variations, a difference-in-differences (DD) design can estimate the effect of the incentive change on accident reporting. Additionally, the variation in reporting costs between the transportation sector and others allows me to examine the effect of peer monitoring on reported accidents.

Exactly why the four provinces adopted the death cap earlier than others remains unclear, which may lead to concerns over selection bias. There are no known initiatives from the central government to "experiment" in the four provinces. SAWS' official media outlet "China Work Safety News" only reported that the four provinces were the first to incorporate the death cap without explaining the rationale for their decisions (Wang, 2004b). For the four provinces, I did find prefecture- and county-level official documents supporting the claim that the death cap was implemented on the ground starting 2004.[5]

The identification of the average treatment effect on the treated (ATT) via DD relies only on the parallel trend assumption, not on the balanced covariates of the treated and the untreated group, which is the main attraction of this method (Imbens and Wooldridge, 2009, p.67). The quantity of interest is the effect of the death cap on those provinces that implemented them. Hence, the estimates tell us the *changes* in the reported accidents due to the incentive scheme, not the actual levels of accidents. While the parallel trend assumption cannot be tested directly, it can be made more plausible by carefully choosing

[4]This was the time that the Eleventh National Five-Year Plan incorporated the related work safety targets, that the death cap calculation formula was made public, and the National Bureau of Statistics decided to report the data regarding work safety accidents publicly (Wang, 2006).

[5]For example, see Wuhua District Government (2005). Media reports and provincial leaders' speeches at major work safety conferences corroborated the actual implementation in the four provinces (Li, 2004; Tang, 2005). Some sources suggest that other provinces adopted measures to improve work safety governance as early as 2000 (Wang, 2004a). However, precisely what they did was unclear, and the term death cap "veto target" was not used. If other places did adopt the death cap during the same period, it would have biased the effect downward.

the control group and appropriately adjusting for covariates, both of which I will explain further below.

Two types of selection problems can still lead to biased results (Stuart *et al.*, 2014). The first type originates from changing the group composition over time. In my application, this means that some of the provinces treated in 2004 chose to discontinue the death cap incentive before 2006, or untreated provinces decided to become part of the treated group prior to 2006. To my knowledge, neither of these scenarios is likely. The second type results from selection bias across groups. For example, if the pre-treatment characteristics associated with the dynamics of the outcome variables *over time* are not accounted for, such as changes in population, the parallel trend assumption is then violated (Abadie, 2005). This is a valid concern for my study.

I did the following exercises to mitigate the selection bias across groups. First, I collected information on variables that could impact both the decision to be an early adopter and the outcome variables, and incorporated them into all the specifications. Additionally, as a standard setup of the DD estimator, the two-way linear fixed effects account for unobserved unit-specific and time-specific confounders. Unobserved trends and time-invariant characteristics are accounted for.[6] Second, I also used multiple empirical strategies in later sections to rule out alternative hypotheses. Lastly, I re-did the analysis with different outcome variables in robustness checks.

Specification

With the following specification, I use a standard DD estimator with two-way fixed effects to examine the effect of death caps on reported workplace accidents and deaths:

$$Y_{it} = \alpha + D_{it}\beta + T_{it}\gamma + D_{it} \times T_{it}\theta + \mathbf{X}\boldsymbol{\delta} + \xi_t + \eta_i + \epsilon_{it}, \tag{1}$$

where i indicates provinces and t indicates years. Y_{it} is the outcome of interests, reported accident counts and death toll. D_{it} is a dummy that equals 1 for the four provinces in the treated group and 0 in the control group. T_{it} is a post-treatment time indicator. The interaction term $D_{it} \times T_{it}$ generates the coefficient of interest θ, which is the average treatment effect on the treated (ATT). $\mathbf{X}$ is a covariate matrix. ξ_t is the year fixed effect, and η_i is the province fixed effect. Year fixed effect captures the influence of aggregate trends in a particular year, such as leadership transition at the national level. The province fixed effect captures the unobserved time-invariant characteristics

[6] Other more refined methods, such as generalized synthetic control (Xu, 2017), trajectory balancing (Hazlett and Xu, 2018), and panel matching (Imai *et al.*, 2021) cannot be employed here due to the small time-series cross-sectional data. There is not enough information to satisfy the modeling assumptions.

specific to each province. Throughout this paper, I use robust standard errors clustered at the province to account for autocorrelation.

Data

Outcomes

I compiled an original event-based dataset on reported workplace accidents in China between July 11, 2000, and December 31, 2005. The raw data contains 13,382 entries in total. Each entry records the time, location, technical cause, and number of deaths. Because the treatment took place at the province level, I collapsed the data by province-year, resulting in an unbalanced panel dataset with 176 observations. The outcome variable is the logged-sum[7] of the reported death toll and accident count in province i and year t. To test the heterogeneous effect, I further disaggregated the event data by traffic accidents and others. Online Appendix A discusses in detail the data collection, coding procedures, and why this source is preferable to alternatives for this study.

Covariates

As noted earlier, because the treatment assignment process is unknown, provinces may self-select into the treated group due to province-specific or time-varying attributes. Both systematic and idiosyncratic reasons affect the provinces' inclination to adopt the death cap. Systematically, *population size* and *coal production* turned out to be important criteria for determining an administrative unit's death cap in some places. I used the provincial population's logarithm as a proxy for population size, and the annual coal production volume in every province as a measure for the latter. Additionally, whether a province has the capacity to improve work safety governance should influence its decision to be an early adopter. I included *economic development* and *tax revenue per capita* to account for the variation in state capacity. In particular, following Landry et al. (2018) I used the logarithm of night light data as a proxy for economic development to avoid concerns over unreliable GDP statistics. Also, the willingness of managers and workers to comply with stringent safety protocols should affect the adoption schedule. I added *rural population*, the percentage of the rural population, as a proxy for human capital in each province. Because provinces with lower human capital may experience more workplace accidents, governors may be pressed to adopt the incentive earlier. I also considered *sectoral wages*, the average yearly salaries in the mining industry and transportation section. It measures labor quality, which is a proxy for the potential for improving work safety in specific sectors.

[7] I used raw data as part of the robustness check.

