Filing Not Found:
Which Injuries Go Unreported to Worker Protection Agencies, and Why?

Alison Morantz
Stanford Law School

Abstract:

The underreporting of occupational injuries and illnesses to worker protection agencies has become a topic of great concern to researchers and policymakers. Although numerous studies have quantified the prevalence of the phenomenon, which specific *types* of injuries and establishments are most susceptible to underreporting is poorly understood. As a consequence, regulators have very little capacity to "red flag" employers that are likely to underreport the most injuries. This study begins to fill this gap in existing literature in four interrelated ways. First, I develop a simple theoretical model of the relationship between regulatory intensity, injury type, and underreporting. The model yields a number of concrete predictions about how the frequency of injuries, and mix of injury types, will respond to changes in the frequency and/or stringency of audits. Secondly, I propose a scheme for classifying different types of injuries by their relative "detectability." Third, using a dataset comprised of granular audit data obtained from the Occupational Safety and Health Administration (OSHA) and the Mine Safety and Health Administration (MSHA), I test the model's predictions regarding which types of injuries will be underreported the most across regimes and over time. Finally, I explore whether any observable, establishment-level covariates – such as the percentage of injuries contained in regulatory filings that are highly detectable – could be used, in a manner akin to the IRS, to identify likely violators. Overall, the results provide considerable grounds for optimism that mining injury data in this fashion could provide useful insights. Not only do my findings bear out most predictions of the model, but they also suggest that empirical algorithms could be devised, based exclusively on observable firm- and injury-level characteristics, to help labor regulators identify employers that hide workplace injuries.

This paper was prepared with funding from the U.S. Department of Labor. The views expressed are those of the authors and should not be attributed to the Federal Government or the Department of Labor.

I. Introduction[1]

In recent years, the underreporting of occupational injuries and illnesses to federal agencies has become an issue of great concern to Department of Labor (DOL) officials, Congress, and labor scholars alike. A growing body of scholarship suggests that injury and illness statistics recorded by employers at the behest of Department of Labor agencies, such as the Bureau of Labor Statistics (BLS), the Occupational Safety and Health Administration (OSHA), and the Mine Safety and Health Administration (MSHA), significantly undercount the true rates of occupational injury and illness.[2] Some researchers have estimated that as many as 70 percent of all injuries go unreported.[3] The underreporting phenomenon is not confined to minor injuries: one recent study suggests that at least one quarter of all amputations are not reported to OSHA.[4]

The deficiencies in the current injury and illness surveillance system compromise regulators' capacity to keep workers safe in several critical ways. First, customized injury prevention programs

[1] This project was funded by a grant from the Department of Labor's DOL Scholars Program. I could not have completed the project without the skilled research assistance of Austin Alleman, Rajlakshmi De, Kathleen Choi, and Alex Weiss. I am also deeply indebted to my colleague, Dr. Mark Cullen, Chief of Stanford University's Division of General Internal Medicine, for his extraordinary generosity in helping me devise a medically sound typology for identifying "hard-to-attribute" injuries.

[2] See, e.g., William J. Wiatrowski, "Examining the Completeness of Occupational Injury and Illness Data: An Update on Current Research," *Monthly Labor Review* (June 2014); Xiuwen S. Dong, et al., "Injury Underreporting Among Small Establishments in the Construction Industry," *American Journal of Industrial Medicine* 54 (2011): 339-349; Tahira M. Probst, Ty L. Brubaker, and Anthony Barsotti, "Organizational Injury Rate Underreporting: The Moderating Effect of Organizational Safety Climate," *Journal of Applied Psychology* 93 (2008): 1147-54; Monica Galizzi, et al., "Injured Workers' Underreporting in the Health Care Industry: An Analysis Using Quantitative, Qualitative, and Observational Data," *Industrial Relations* 49 (2010): 22-43; S. A. McCurdy, et al., "Reporting of Occupational Injury and Illness in the Semiconductor Manufacturing Industry," *American Journal of Public Health* 81 (1991): 85-89; John W. Ruser, "Examining Evidence on Whether BLS Undercounts Workplace Injuries and Illnesses," *Monthly Labor Review* (August 2008): 20-32; Alison Morantz, "Coal Mine Safety: Do Unions Make a Difference?" *Industrial and Labor Relations Review* 66 (2013): 88-116; Leslie Boden, and Alexander Ozonoff, "Capture-Recapture Estimates of Nonfatal Workplace Injuries and Illnesses," *Annals of Epidemiology* 18 (2008): 500-506; Leigh J. Paul, James P. Marcin, and Ted R. Miller, "An Estimate of the US Government's Undercount of Nonfatal Occupational Injuries," *Journal of Occupational and Environmental Medicine* 46 (2004): 10-18; K. D. Rosenman, et al., "How Much Work-Related Injury and Illness Is Missed By the Current National Surveillance System?" *Journal of Environmental Medicine* 48 (2006): 357-365.

[3] See Ruser (2008).

[4] See Wiatrowski (2014). As part of new severe injury and illness reporting requirements that went into effect in January of 2015, OSHA began requiring all employers under its jurisdiction to notify the agency within 8 hours of any work-related fatality, and to notify it within 24 hours of any hospitalization, amputation or enucleation. For a summary, see http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=NEWS_RELEASES&p_id=26673.

designed to assist workers in hazardous occupations will be of limited value if the data paint a distorted portrait of the distribution of injuries across industries and tasks.[5] Secondly, underreporting undermines the experience rating system upon which workers' compensation is based, whereby the most dangerous employers pay the highest insurance premiums. If firms can effectively shield workplace injuries from scrutiny – so that many injuries are neither reported to DOL nor trigger the filing of workers' compensation claims – the system will fail to induce the riskiest employers to invest more in accident prevention.[6] Finally, if firms that underreport injuries disproportionately violate other labor laws, the incapacity to identify the worst violators will hamper any coordinated regulatory strategy to protect workers' rights.[7]

Awareness of the underreporting problem has grown considerably since the turn of the millennium. In the mid-2000s, Congress held hearings on the issue and allocated funds to OSHA, BLS, and the National Institute of Occupational Safety and Health (NIOSH) for the express purpose of "follow[ing] up and expand[ing] on the previous research so as to understand the nature and magnitude of any undercount [of occupational injuries and illnesses] and attempt to identify solutions."[8] Yet to date, scholars still know relatively little about the prevalence and distribution of underreporting behavior across the U.S. economy. Most prior literature has compared employer-reported injury data to state workers' compensation filings to determine how many injuries are omitted from each data source. Much prior research also confines the scope of inquiry to a single

---

[5] William J. Wiatrowski, "Using Workplace Safety and Health Data for Injury Prevention," *Monthly Labor Review* (October 2013).