Idiosyncratic reasons, such as the characteristics of provincial leaders (provincial party secretaries and governors—the most influential political figures in a province), also play a crucial role. Research shows that officials' attributes affect their policy orientation and even the tendency to manipulate data (Jiang and Wallace, 2017). Provinces with leadership turnover on the horizon may devote more resources to keeping the performance records in order to boost the leaders' promotion prospects (Wallace, 2016). These provinces may promulgate stringent work safety regulations or adopt measures to incentivize lower-level officials' performance. I collected information on provincial party secretaries' and governors' *total years in office at their current position*, *age*, and *whether they were promoted at the beginning of the following year* to control the variation.

Covariate information comes from *China Statistical Yearbooks*, the *China Coal Industry Yearbooks*, the *Labour Statistical Yearbooks*, and the existing dataset collected by other researchers. The summary statistics of all variables are presented in Online Appendix Table A2.

Findings

Overall, I find evidence supporting the peer monitoring hypothesis: While the death cap reduces reported accidents, the reduction is concentrated in non-traffic sectors. For traffic accidents, there is no statistically significant reduction in reported deaths or accidents. Figure 1 summarizes the estimated DD coefficients. Those coefficients indicate the *change* in the outcome variables within the treated and control provinces. In aggregate, the death cap rule reduces the reported casualties by 33.8% ($\exp(-0.412)$) and the reported accident counts also drop by 29.4% ($\exp(-0.348)$). Both results are statistically significant at the 5% level. Comparing traffic accidents with other sectors reveals the heterogeneous effect of death caps on reporting. The death cap does not lead to any statistically significant change in reported casualties or accident counts for traffic accidents, regardless of how the model is specified. For other sectors, however, the death cap reduces the reported casualties by 42.4% ($\exp(-0.552)$) and the accident counts by 41.3% ($\exp(-0.532)$). The results are statistically significant at 1% and 5%, respectively.[8] The contrast suggests that the reduction in other sectors alone drives the aggregate decline following the introduction of the death cap. Next, I evaluate additional empirical data against alternative explanations to the pattern that I discovered.

[8] Online Appendix Table A3, Table A4, and Table A5 are the corresponding the regression tables.

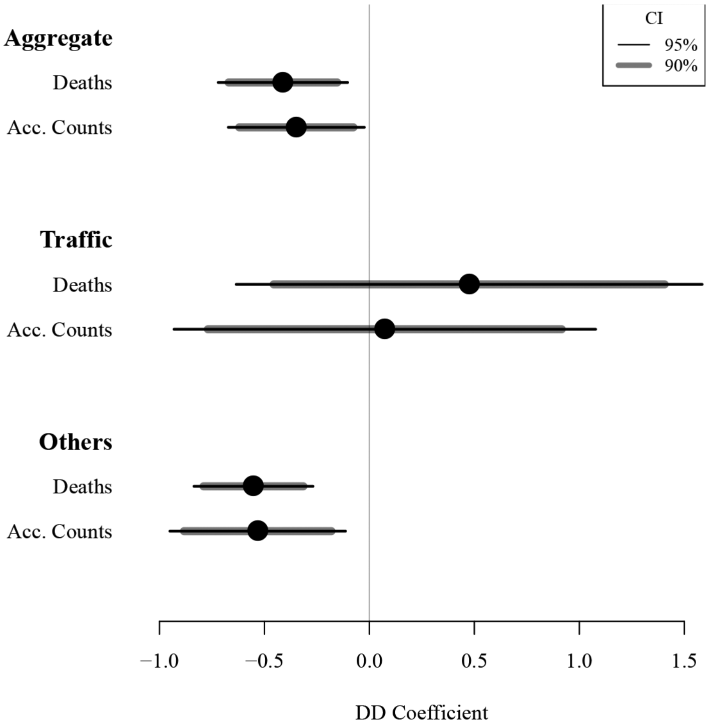


Figure 1: The DD Estimates of the Effect of the Death Cap on Aggregate Reported Workplace Accident Deaths and Accident Counts, on Traffic Accidents, and on Accidents in Other Sectors. The wide horizontal bars represent 90% asymptotic confidence intervals, and narrow horizontal bars represent 95% confidence intervals. The vertical bar marks zero.

Timing: Abrupt Improvement in Governance?

Implicit in the peer monitoring hypothesis is that the reported traffic accident is more reliable than other accidents, as agents are likely to observe their peers' traffic accidents and are incentivized to report them truthfully. Although it is impossible to determine the true number of workplace accidents, I can speculate what a genuine improvement in occupational safety and truthful reporting look like. Using common sense, we know that genuine improvement in governance is unlikely to occur *immediately* after adopting a new evaluation standard because it often takes time for regulatory adjustments to yield noticeable change. Hence I expect to see a flat line of reported accidents around the death cap announcement if the data are free from manipulation.

I inspected a 16-month trend around its announcement on February 3, 2004. While the focus is February, a longer time window is preferred for two reasons. The first is that this allows us to show a year-over-year comparison. Second,

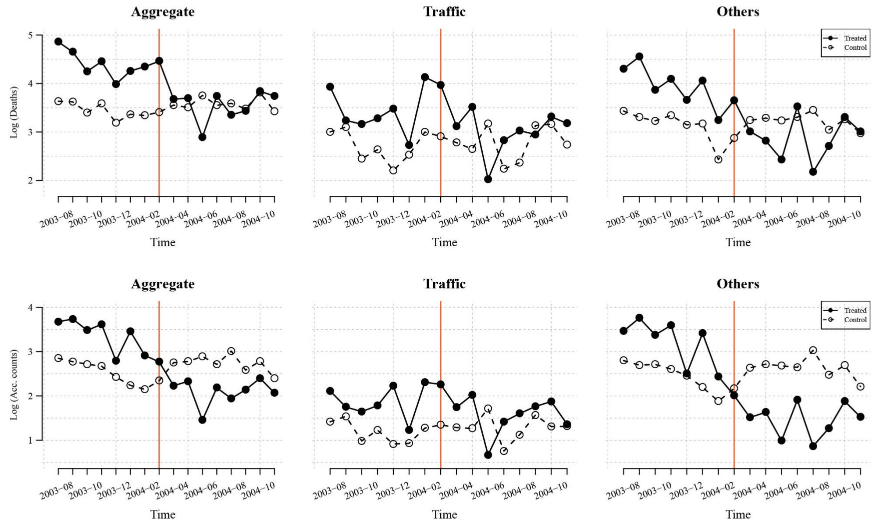


Figure 2: 16-Month Trend of Reported Workplace Accident Deaths and Accident Counts around the Death Cap Announcement on February 3, 2004 for the Treated and Control Groups. The outcome variables for the three figures at the top are accident deaths. The outcome variables for the figures at the bottom are accident counts. The vertical lines mark the announcement of the death cap in February 2004.

a more extended time frame could alleviate the problem of "political cycle" reporting. Researchers find that many government activities and indicators in China fluctuate with major political events (Nie *et al.*, 2013; Truex, 2016). The death cap announcement coincided with the annual National People's Congres and the National Committee of the Chinese People's Political Consultative Conference. A longer time frame mitigates the impact of these major political events on reporting.

Figure 2 plots casualties and accident counts for all accidents (left panel), traffic accidents (middle panel), and accidents in other sectors (right panel). The drop happened immediately after the announcement of the death cap in the treated group. However, the statistics for traffic accidents bounced back almost to the initial level over time. I did not observe any noticeable change in the control group. The immediate drop in reported accidents around the announcement window adds to the evidence against genuine improvement.

Borderline Cases

Public eyewitnesses are an obvious alternative explanation for the heterogenous effects on reported accidents. Because traffic accidents usually occur in public spaces, citizens—apart from competing officials—can observe and report the accident. This concern is mitigated for three reasons. First, filing an accident as a workplace safety accident that counts towards the death cap differs from reporting an accident to the local police. Access to the former is restricted to government officials. Second, the time period of this research predates the Internet era and citizen journalism in China. These unconventional methods of exposure were limited. Third, while government officials' reporting behavior may be affected by incentives, it is likely that citizens' propensity to report accidents will remain unchanged. I find that reported borderline accidents, i.e., those fall near the cutoff of the reporting protocol, exhibit a distinct pattern in the treated province. I elaborate this idea below.

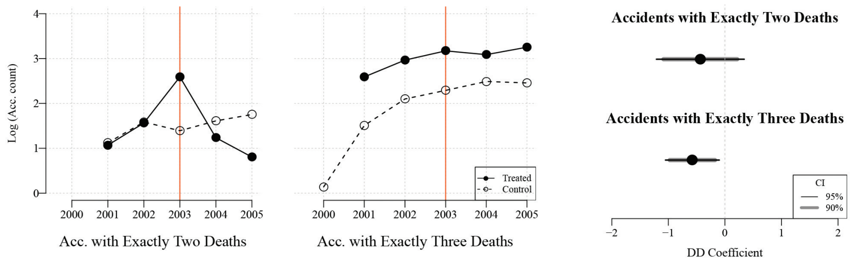
Depending on the scale of the accident, the death cap designates authorities at various levels to supervise post-accident investigations. Accidents are categorized differently according to a casualty threshold: *minor accidents* are those with two or fewer deaths; *major accidents* are those with three to nine fatalities; *serious accidents* are those with more than ten casualties.[9] Counties and district-level governments have the full authority to deal with the aftermath of minor accidents. For accidents with three or more deaths, authorities above the county level have to be notified. Once informed, the higher-level authorities are expected to play an active role in the investigation (The State Council, 2004b).[10] If they are found to have covered up an accident, they are subject to penalties by their superiors. Here I focus on the count of reported accidents with *exactly two deaths* and *exactly three deaths*, the lowest threshold for upward reporting.[11] Under official reporting, these borderline cases are more likely to be manipulated. Under citizen reporting, I should observe no difference in accidents below or above the threshold.

Figure 3a plots the accident counts for the two categories. Accident counts with two deaths (just below the threshold) fall sharply in the treated group, but there is no such decline in the control group. There is no apparent drop in accidents with three deaths (just above the threshold). Figure 3b presents the DD estimates for the reported borderline cases with the full specification. Overall, the negative coefficients suggest a different reporting pattern for both types of accidents in the treated provinces. The estimate is statistically significant for accidents with exactly three deaths but not for those with two deaths.

[9]The three-level categorization is in document GB/T15236-1994 issued by the State Bureau of Technical Supervision in 1994. This rule was effective between 1995 to 2007. In 2007, the three-level classification was revised with a new threshold: "serious accidents" are those with 10 to 29 casualties, and accidents that result in more than 30 deaths are categorized as "especially serious accidents".

[10]Online Appendix Figure A2 illustrates the government structure and workplace accident reporting protocols.

[11]In the event-based dataset, there were about 44% of cases with three or more deaths, and only about 4% that resulted in ten or more deaths. Because fewer observations lead to more noise, a lower cutoff is preferred.



(a) Workplace Accident Counts with *Exactly Two Deaths* and *Exactly Three Deaths* in the Treated and Control Group.

(b) The DD Estimates of Reported Borderline Workplace Accidents.

Figure 3: Reported Borderline Workplace Accidents in the Treated and Control Group. Missing data points indicate no such accident was recorded in that year in the data set. In panel (a), the solid lines connect the annual average for the treated group. The dash lines connect the annual average for the control group. The vertical lines separate the pre- and post-treatment period by marking 2003, which is one year before the treatment period. In panel (b), all the solid circles represent DD estimates with control variables, year and province fixed effects.

Multiple factors could give rise to the non-significant result. One potential cause is the unbalanced panel dataset. Because I focused on accidents with exactly two or three deaths, many observations dropped out because they did not meet the criteria. Fewer observations result in larger standard errors of the estimates. Specifically, accidents resulting in two deaths have missing outcomes in more than one-third of province-years, while those with three deaths have about one-sixth of outcomes missing.[12] Nevertheless, it is reassuring that the direction of the estimates is robust to different model specifications. Online Appendix Table A6 presents all the regression results. Compared with the control group, both borderline cases just above and below the threshold are likely to decrease. Hence, the result is more consistent with the hypothesis on peer reporting than citizen reporting.