[6] See Boris Kralj, "Employer Responses to Workers' Compensation Insurance Experience Rating," *Industrial Relations* 49 (Winter 1994): 41-61; Sidney A. Shapiro, "Occupational Safety and Health Regulation," *Encyclopedia of Law and Economics, Cheltenham, Edward Elgar* 5540 (2000): 596-625.

[7] "[O]rganizations that have a lower commitment to safety are more likely to inadvertently or otherwise skew their injury data such that they appear to have similar safety outcomes as organizations with a positive safety climate." See Probst, Brubaker, and Barsotti (2008).

[8] See Wiatrowski (2014): 1-3.

industry or injury. Although often yielding valuable insights, these approaches sometimes yield biased estimates of the prevalence of underreporting to federal agencies, while failing to illuminate which *types* of injuries and workplaces are most vulnerable to underreporting.[9]

A few scholars have generated, and occasionally tried to test, hypotheses about which types of injuries are most likely to be underreported. As early as 1982, a study by the National Research Council conjectured that a subset of mining injuries that it labeled "intermediate" was less prone to underreporting than total injuries, although this assumption has never been empirically verified.[10] Several studies of particular industries[11] or discrete regulatory contexts[12] seem to bear out the notion that certain injuries – such as musculoskeletal injuries[13] and injuries occurring in smaller firms[14] and certain industrial sectors[15] – are highly vulnerable to underreporting. These findings are reminiscent of a parallel body of literature in the workers' compensation arena suggesting that some injuries,

---

[9] For a general, detailed discussion of these biases, see Wiatrowski (2014): 1-5. A recent (unpublished) report prepared by Eastern Research Group, and presented to the U.S. Department of Labor, also illustrates the pitfalls of relying exclusively on comparisons of workers' compensation claims to injuries reported to regulatory agencies. The report tabulated differences in "match rates" across workers' compensation records and MSHA Part 50 data by type of injury across two jurisdictions. See Eastern Research Group, *Final Report (Revised): Evaluation of the Accuracy and Completeness of Nonfatal Injury and Illness Reporting in the Mining Industry* (Lexington, MA: Eastern Research Group, June 11, 2013): 24-25, 34-36, http://www.dol.gov/asp/evaluation/reports/MSHA-Part50-Underreporting.pdf. Since the workers' compensation records used to conduct the matches were highly incomplete and themselves prone to underreporting, and there was no one-to-one correspondence between the categories used across the two datasets, no credible inferences could be drawn regarding whether certain types of injuries were more likely to be underreported than others. See also the report released by the U.S. Department of Labor, Office of the Inspector General (OIG), *MSHA Has Taken Steps to Detect and Deter Underreporting of Accidents and Occupational Injuries and Illnesses, But More Action Is Still Needed* (March 31, 2014), http://www.oig.dol.gov/public/reports/oa/2014/05-14-001-06-001.pdf.

[10] National Research Council, *Toward Safer Underground Coal Mines* (Washington, D.C.: National Academy Press, 1982).

[11] See Alison Morantz and Alexandre Mas, "Does Post-Accident Drug Testing Reduce Injuries? Evidence from a Large Retail Chain," *American Law and Economics Review* 10 (2008): 246-302, and Morantz (2013).

[12] See, e.g., Rosenman et al. (2006), which generates capture-recapture estimates of injuries reported in Michigan to multiple surveillance system, including the Bureau of Labor Statistics.

[13] See, e.g., Christine Daniels, and Peter Marlow, "Literature Review on the Reporting of Workplace Injury Trends," *Health and Safety Laboratory* 36 (2005); Wiatrowski (2014).

[14] See, e.g., Katherine L. Hunting, and James L. Weeks, "Transport Injuries in Small Coal Mines: An Exploratory Analysis," *American Journal of Industrial Medicine* 23 (1993): 391-406; Dong et al. (2011); Daniels and Marlow (2005); A. Oleinick, J. V. Gluck, and K. E. Guire, "Establishment Size and Risk of Occupational Injury," *American Journal of Industrial Medicine* 28 (1995): 1-21.

[15] See, e.g., Daniels and Marlow (2005).

especially hard-to-diagnose ones, are unusually prone to moral hazard effects.[16] Since approximately 2009, the Bureau of Labor Statistics has funded several independent, state-level studies that compare BLS non-fatal injury records with state workers' compensation data, match multiple-source data for cases involving work-related amputations and carpal tunnel syndrome, and interview employers about recordkeeping practices. Some of these studies have begun bearing fruit.[17] Finally, a few scholars in the field of occupational medicine have explored which injuries are most likely to go untreated because they are never reported to health care providers.[18] Yet no comprehensive mapping of the relationship between injury type and underreporting – or the practical implications of such relationships for the enforcement of occupational safety and health laws – has been attempted in prior work.

The paucity of knowledge regarding which types of injuries are most likely to evade detection limits the capacity of worker protection agencies to combat underreporting. The injury logs submitted to DOL contain granular detail on the frequency, type, and cause of occupational injuries and illnesses that occur each calendar year (or quarter). Yet like individuals that file income tax returns, employers have strong incentives to hide information they are statutorily obliged to report. In theory, DOL's worker protection agencies could emulate the Internal Revenue Service's fraud detection division and use statistical algorithms to target the worst violators. Just as an IRS official

[16] See, e.g., Georges Dionne and Pierre St-Michel, "Workers' Compensation and Moral Hazard," *The Review of Economics and Statistics* 73 (May 1991): 236-244.

[17] See, e.g., Wiatrowski (2014).