Expanding Transportation Sector?

The rapid expansion of the transportation sector is another confounding factor. If the transportation sector is outpacing the government's capacity to regulate it, then the absence of a decline in traffic accidents may already be a genuine, improved outcome. For example, if increasing numbers of vehicles are sold and used during the research period, then traffic accidents should increase too. The stable trend I observed is attributable to improved work safety governance without which it could have been much worse.

An implicit assumption of this argument is that improvements in regulations, safety technology, and transportation infrastructure falls behind the speed of

[12] For accidents with two deaths, 61 out of 176 province-year are missing. For those resulting in three deaths, 31 province-year are missing.

market expansion. Whether it is true is an empirical question that goes beyond the scope of this paper. Here I provide two additional pieces of information that help us better appreciate the context. First, reckless driving, a leading cause of traffic accidents in China, became a criminal offense with the enacting of the first traffic safety law in May 2004 (O'Meara, 2020). The timing coincided with the announcement of the death cap incentive. If anything, the legal change should have decreased the likelihood of traffic accidents, *ceteris paribus*. Second, other data sources on traffic accidents and the death toll did not show an upward trend during the research period (Ma *et al.*, 2012). Thus, the evidence does not support the competing view.

The "Ungovernable" Transportation Sector

In addition to regulatory agencies, improving work safety outcomes requires efforts by factory managers and workers, and other conditions need to fall in line as well. The transportation sector, unlike others, is more likely to suffer from random shocks like extreme weather conditions. The death toll of traffic accidents may have a smaller variance due to vehicles' fixed carrying capacity. If these known and unknown sector-specific characteristics make traffic accidents less responsive to improved official regulatory measures, I will observe no improvement despite the governments' best efforts.

To assess this alternative explanation, I looked into the *average deaths per accident* across sectors, which is the total deaths divided by the total accident counts in a sector in a given year. The measure is meaningful because it captures three channels through which genuine improvement occurs. First, if I keep the denominator accident counts unchanged, reducing the total casualties will reduce the average. Second, improvement can occur by breaking down a workplace safety catastrophe into minor accidents. That is, by holding the total deaths constant but increasing the number of accidents, the average decreases, too. Lastly, the average drops when accident casualties decrease and the counts increase. Thus, if some sector-specific characteristics make the transportation sector inherently challenging to govern and improve, the average death per accident should remain constant post-treatment. In contrast, other sectors should experience a *decline* if they are more responsive to the stricter regulatory measures.

Figure 4 presents a different pattern.[13] The average death per accident remains relatively stable for traffic accidents, but it *increases* for other sectors in the treated group. One plausible explanation is that the post-treatment death toll increases in other sectors, and another is the post-treatment accident

[13] Accidents with more than 30 deaths were dropped due to their rare occurrence but large impact on the average. I choose 30 because the 2007 regulation sets 30 as the threshold for "especially serious accidents," the most severe category. In the data set, accidents that resulted in 30 or more casualties account for about 0.005% of all cases.

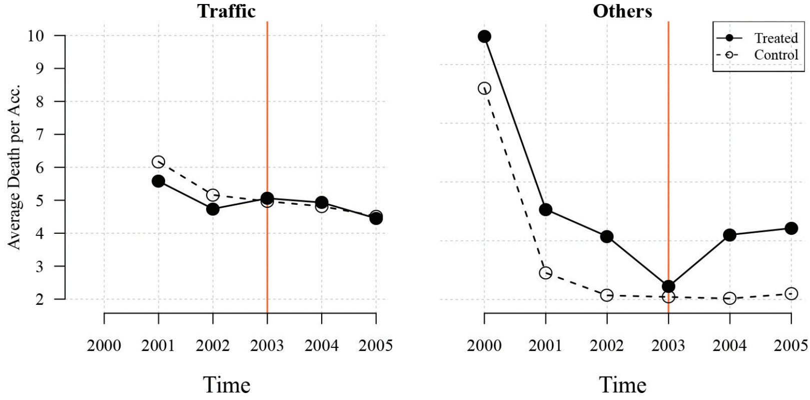


Figure 4: The Reported *Death per Accident* for Traffic Accidents and Accidents in Other Sectors in the Treated and Control Group. Missing data points indicate that no such accident was recorded in that year in the data set. The vertical line separates the pre-and post-treatment period by marking 2003, which is one year before the treatment period.

counts decrease. The former contradicts the empirical evidence of a significant decline in accident deaths. The latter is consistent with the evidence. But as I demonstrated earlier, the reduction tends to concentrate on minor accidents that are vulnerable to manipulation. The divergent pattern lends credence to the hypothesis that the improvement was due to manipulation rather than unobservable sector-specific characteristics.

Alternative Source: Occupational Accident Insurance

Reported accident statistics are prone to manipulation because the numbers are a metric management tool that determines officials' payoffs. Unlike accident counts, occupational insurance statistics are not subject to intense political scrutiny and therefore are less likely to be manipulated. Insured workers who suffered from occupational injuries have strong incentives to claim compensation. If they make a claim, government agencies will likely keep the record because the compensation is drawn from public funds.

Theoretically, the frequency of workplace injuries should positively correlate with accident counts and casualties. If the reported accidents are truthful, I would observe a decline in insurance claimants during the same period. Figure 5 shows the percentage of work injury insurance beneficiaries for the treated and control group, respectively. The trend remained stable after the adoption of the death cap. Departing from the pattern in reported accidents, the treated group did not experience any significant decline in workplace injuries. The most stringent DD estimator's coefficient of interest is -0.003 (Std. Error

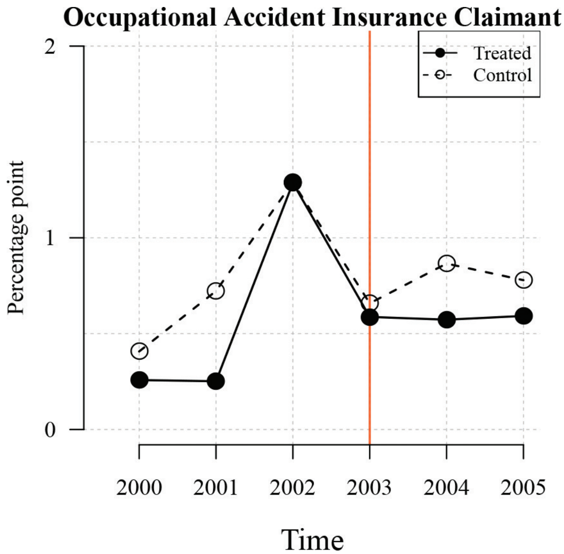


Figure 5: The Normalized Occupational Insurance Beneficiaries in the Treated and Control Group. The vertical line separates the pre-and post-treatment period by marking 2003, which is one year before the treatment period.

0.003).[14] This is suggestive evidence that the drop in reported accidents may be exaggerated.

Complementary Interviews

Making an inference about data manipulation is intrinsically difficult because the benchmark of truthful data is unknown. To complement the statistical analysis, I conducted forty semi-structured interviews on the topic of interest with government officials, scholars, and factory managers in Shanghai and Sichuan in 2017. Although the interviews were far from a representative sample,[15] these informative conversations showed how workplace accident reporting works in practice.

During an interview, the deputy head of a county SAWS with over twenty years of experience in this area remarked, "The first thing we check when we arrive at the scene of a traffic accident is where the vehicles are registered. This is important because it determines who will be held accountable for the accident—us or officials elsewhere."[16] According to him, the off-site reporting

[14]Insurance data by sectors would be ideal, but that breakdown is not available.

[15]I relied on snowball sampling to get the contact information of interviewees, starting with a few scholars and local officials. See Online Appendix A.5 for more details.

[16]Interview: YXsh90507102.

rules were followed by local agents in the early 2000s, even though officials and business owners considered the system "ridiculous" and "not reasonable at all."[17] They often attributed factors beyond their control, such as poor road conditions in the "witness" locale, as the primary cause of the traffic accidents.[18]

Strategic reasoning also underscores borderline case reporting. According to a former department head in a prefecture-level SAWS, "The unspoken rule (*qian guize*) is that as long as authorities at a certain level are informed of an accident, officials below them will definitely be punished. This is higher-level authorities' (*shangji lingdao*) way of protecting themselves from being disciplined by their supervisors."[19] This statement was cross-validated by multiple other interviews.

It is worth acknowledging that if officials are willing to manipulate statistics to make a good impression on their superiors, it is unlikely that they will admit to any wrongdoing to researchers, even in private. As expected, none of the officials in the study did so. But the interviews provided valuable insights from officials who are at the forefront of implementing the death cap and lent credibility to the main hypotheses.

Robustness Checks

The Parallel Trend Assumption

A fundamental identification assumption of the DD estimator is that the outcome variable for the treated and control units follows the same trend post-treatment. In the absence of the death cap, reported accident counts and death tolls in the treated and control groups should remain constant over time. Only if this assumption is satisfied can untreated units provide a valid counterfactual for the treated ones. While this assumption is not testable, Figure 6 presents the dynamic event studies plot for all outcome variables, which graphically illustrates the point estimates and 95% confidence interval for each time period. The results are estimated from a fully-saturated model with all possible lags and leads (Clarke and Tapia-Schythe, 2021). The point is to check if the coefficients on the pre-treated periods are statistically *in*significant. While most estimates are reassuring, I do observe several significant differences in lag four for aggregate deaths and accident counts, as well as accident counts for other sectors. Note that these lags are the farthest away from the treatment

[17] Interview: XGsh22507102, MMsh90507102.

[18] Interview: YXsh90507102. According to my data set, however, many traffic accidents are due to passenger and cargo overloading. These causes are attributable to lax government oversight in places where the transportation companies are registered, and not necessarily to poor conditions where accidents occured.

[19] Interviews: DYsh12307102, XGsh22507102.

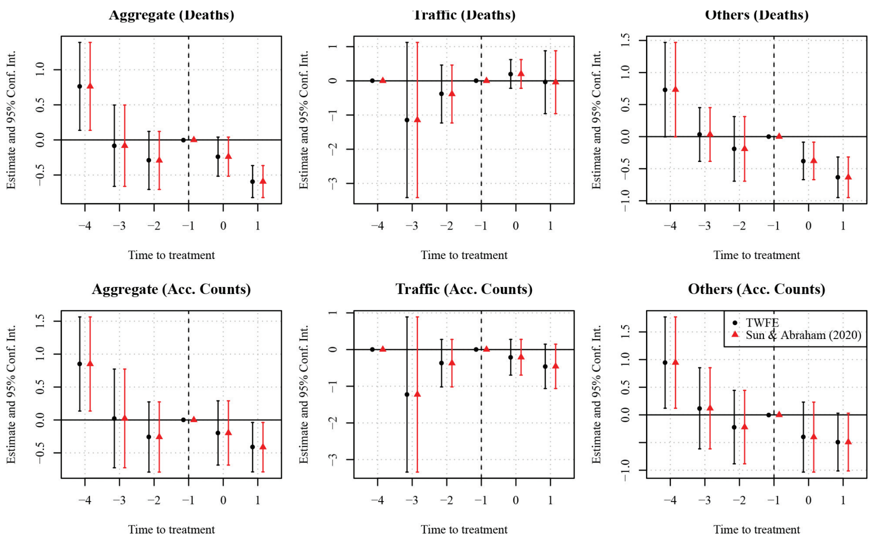
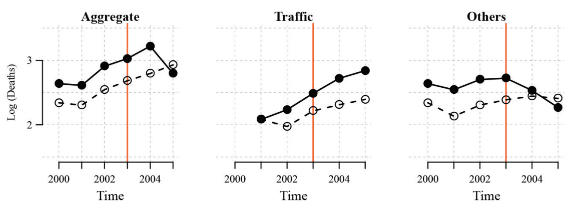
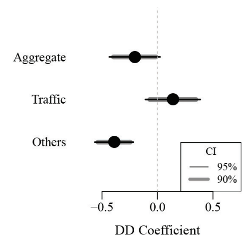


Figure 6: Dynamic Event Studies Plot with Standard TWFE and Cohort-specific Effects. TWFE point estimates are plotted along with the 95% asymptotic confidence intervals in black. Sun and Abraham (2021) point estimates are plotted along with the 95% asymptotic confidence intervals in red. The vertical dash line indicates the baseline (omitted) period, which is one year before the death cap adoption in the four provinces. For traffic accidents, the -4 period data is missing. The outcome variables for the three figures at the top are workplace accident deaths. The outcome variables for the figures at the bottom are workplace accident counts.