[18] For example, see K. D. Rosenman, et al., "Why Most Workers with Occupational Repetitive Trauma Do Not File for Workers' Compensation," *Journal of Occupational and Environmental Medicine* 42 (2000): 25–34; Laura Welch, and Katherine Hunting, "Injury Surveillance in Construction: What is an 'Injury', Anyway?" *American Journal of Industrial Medicine* 44 (2003): 191-196; David L. Parker, et al., "Characteristics of Adolescent Work Injuries Reported to the Minnesota Department of Labor and Industry," *American Journal of Public Health* 84 (1994): 606-611; Kris Siddharthan, et al., "Under-Reporting of Work-Related Musculoskeletal Disorders in the Veterans Administration," *International Journal of Health Care Quality Assurance* 19 (2006): 463-476; Oleinick, Gluck, and Guire (1995); Lenore S. Azaroff, Charles Levenstein, and David H. Wegman, "Occupational Injury and Illness Surveillance: Conceptual Filters Explain Underreporting," *American Journal of Public Health* 92 (2002): 1421-1429; Z. Joyce Fan, et al., "Underreporting of Work-Related Injury or Illness to Workers' Compensation: Individual and Industry Factors," *Journal of Occupational and Environmental Medicine* 48 (2006): 914-922.

might compare self-employment income or restaurant tips to state medians to help ferret out tax evasion, a DOL official could scrutinize the composition of injury logs for telltale signs of underreporting. Yet to date, no concerted attempt has been made to devise statistical algorithms that could help OSHA or MSHA channel regulatory resources toward the most likely offenders.

The present study opens up this line of inquiry in several interrelated ways. First, I develop a simple theoretical model of the relationship between regulatory intensity, injury type, and underreporting behavior. Starting with the simple observation that injuries differ in their relative chances of detection, the model presumes that the likelihood of underreporting will decline monotonically as detectability increases. It also yields a number of concrete predictions about how the frequency of injuries and mix of injury types will respond to changes in the frequency and/or stringency of audits. For example, the more intense the regulatory regime, the less detectable the type of injury that will best predict underreporting. In contrast to much existing literature, the model also implies that *ceteris paribus*, an increase in the frequency of audits and in their relative stringency can have very different effects. Whereas an increase in inspection frequency can have a similar deterrent effect on all types of injuries, a change in the thoroughness of inspections will have a disproportionate effect on the reporting of less-detectable injuries.

The paper's second important contribution is to propose a scheme, grounded in occupational medicine, for categorizing non-fatal injuries by their relative "detectability." I classify all occupational injuries into four types using three criteria: whether they are severe, whether they are traumatic, and whether they are easily attributed to activities performed on the job.

Third, using a dataset comprised of granular audit data from OSHA and MSHA, I test the model's predictions regarding which types of injuries will be underreported the most across regimes and over time. This task is not straightforward for two reasons. First, since the number of reporting violations detected is largely a function of the stringency of audits performed, "audit stringency bias" can confound my empirical estimates of the magnitude of underreporting. Secondly, the prevalence of behavioral incentive programs that penalize workers for reporting injuries (or reward them for not doing so) can cause "stickiness" in firms' responses to fluctuations in audit stringency. Both factors complicate my identification strategy. Nevertheless, my empirical findings broadly bear out the predictions of the model.

Finally, I explore whether any establishment-level covariates – such as the percentage of reported injuries that are highly detectable – could be used as red flags, in a manner akin to the IRS, to target the most likely violators. Several of the red flags explored do turn out to have significant predictive value. For example, a reported percentage of *severe* injuries that places an employer in the top quartile of the sample is a significant predictor of the frequency of underreported injuries in the OSHA environment, and a reported percentage of *easy-to-attribute* injuries that places the employer in the top quartile of the sample is a significant predictor of the frequency *and* percentage of underreported injuries in the MSHA environment. On the basis of these findings, I suggest that DOL's worker protection agencies would do well to emulate their peers at the IRS by developing new data-mining techniques to combat underreporting. Although the likelihood that firms will behave strategically in a more dynamic, iterative enforcement environment complicates the task of translating theory into practice, I suggest that this challenge is probably surmountable.

The next section, Section Two, describes the origin and construction of the datasets analyzed. Section Three presents a simple, non-mathematical model of the relationship between injury detectability, inspection intensity, and underreporting. Section Four describes my empirical methodology. Section Five presents the study's key findings. Section Six considers several practical challenges involved in policy implementation, and Section Seven concludes.

II. Description of Data

Each year, MSHA and OSHA collect self-reported injury and illness data from tens of thousands of private establishments. MSHA receives Part 50 data on a quarterly basis from every mine nationwide, and OSHA receives similar information from a sample of about 80,000 establishments in high-risk industries selected for inclusion in its annual OSHA Data Initiative (ODI) survey.

The OSHA data used throughout the analysis were obtained from two auditing programs that jointly encompass the years 1997 through 2012. From October 1997 through February 2009, OSHA audited injury and illness records from high-hazard industries through its Audit and Verification Program - Occupational Injury and Illness Records ("Recordkeeping" or "RK") program. Since each audit typically examined injury logs from two years prior, the RK audits collectively examined injury logs from 1996-2006. In 2009, OSHA suspended the RK program and instead began conducting audits under the auspices of its short-lived Injury and Illness Recordkeeping National Emphasis Program ("NEP"). The NEP program, which ran from September 2009 through February 2012, audited injury records from 2007 through 2009. (Whereas each RK audit examined just one year of an establishment's injury and illness history, NEP audits spanned two calendar years.) The average annual frequency of audits declined slightly under the NEP program.[19] By combining information from these two OSHA programs, I created a dataset encompassing injury logs from the years 1996-2009.