(a) Parallel-Trend Visualization in the Treated and Control Group for the Prefecture-Year Analysis.



(b) The DD Estimates of the Prefecture-Year Analysis.

Figure 7: Prefecture-Year Analysis. Panel (a) plots workplace accident deaths in aggregate, traffic, and other sectors. Missing data points indicate no such accident was recorded in the dataset. The solid lines connect the annual average for the treated group. The dash lines connect the annual average for the control group. The vertical line separates the pre-and post-treatment period by marking 2003, which is one year before the treatment period. In panel (b), all the solid circles represent DD estimates with TWFE. The wide horizontal bars in panel (b) represent 90% asymptotic confidence intervals, and the narrow horizontal bars represent 95% confidence intervals. The vertical line marks zero.

year during the research window, and not all the treated units are observed. Some compositional changes may correlate with the differences. Sun and Abraham (2021) argue that the standard two-way fixed effects (TWFE) are vulnerable to problems when there is a staggered treatment rollout. This is not a problem for this study because the setup is relatively simple. But following Sun and Abraham (2021), I estimated the aggregated "cohort-specific" effects. The results were identical. Online Appendix Figure A3 also plots the treated and control group average for all outcome variables. The trends between the treated and control groups were heading in the same direction before 2004.

Prefecture-Year Analysis

As a robustness check, I aggregated the event data to prefecture-year and redid the analysis with a simple TWFE specification. Control variables are not included because many covariates at the prefecture-level are not available during the research period. Figure 7 plots the outcome variable and coefficients. The heterogeneous effects across sectors are consistent with the main results and are statistically significant at conventional levels. Online Appendix Section C.2 presents the regression results with prefecture-year and province-month as units of analysis and discusses the related issues in detail.

Other Tests

I also performed a placebo test to show that the main result is not a "luck" draw from all possible combinations of treated and control group assignments. In addition, I repeated the same analysis with the raw data (without logarithm transformation of the outcome variables). The results largely remain consistent with the main hypotheses. More details can be found in Online Appendix Section C.3 and Section C.4, respectively.

Conclusion

This paper explores the effect of a metrics management incentive on reported workplace accidents in China. The introduction of the "death cap"—an upper limit for workplace accident counts and total fatalities in each locale—as a career incentive metric in China has heterogeneous effects on reported accidents. While there is a statistically significant and substantively meaningful decline in aggregated accidents and the death toll, the decline is driven only by sectors where agent cross-reporting is costly. There is no reduction in traffic sector where agents can observe others' outcomes with ease. In contrast to other studies that suggest high-stake incentives lead to manipulated statistics, I argue that when incentives enable peer monitoring, governing outcomes that are observable by those with competing interests are less likely to be manipulated.

Although China is the empirical reference, the findings have important implications for research on bureaucratic control more broadly. First, strategic interaction among multiple agents affords a different perspective in approaching the problem of informational asymmetries in the principal-agent relationship. The classic research on bureaucracy offers rich descriptive accounts of the complexities of diverse interests in seemingly mundane bureaucratic operations, regardless of political system (Hough and Fainsod, 1979; Wilson, 1989; Lieberthal and Oksenberg, 1988). As more recent research has shown, strategic interaction among agents affects how they respond to directives from above: agents' collusive, imitative, and adaptive behavior influences what information is revealed by agents, as well as when and how much in formation is revealed (Whitford, 2003; Zhou, 2010; Rundlett and Svolik, 2016). This study suggests competing agents who are capable of observing each other's performance are incentivized to disclose more honest information.

Second, lacking truly independent institutions, when do agents truthfully report decidedly unfavorable information like workplace accidents and deaths in authoritarian contexts? This paper explains how and when intra-agent competition mitigates the information problem in the most challenging circumstances, highlighting the possibility that principals can design institutions to harness the competition to solicit more information. A growing literature has explored various aspects of executive politics in non-democracies and emerging democracies. While bureaucratic organizations in these contexts do not conform to the Weberian legal-rational ideal and are often perceived as weak or corrupt, they are capable of helping their leaders meet key policy and political goals (Ang, 2016; Mattingly, 2019; Brierley, 2020; Hassan, 2020; Ding, 2020). Apart from the synergy between political principals and administrative agents, the interaction among a multitude of local underlings is an important piece of the puzzle. Intra-agent competition may be less visible compared with electoral competition, but it has significant consequences for elite management, tax revenue collection, and the well-being of ordinary citizens (Lü and Landry, 2014; Shi and Xi, 2018; Chen *et al.*, 2020).

The study also suggests directions for future research. A limitation of this analysis is that the actual number of accidents and death tolls are unobservable. Also, the source of the reporting, who registered each accident entry, is not identified. I tried to engage with selection problem and alternative explanations and explain why I could, with caution, interpret the difference as manipulation and attribute it to the threat of peer monitoring. More fined-grained qualitative data would allow us to distinguish the mechanisms: whether monitoring works through the channel of reporting on others, self-reporting, or both. The analysis will benefit from a comprehensive dataset covering a more extended period. For example, an outstanding question is how long the effect of peer monitoring on information revelation lasts. In other words, is the incentive-induced intra-agent monitoring a transient phenomenon? After the initial shock, agents may quickly determine that collusion is an optimal strategy, and the equilibrium will change accordingly. In fact, Fisman and Wang's (2017) careful analysis of the same incentive scheme suggests this is likely to be the case.[20] Using the province-level quarterly reported deaths in *People's Daily* and the corresponding death cap between 2005 and 2012, they uncovered a sharp discontinuity in the reported data at the prespecified "death ceiling" threshold with McCrary's density test, but there is no distinct sectoral difference during that time frame. Understanding the conditions under which incentive-induced monitoring can be sustained will help scholars better appreciate the limitations and promises of peer monitoring in bureaucratic behavior.