The RK and NEP programs were substantively similar in many regards. Both required auditors to ascertain whether the employer's log of total injury and illness counts, employment numbers, and hours worked for the audited year matched the summary data submitted to OSHA; to

[19] Whereas the number of audits conducted from 1996-2006, under the auspices of the RK auditing program, exceeded 225 in every year except 1996, the average number of audits conducted annually under the NEP program (2007-2009) was 200. (Calculations are based on an analysis of data received from Dave Schmidt, OSHA.)

check the accuracy and completeness of records for a random sample of employees drawn from the pertinent employee roster; and to interview the establishment's designated record-keeper to "assess [the] recordkeeper's knowledge of the OSHA injury/illness recordkeeping requirements and to determine whether recordkeeping problems exist."[20]

Yet the programs also differed in critical respects. First, the selection criteria used to determine which establishments were audited varied somewhat across the two programs. While both programs generally focused on firms in high-hazard industries, the NEP program added an additional selection criterion that compared each firm's "DART" rate (the rate of injuries and illnesses resulting in days away from work, restricted work, or transfer to another position) to the mean DART rate among high-hazard industries.[21] Secondly, unlike their RK counterparts, NEP auditors were obliged to interview managers to determine the existence of "any incentive or disciplinary programs that may influence recordkeeping."[22] Third, although employee interviews were optional under the RK program, they were mandatory under the NEP program, and the NEP auditor was further instructed to "focus interviews on employees likely to be injured or ill."[23] Finally, if an NEP auditor found any evidence of underreporting, (s)he had the right to "expand the records inspection beyond the

[20] See, e.g., U.S. Department of Labor, Occupational Safety and Health Administration, *Audit and Verification Program of Occupational Injury and Illness Records,* Directive Number: CPL 02-00-138, January 12, 2006, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=DIRECTIVES&p_id=3329; U.S. Department of Labor, Occupational Safety and Health Administration, *Injury and Illness Recordkeeping National Emphasis Program (RK NEP)*, Directive Number: 09-08 (CPL 02), September 30, 2009, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_09-08.pdf; and U.S. Department of Labor, Occupational Health and Safety Administration, *Injury and Illness Recordkeeping National Emphasis Program (RK NEP)*, Directive Number: 10-07 (CPL 02), September 28, 2010, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_10-07.pdf.

[21] Ibid. For example, while the RK's sample frame included *all* establishments above a given size cutoff that were part of the OSHA Data Initiative, the NEP program used a firm's reported Days Away and Restriction or Transfer (DART) rate as an additional selection criterion. The minimum size threshold in effect under the RK program was also lowered twice: once in late 1998 (from 60 to 50 employees) and again in December of 1999 (from 50 to 40 employees).

[22] Ibid.

[23] Ibid.

sampled employees," whereas RK auditors had no such authority.[24] In short, the NEP audits inspections were more stringent and comprehensive than RK audits. The data confirm, as one would expect, that reliance on employee interviews increased dramatically under the NEP program.[25]

Changes undertaken partway through the implementation of the NEP program, however, created sharp discontinuities in the mix of firms audited. During the first year of the program (Sept. 2009-Sept. 2010), which audited injury logs from 2007,[26] the agency targeted firms whose reported injury rates fell *below* the mean for high-hazard industries, 4.2 per 100 full-time employees. In the final 17 months of the program (Sept. 2010 – Feb. 2012), however, which audited injury logs from 2008 and 2009, the agency took the opposite approach of targeting firms whose reported DART rates fell *above* 4.2. As a consequence, injury rates from before and after January 2008 cannot be meaningfully compared. [27]

Information gleaned from OSHA audits is collected and stored at the audit level, making it very well suited to a study of this type. For each audit conducted, the dataset records information on both reported and unreported injuries. By examining the complete "injury set" for a given year, one

[24] Ibid.

[25] An average of 0.24 employee interviews per audit were conducted under the RK program, while the comparable figure for NEP audits was 3.19.

[26] In addition to the 2007 logs, injury logs from 2008 were also audited from February through September of 2010. See US DOL, OSHA, Directive Number: 10-07 (CPL 02) (2010); U.S. Department of Labor, Occupational Safety and Health Administration, *Injury and Illness Recordkeeping National Emphasis Program (RK NEP)*, Directive Number: 10-02 (CPL 02), February 19, 2010, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_10-02.pdf.

[27] In addition to the dramatic change in selection criteria after the 2007 injury logs were audited, firms evidently were given considerable *de facto* leeway to "correct" injury logs from prior years. Upon arriving at a workplace, each OSHA inspector was instructed to re-calculate the DART rate using the OSHA 300 Log for 2007 and "[i]f the re-calculated 2007 DART rate [fell] above the cutoff rate of this NEP (i.e., >4.2) the records inspection [would] not be conducted." See US DOL, OSHA, Directive Number: 09-08 (CPL 02) (2009), and US DOL, OSHA, Directive Number: 10-02 (CPL 02) (2010). In other words, if a firm responded to the announcement of the NEP program by retroactively recording more injuries on its 2007 logs, the audit would not be conducted and the firm would have been dropped from the dataset. A phone conversation with an employee of OSHA's Directorate of Enforcement Programs, September 4, 2014, also confirmed that this sometimes might have taken place. Unfortunately, I cannot discern from the data how frequently, if ever, such injury "backfilling" occurred. If it did occur, and firms with the most unreported injuries were especially likely to "backfill" their injury logs during the NEP program, then the figures presented here would probably *underestimate* the total number of injuries from 2007 that went unreported.

can ascertain how many (and what proportion of) injuries the audited firm failed to report. Moreover, for each audited establishment, the dataset records state, union status, total workers employed, and four-digit SIC code.[28] Importantly, however, the injury sets do *not* contain unique firm-level identifiers, nor do they contain the name or address of audited establishments. Therefore, audit results cannot be linked to other historical data on the same establishment, such as violations cited during prior OSHA inspections. Nor is it possible to determine if any firms were audited more than once.[29]

The MSHA dataset, encompassing the years 1992-2012, contains information gleaned from three different audit-like activities.[30] First, "compliance checks" performed during ordinary inspections include at least a cursory review of injury and illness records, which sometimes leads to the discovery of unreported injuries or illnesses. The second type, "Part 50" audits, have been undertaken since 1979 to assess mine operators' adherence with their statutory obligation to report all injuries and illnesses on a quarterly basis.[31] Thirdly, Potential Pattern of Violation ("PPOV") audits have been conducted since 2007 as part of a broader effort to single out the most dangerous mines for enhanced regulatory scrutiny. Mines that qualify for PPOV status on every criterion except their reportedly low rate of injury severity (total lost workdays X 200,000 / total hours worked) are targeted for PPOV audits on the theory that their numbers look "too good to be true" and thus could

[28] An SIC manual containing a list of these codes can be found on the OSHA website, https://www.osha.gov/pls/imis/sic_manual.html.

[29] One can obtain an upper bound for the prevalence of repeat audits by analyzing combinations of two establishment-specific fields: state and four-digit Standard Industrial Classification (SIC) code. If a given state and SIC code combination appears exclusively in a single audit-year, one can rest assured that the establishment in question was only audited once. Of the 2,019 unique combinations of state and four-digit SIC that appear in my data, 74.9% appear exclusively in a single audit-year. Therefore, no more than a quarter of the OSHA data utilized for the study could involve establishments that were audited repeatedly.

[30] Because of the small number of observations and somewhat ad hoc nature of its construction, the MSHA dataset does not include data from every year in this range. The final version includes injuries from the years 1992, 1994, 1996, 1997, and 2001-2012. Before 2001, however, the data only includes one mine for each year.

[31] According to Beth Nettles at MSHA, Part 50 audits dating back to this year appear in the MSHA database.

be inaccurate. When carrying out a PPOV audit, MSHA officials examine the most recent year (i.e. four quarters) of injury data.[32]

The scope and rigor of these three types of MSHA audits vary substantially. At one extreme are the PPOV audits, which are conducted by specially trained auditors and/or Education and Field Safety (EFS) staff, and scrutinize a broad array of data sources including medical records, medical claims forms, drug screening documents, employee interviews, and Part 50 information filed by mine operators. Since the program's inception, most (83%) of the mines subjected to PPOV audits have been underground mines.[33] At the opposite extreme are compliance checks conducted by ordinary MSHA inspectors, who simply compare data from injury logs (1000-1 and 7000-2 forms) and timesheets to the Part 50 information filed with MSHA's Denver office. Part 50 audits lie somewhere between these two extremes: although auditors are specifically tasked with finding unreported injuries, and scrutinize some of the same data sources examined in PPOV audits, they need not, and rarely do, examine them all.[34]

Unlike OSHA, MSHA does not store audit data at the injury level. For a given audited establishment, therefore, there is no straightforward way to compare injuries that were reported to those that were not. The only way to do so is to compare injuries that were *timely* reported in Part 50 filings with injuries that were *untimely* reported after the close of the fiscal (reporting) year. MSHA personnel indicated that injuries reported after the close of the fiscal year typically were reported at the behest of regulators following a compliance check or formal audit.[35] For the purposes of the

[32] Correspondence with Jay Mattos, MSHA, September 26, 2014.

[33] See Pattern of Violations Fact Sheet, http://www.msha.gov/POV/POVsinglesource.asp (last visited 9/26/2014) and "Mines Having Received POV or PPOV Notifications," available at http://www.msha.gov/POV/povmines.asp (last visited 9/26/2014).

[34] These institutional details are summarized in Eastern Research Group, *Final Report (Revised)* (2013).

[35] Telephone conference with Beth Nettles, MSHA, September 23, 2013.

empirical analysis, therefore, I compare two groups of injuries: those that were reported *after* the close of the fiscal year, and those that were reported by the same mine(s) in the same year(s) that the late-reported injuries occurred. For example, if a mine reported in 2013 an injury that took place in 2010, that injury would be compared to all injuries that the same mine timely reported in 2010. Because of these limitations, it is impossible to determine which respective proportions of late-reported injuries came to light as a result of PPOV audits, Part 50 audits, and compliance checks. Nor can I ascertain with certainty what number (or proportion) of injuries a given firm underreported in a given calendar year. Although mindful of these limitations, I create a "quasi injury set" for each mine that reported an injury after the close of the fiscal year. The quasi injury set contains both timely- and untimely-reported injuries from the year in question. Despite the likelihood of some measurement error, this methodology enables me to observe detailed injury information from all mines in the sample.

The substantive content of the OSHA and MSHA databases is largely similar. Both contain information on the source of injury, degree of injury, nature of injury, and affected body part. Both also enable me to observe the number of employees, state, year, and industry code. Yet there are a few salient differences. Although union status is available for all establishments audited by OSHA, in the mining sector union status is only available for coal mines.[36] Finally, the MSHA data contain information on "canvass code" (an industry coding scheme which the agency deems more useful than NAICS code), and on whether the mine operates at the surface or underground.

Both datasets are also susceptible to selection bias. The injury-reporting behavior of firms that are singled out for PPOV audits (in the case of MSHA) or NEP audits (in the case of OSHA) may differ from the behavior of firms that are not so targeted. Moreover, in the case of MSHA, firms that

[36] MSHA does not record data on union status, but the Department of Energy's Energy Information Administration (EIA) does record union status for coal mines on an annual basis. Using unique mine-level identifiers, I was able to append the union status field to the MSHA dataset.

underreported *no* injuries, or whose unreported injuries were never brought to light, are excluded entirely from the sample. Although the likelihood of selection bias (especially in the MSHA sample) may compromise the validity of out-of-sample predictions, the data still enable me to test several important predictions of the model, and may serve as useful bases for comparison in future studies that apply similar methodologies to a sample of randomly-selected establishments.

III. Injury Detectability, Regulatory Intensity, and Underreporting: A Simple Model

Although individuals may occasionally deviate from rational decision making due to cognitive biases or other behavioral limitations, most economic models presume that for-profit corporations operating in competitive markets engage in profit-maximizing behavior.[37] In the workplace safety context, profit maximization implies that a firm will compare the anticipated costs and benefits of abating a particular workplace hazard with the costs and benefits of allowing the hazard to persist. *Ceteris paribus*, an increase in the frequency or rigor of regulatory scrutiny raises the odds that a given hazard will be detected and penalized, thereby inducing rational firms to invest more in abatement. A sizable body of empirical literature has borne out these simple theoretical predictions.[38]

Parallel logic applies to the *reporting* of occupational injuries, which although mandated by statute (or regulation) is justifiably perceived as costly to employers. For example, reporting an

---

[37] See, e.g., Fritz Machlup, "Theories of the Firm: Marginalist, Behavioral, Managerial," *The American Economic Review* 57 (1967): 1-33; Jean-Jacques Laffont and David Martimort, *The Theory of Incentives: The Principal-Agent Model* (Princeton, NJ: Princeton University Press, 2002). Principal-agent theory – that is, the potential divergence between the incentives of a corporate entity *per se* and the incentives of the individual executives who make decisions on its behalf – is the leading paradigm used to explain instances in which corporations *do* exhibit apparent deviations from rationality. The pressure of market competition can play a significant role in mitigating such market imperfections.