References

Abadie, A. 2005. "Semiparametric difference-in-differences estimators". *The Review of Economic Studies*. 72(1): 1–19.

Ang, Y. Y. 2016. *How China Escaped the Poverty Trap*. Cornell University Press.

Beim, D., A. V. Hirsch, and J. P. Kastellec. 2014. "Whistleblowing and compliance in the judicial hierarchy". *American Journal of Political Science*. 58(4): 904–918. DOI: 10.1111/ajps.12108.

Brierley, S. 2020. "Unprincipled principals: Co-opted bureaucrats and corruption in Ghana". *American Journal of Political Science*. 64(2): 209–222. DOI: 10.1111/ajps.12495.

Carlitz, R. D. and R. McLellan. 2020. "Open data from authoritarian regimes: New opportunities, new challenges". *Perspectives on Politics*. June: 1–11. DOI: 10.1017/S1537592720001346.

Carpenter, D. 2014. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton University Press.

Chen, T. and J. Y. Hong. 2021. "Rivals within: Political factions, loyalty, and elite competition under authoritarianism". *Political Science Research and Methods*. 9(3): 599–614.

Chen, W., X. Chen, C.-T. Hsieh, and Z. Song. 2019. "A forensic examination of China's national accounts". *Tech. rep.* National Bureau of Economic Research.

Chen, X., B. Huang, and Y. Yu. 2020. "Peer effect, political competition, and eco-effciency: Evidence from city-level data in the People's Republic of China". *Tech. rep.* ADBI Working Paper Series.

[20]Fisman and Wang (2017) call the incentive scheme as "death ceiling program".

China Legal Publishing House. 2005. *The Code of Workplace Safety (New Edition) [Xinbian anquan shengchan shiyong fadian]*. China Legal Publishing House.

Clarke, D. and K. Tapia-Schythe. 2021. "Implementing the panel event study". *The Stata Journal*. 21(4): 853–884.

Cook, S. J. and D. Fortunato. 2023. "The politics of police data: State legislative capacity and the transparency of state and substate agencies". *American Political Science Review*. 117(1): 280–295.

Crawford, V. P. and J. Sobel. 1982. "Strategic information transmission". *Econometrica*. 50(6): 1431–1451.

Ding, I. 2020. "Performative governance". *World Politics*. 72(4): 525–556.

Eckhouse, L. 2022. "Metrics management and bureaucratic accountability: Evidence from policing". *American Journal of Political Science*. 66(2): 385–401.

Fisman, R. and Y. Wang. 2017. "The distortionary effects of incentives in government: Evidence from China's "Death Ceiling" program". *American Economic Journal: Applied Economics*. 9(2): 202–218.

Gailmard, S. and J. W. Patty. 2012. "Formal models of bureaucracy". *Annual Review of Political Science*. 15(1): 353–377. DOI: 10.1146/annurev-polisci-031710-103314.

Ghanem, D. and J. Zhang. 2014. "'Effortless perfection:' Do Chinese cities manipulate air pollution data?" *Journal of Environmental Economics and Management*. 68(2): 203–225. DOI: 10.1016/j.jeem.2014.05.003.

Hassan, M. 2020. *Regime Threats and State Solutions: Bureaucratic Loyalty and Embeddedness in Kenya. Cambridge Studies in Comparative Politics*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108858960.

Hazlett, C. and Y. Xu. 2018. "Trajectory Balancing: A General Reweighting Approach to Causal Inference With Time-Series Cross-Sectional Data". *SSRN Scholarly Paper* No. ID 3214231. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.3214231.

Holmstrom, B. 1982. "Moral Hazard in Teams". *The Bell Journal of Economics*. 13(2): 324–340.

Hough, J. F. and M. Fainsod. 1979. *How the Soviet Union Is Governed*. Harvard University Press.

Imai, K., I. S. Kim, and E. H. Wang. 2021. "Matching methods for causal inference with time-series cross-sectional data". *American Journal of Political Science*.

Imbens, G. W. and J. M. Wooldridge. 2009. "Recent developments in the econometrics of program evaluation". *J. Econ. Lit.* 47(1): 5–86.

Jia Ruixue Nie, H. 2017. "Decentralization, collusion, and coal mine deaths". *Review of Economics and Statistics*. 99(1): 105–118.

Jiang, J. and J. Wallace. 2017. "Informal Institutions and Authoritarian Information Systems: Theory and Evidence from China". *SSRN Scholarly Paper* No. ID 2992165. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.2992165.

Kaufman, H. 1960. *The Forest Ranger: A Study in Administrative Behavior*. Routledge.

Landry, P. F., X. Lü, and H. Duan. 2018. "Does performance matter? Evaluating political selection along the Chinese administrative ladder". *Comparative Political Studies*. 51(8): 1074–1105.

Lazear, E. 1985. "Relative performance, sabotage and industrial politics". *Tech. rep.* University of Chicago.

Li, H. 2004. *Yunnan Establishes Work Safety Veto Targets, Negligent Officials Will Be Asked to Resign*. Xinhua News.

Li, H. and L.-A. Zhou. 2005. "Political turnover and economic performance: The incentive role of personnel control in China". *Journal of public economics*. 89(9–10): 1743–1762.

Lieberthal, K. and M. Oksenberg. 1988. *Policy Making in China: Leaders, Structures, and Processes*. Princeton University Press.

Lü, X. and P. F. Landry. 2014. "Show me the money: Interjurisdiction political competition and Fiscal extraction in China". *American Political Science Review*. 108(3): 706–722. DOI: 10.1017/S0003055414000252.

Ma, S., Q. Li, M. Zhou, L. Duan, and D. Bishai. 2012. "Road traffic injury in China: A review of national data sources". *Traffic Injury Prevention*. 13(sup1): 57–63. DOI: 10.1080/15389588.2011.633945.

Mattingly, D. C. 2019. *The Art of Political Control in China*. *Cambridge Studies in Comparative Politics*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108662536.

McCubbins, M. D. and T. Schwartz. 1984. "Congressional oversight overlooked: Police patrols versus fire alarms". *American Journal of Political Science*. 28(1): 165–179. DOI: 10.2307/2110792.

Moe, T. M. 2012. "Delegation, control, and the study of public bureaucracy". In: *The Forum*. Vol. 10. No. 2. De Gruyter.

Nakamura, E., J. Steinsson, and M. Liu. 2016. "Are Chinese growth and inflation too smooth? Evidence from Engel Curves". *American Economic Journal: Macroeconomics*. 8(3): 113–44.

Nie, H., M. Jiang, and X. Wang. 2013. "The impact of political cycle: Evidence from coalmine accidents in China". *Journal of Comparative Economics*. 41(4): 995–1011. DOI: 10.1016/j.jce.2013.04.002.

O'Meara, S. 2020. "The road to reducing traffic accidents in China". *Science*. 369(6501): 338–340.

Pan, J. and K. Chen. 2018. "Concealing corruption: How Chinese offcials distort upward reporting of online grievances". *American Political Science Review*. 112(3): 602–620. DOI: 10.1017/S0003055418000205.

People's Daily. 2006. "Background: The Formation of China's Emergency Response System". URL: http://politics.people.com.cn/GB/1026/4009970.html (accessed on 05/05/2020).

Pepinsky, T. B., J. H. Pierskalla, and A. Sacks. 2017. "Bureaucracy and service delivery". *Annual Review of Political Science*. 20(1): 249–268. DOI: 10.1146/annurev-polisci-051215-022705.

Potter, R. A. 2019. *Bending the Rules: Procedural Politicking in the Bureaucracy*. University of Chicago Press.

Rundlett, A. and M. W. Svolik. 2016. "Deliver the vote! Micromotives and macrobehavior in electoral fraud". *American Political Science Review*. 110(1): 180–197.

Sandefur, J. and A. L. Glassman. 2014. "The political economy of bad data: Evidence from African survey & administrative statistics". *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2466028.

Shi, X. and T. Xi. 2018. "Race to safety: Political competition, neighborhood effects, and coal mine deaths in China". *Journal of Development Economics*. 131: 79–95. DOI: 10.1016/j.jdeveco.2017.10.008.

Slack, D. 2016. "VA Bosses in 7 States Falsified Vets' Wait Times for Care". *USA TODAY*.

Stuart, E. A., H. A. Huskamp, K. Duckworth, J. Simmons, Z. Song, M. Chernew, and C. L. Barry. 2014. "Using propensity scores in difference-in-differences models to estimate the effects of a policy change". *Health Services and Outcomes Research Methodology*. 14(4): 166–182.

Sun, L. and S. Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". *Journal of Econometrics*. 225(2): 175–199.

Tang, Z. 2005. "Speech at the national work safety statistics working conference [Tigao renshi guifan guanli nuli zuohao anquan shengchan diaodu tongji gongzuo]".

The State Council. 2004a. "Decisions of the state council on further enhancing work safety [Guowuyuan guanyu jinyibu jiaqiang anquan shengchan gongzuo de jueding]".

The State Council. 2004b. "Opinions on establishing work safety governance performance indicators [Guanyu jianli anquan shengchan kongzhi zhibiao tixi de yijian]".

The State Council. 2006. "The national work place safety accidents emergency response plan [Guojia anquan shengchan shigu zainan yingji yuan]". URL: http://pkulaw.cn/fulltext_form.aspx?Db=chl&Gid=e245c9d7de2ef607bdfb&keyword=&EncodingName=&Search_Mode=&Search_IsTitle=0. (Accessed on 05/05/2020).

The State Council. 2007. "Regulations on the Reporting, Investigation and Disposition of Work Safety Accidents [Shengchan anquan shigu baogao he diaocha chuli tiaoli]". URL: http://www.gov.cn/zwgk/2007-04/19/content_588577.htm.

Ting, M. M. 2003. "A strategic theory of bureaucratic redundancy". *American Journal of Political Science*. 47(2): 274–292.

Truex, R. 2016. "The temporal logic of repression in China: A political calendar approach". *Working Paper*.

Wallace, J. 2016. "Juking the stats? Authoritarian information problems in China". *British Journal of Political Science*. 46(1): 11–29.

Wallace, J. L. 2022. *Seeking Truth and Hiding Facts*. Oxford University Press.

Wang, X. 2004a. "Notice on the work safety performance indicators from Xianzheng Wang [Wang Xianzheng dui 'guanyu jianli anquan shengchan kongzhi zhibiao tixi de yijian' de shuoming]".

Wang, X. 2006. "Notice on the 2006 national work safety performance indicators [Guanyu 2006 nian quanguo anquan shengchan kongzhi kaohe zhibiao de shuoming]". URL: http://www.gov.cn/gzdt/2006-01/25/content_171305.htm.

Wang, Z. 2004b. "Speed up the establishment of national work safety governance institutions [Quanguo anjian jigou jianshe bufa jiakuai]". China Work Safety News.

Wang, Z. 2004c. "To accelerate the pace of work safety institutions [Guojia anjian jigou jianshe bufa jiakuai]". *China Work Safety News*.

Whitford, A. B. 2003. "Adapting agencies: Competition, imitation, and punishment in the design of bureaucratic performance". *Politics, Policy, and Organizations: Frontiers in the Scientific Study of Bureaucracy*: 160–186.

Wilson, J. Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. URL: http://hdl.handle.net/2027/mdp.39015015321543.

Wuhua District Government. 2005. "Rewarding outstanding individuals in work safety performance". *Work Safety Bureau in Wuhua District, Yunan*. URL: http://wh.km.gov.cn/c/2005-03-28/726025.shtml.

Xu, Y. 2017. "Generalized synthetic control method: Causal inference with interactive fixed effects models". *Political Analysis*. 25(1): 57–76.

Zhou, X. 2010. "The institutional logic of collusion among local governments in China". *Modern China*. 36(1): 47–78.

Zuo, C. 2015. "Promoting city leaders: The structure of political incentives in China". *China Quarterly*. 2015(224): 955–984.