[38] See, e.g., S. Lewis-Beck, and John R. Alford, "Can Government Regulate Safety? The Coal Mine Example," *The American Political Science Review* 74 (1980): 745-756; Wayne B. Gray, and John T. Scholz, "Does Regulatory Enforcement Work? A Panel Analysis of OSHA Enforcement," *Law and Society Review* 27 (1993): 177-214; David Weil, "If OSHA Is So Bad, Why Is Compliance So Good?" *The Rand Journal of Economics* 27 (1996): 618-640; Leon S. Robertson, and J. Philip Keeve, "Worker Injuries: The Effects of Workers' Compensation and OSHA Inspections," *Journal of Health Politics, Policy and Law* 8 (1983): 581-597.

injury makes it far more likely that the injury will be processed through the workers' compensation system, which (for all but the smallest employers) increases experience-rated premiums.[39] Full reporting of injuries may also harm the firms' eligibility for lucrative contracts[40], undermine workplace morale[41], place upward pressure on wages[42], and tarnish the firm's reputation.[43] Firms may also justifiably believe that reporting each and every injury will subject them to greater regulatory scrutiny, especially if underreporting is commonplace among their competitors.[44] For all these reasons, firms have strong incentives *not* to report all occupational injuries to worker protection agencies.

---

[39] Since it is well understood that experience rating itself encourages firms to underreport injuries, some researchers have even questioned whether the system has played an important role in lowering overall injury rates. See M. Harcourt, H. Lam, and S. Harcourt, "The Impact of Workers' Compensation Experience-Rating on Discriminatory Hiring Practices," *Journal of Economic Issues* 41 (2007): 681-99.

[40] A GAO report found, "Many employers did not report workplace injuries and illnesses for fear of increasing their workers' compensation costs or hurting their chances of winning contracts." Stephen Greenhouse, "Work-Related Injuries Underreported," *The New York Times,* November 16, 2009, http://www.nytimes.com/2009/11/17/us/17osha.html. Similarly, a Lockton Companies' report states that OSHA Incidence Rates are often used when deciding to hire a contractor. See Steven Polich, "Do Not Underestimate the Importance of OSHA Incidence Rates," *Lockton Companies,* December 2012, http://www.lockton.com/Resource_/PageResource/MKT/Polich_OSHA%20incidence%20rates_Dec%2012%20update.pdf. A Business Roundtable report likewise recommends that OSHA incidence rates be considered when selecting contractors for construction projects. See "Improving Construction Safety Performance," Report A-3, *The Business Roundtable* (January 1982).

[41] See Julian Barling, E. Kevin Kelloway, and Roderick D. Iverson, "Accidental Outcomes: Attitudinal Consequences of Workplace Injuries," *Journal of Occupational Health Psychology* 8 (2003): 74-85.

[42] See, e.g., Thomas J. Kniesner, and John D. Leeth, "Compensating Wage Differentials for Fatal Injury Risk in Australia, Japan, and the United States," *Journal of Risk and Uncertainty* 4 (1991): 75-90.

[43] For example, one popular business magazine advised its clientele, "Injuries tarnish a company's reputation and erase years of marketing gains." See Phil La Duke, "What Every Entrepreneur Should Know about Worker Safety," *Entrepreneur*, May 30, 2014, http://www.entrepreneur.com/article/234305. Also, Williams and Barrett find a negative relationship between a firm's reputation and the number of OSHA and Environmental Protection Agency (EPA) violations at the firm. See, e.g., Robert J. Williams, and J. Douglas Barrett, "Corporate Philanthropy, Criminal Activity, and Firm Reputation: Is There a Link?" *Journal of Business Ethics* 26 (2000): 341-350.

[44] As noted in a *New York Times* article summarizing the findings of a GAO report, "[m]any employers fear that reporting numerous injuries will prompt a full-scale OSHA inspection." See Greenhouse (2009). OSHA's site-specific targeting program, for example, targets firms for inspection based on their reported incidence rates. See U.S. Department of Labor, Occupational Safety and Health Administration, *Site-Specific Targeting 2014 (SST-14),* Directive Number: 14-01 (CPL 02), March 6, 2014, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02-14-01.pdf. High rates of injury or illness may also subject mines for increased enforcement by MSHA. See US DOL, OIG, *MSHA Has Taken Steps* (2014).

Yet failing to timely report an injury can expose a firm to financial penalties and reputational harms beyond those arising from the injury itself. Under OSHA's NEP program, for example, employers found to have violated their recordkeeping requirements were subjected to substantial monetary penalties, especially in cases where OSHA exercised its discretion to "stack" penalties for multiple violations.[45] In two well-publicized cases, for example, OSHA assessed firms $1,215,000[46] and $2,590,000[47] in penalties for their repeated failure to report occupational accidents.

If the frequency and stringency of regulatory inspections are largely exogenous and firms treat them as such, one would expect risk-neutral, profit-maximizing firms to consider both the costs of reporting and the costs of failing to report in choosing which, and how many, injuries to include in regulatory filings. Any regulatory change that lowers the risk that unreported injuries are detected and penalized, such as a cut in federal funding for recordkeeping audits, should increase the prevalence of underreporting. Conversely, any factor that *increases* the expected cost of underreporting – such as a rise in average penalties or an increase in the likelihood of detection – should lower the prevalence and magnitude of underreporting.

These simple predictions can be usefully extended, however, if one accounts for the fact that not all injuries are created equal. The likelihood that an injury can be successfully hidden from regulatory scrutiny varies by injury type. For example, occupational fatalities are very difficult for

[45] See Patrick Melfi, "New OSHA Initiative Targets Underreporting of Workplace Injuries," *New York Labor and Employment Law Report,* February 22, 2010, http://www.nylaborandemploymentlawreport.com/2010/02/articles/occupational-safety-and-health/new-osha-initiative-targets-underreporting-of-workplace-injuries.

[46] In 2010, OSHA "issued Goodman Manufacturing Co. LP 83 willful citations for failing to record and improperly recording work-related injuries and illnesses at the company's Houston air conditioning cooling facility. Proposed penalties total $1,215,000." See U.S. Department of Labor, Occupational Safety and Health Administration, *US Department of Labor's OSHA Cites Houston Manufacturing Company for Hiding Work-Related Injuries and Illnesses; Fines Exceed $1.2 Million,* Release Number: 10-1179-DAL, September 1, 2010, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=NEWS_RELEASES&p_id=18261.

[47] In 1987, OSHA assessed a $2.59 million fine against IBP Inc., a large meatpacker, for failing to report more than a thousand job-related injuries and illnesses over a two-year period. See Philip Shabecoff, "OSHA Seeks $2.59 Million Fine for Meatpacker's Injury Reports," *The New York Times,* July 22, 1987, http://www.nytimes.com/1987/07/22/us/osha-seeks-2.59-million-fine-for-meatpacker-s-injury-reports.html.

firms to shield from regulatory purview. This is so not only because of their extreme salience to coworkers and surviving family members, but also because of the unparalleled intensity of regulatory scrutiny.[48] Conversely, cumulative low-back injuries whose underlying cause and work-relatedness are often difficult to ascertain may be quite easy for firms to underreport. Not only may managers question whether the injury is truly work-related,[49] but they can easily direct the injured worker to providers outside the occupational safety and health system, such as group health plan clinicians.[50] In the latter scenarios, little if any paperwork is typically filed by the employer. The only way the injury may come to light is if the employee persists in filing a workers' compensation claim, or if an auditor learns of the injury during an interview with an affected worker.

To illustrate the choice confronting an employer, it is helpful to visualize the full range of injuries as falling along an "injury detectability continuum," represented by the horizontal (*x*) axis in Figure 1. The term "detectability," as used throughout this paper, encompasses not merely the extent to which an injury is visible to onlookers, but more broadly, the likelihood that an employer could realistically, and with impunity, avoid disclosing it to regulators. The vertical (*y*) axis represents the probability that the employer will decline to report the injury.

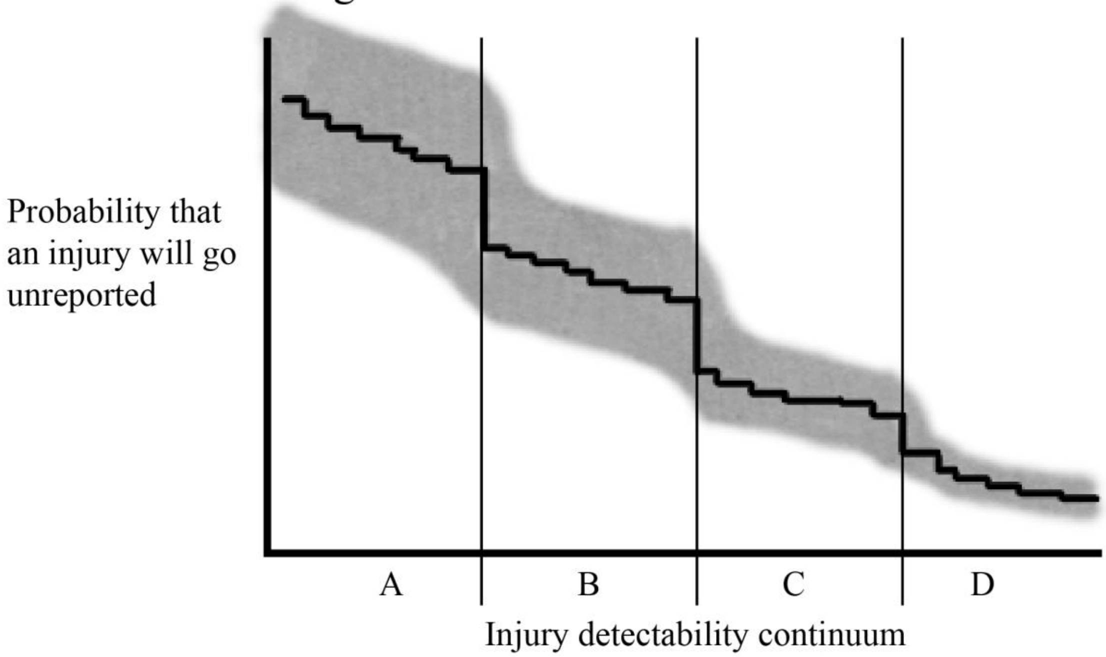
---

[48] The Census of Fatal Occupational Injuries (CFOI), conducted by the Bureau of Labor Statistics, uses extremely sophisticated surveillance techniques in an effort to obtain a comprehensive annual census of all workplace fatalities occurring nationwide. For a description of the program, see http://www.bls.gov/iif/oshfat1.html (last visited August, 13, 2014).

[49] States vary with regard to if injuries stemming from pre-existing conditions are compensable under workers' compensation. For example, in Oregon, a claim is only compensable if "the work injury/exposure was the major contributing (51% or more) cause of the disability or need for treatment,"(see Arthur W. Stevens, "New Comp Law Defines 'Preexisting Conditions'," *Black Chapman Webber & Stevens Attorneys* (blog), June 8, 2010, http://www.blackchapman.com/new-comp-law-defines-preexisting-conditions) while in California, any injury that aggravates a pre-existing condition is compensable (see California Department of Human Resources, *Workers' Compensation Preview* (January 2014), http://www.calhr.ca.gov/Documents/workers-compensation-preview.pdf).

[50] See *Increasing Access to Workers' Compensation Medical Benefits for Low-Income and Immigrant Workers*, NH Coalition for Occupational Safety and Health Research Report, March 31, 2008, http://www.nhcosh.org/pdfs/NHCOSH_WC_Report_3-31-08.pdf.

Figure 1: Basic Theoretical Framework



At the origin are less-detectable injuries that are relatively easy for an employer to underreport. At the far-right end of the *x*-axis are injuries that are very difficult for an employer to hide, such as fatalities. If all aspects of an injury were perfectly observable, the relationship between detectability and the likelihood of underreporting could appear linear, or at least smoothly continuous. In reality, however, many aspects of an injury that affect its detectability are unobservable, even to employers. Therefore, one would expect employers (and researchers) to group injuries into categories based on their observable characteristics. Figure 1 depicts a scenario in which injuries can be grouped into four discrete types, A, B, C and D, that vary in average detectability.

Importantly, within a given type, not all injuries are reported with equal probability. This is because employers have access to more granular information than researchers and are likely to use finer classification schemes. For example, even among nonspecific, cumulative low-back injuries, there will be variations in the likelihood that an employer who chooses not to report the injury will be caught and penalized. (A highly vocal, unionized employee working full-time in a warehouse loading boxes onto trucks, for example, will have a more detectable injury claim than a part-time, non-unionized receptionist who plays amateur rugby.) The "mini-steps" *within* each injury type illustrate this point. For purposes of empirical analysis, however, the key assumptions are simply that average detectability varies across injury types, and that the types observable to researchers are equally observable to employers. Since the costs of underreporting increase with an injury's detectability, the likelihood of underreporting should peak at the origin and then decline monotonically, in stair-step fashion, as one proceeds along the *x*-axis.[51]

Another important feature of Figure 1 is the shaded area surrounding the staircase. This area represents variation across firms in the likelihood that an injury will be reported at each level of detectability. In other words, it represents heterogeneity in how different firms "descend the staircase." As is shown in the figure, the elevation and steepness of the staircase, and the relative steepness of each individual step, can vary widely across employers. Yet the amount of dispersion is also likely to vary by injury type. For less-detectable injuries, routine underreporting may be the profit-maximizing choice. Consequently, many firms will choose to underreport. Yet at the same time, non-economic factors such as social norms, or risk aversion on the part of individual managers, may induce some firms to report even those injuries they could successfully shield from regulatory scrutiny.[52] In short, for less detectable injuries, one would expect to see considerable dispersion across firms in the prevalence of underreporting. For highly detectable injuries, however, the opposite logic applies. Simple cost-benefit analysis will dictate the wisdom of reporting, so one is

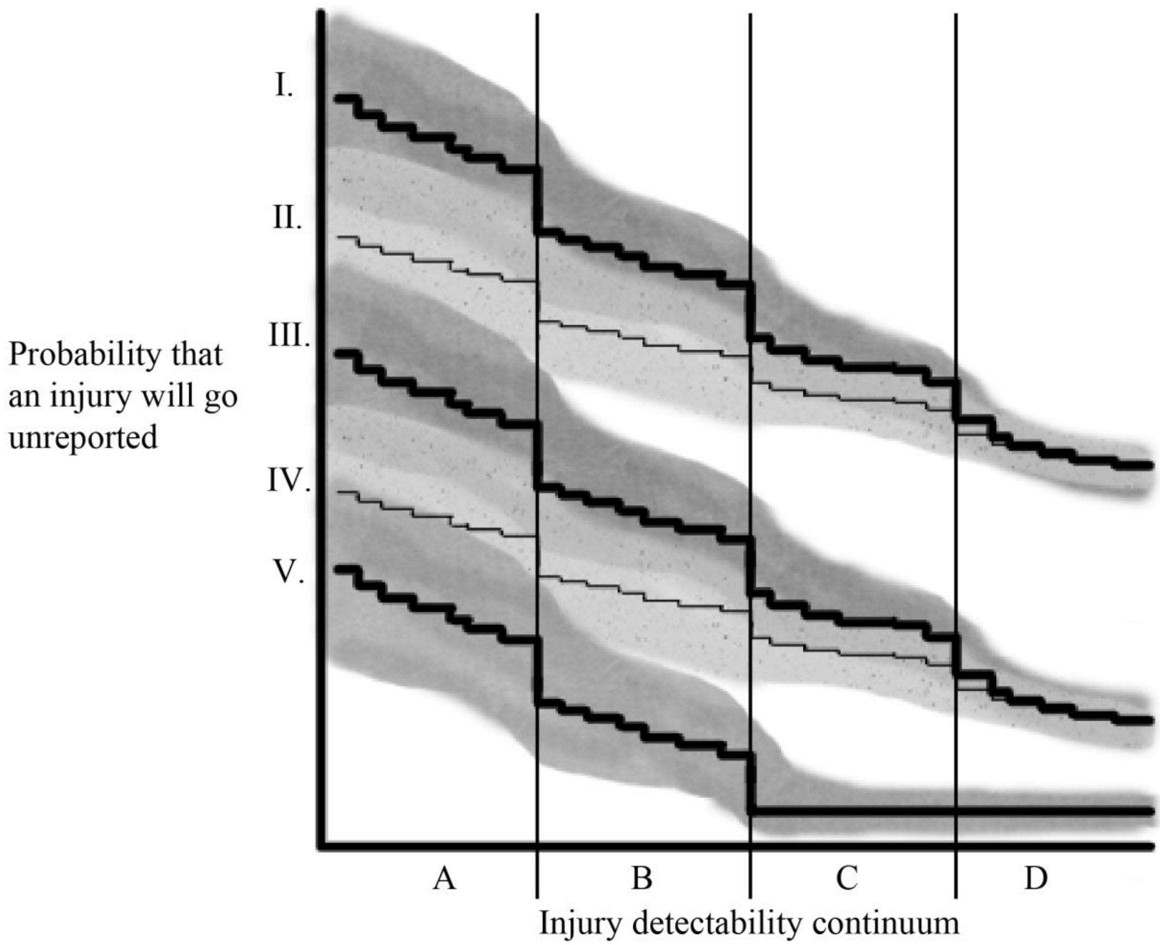
[51] This relationship need not, however, be linear, since the marginal difference in "detectability" may vary at different points along the continuum.

[52] For a discussion of this phenomenon in the OSHA context, see Weil (1996).

likely to see less variation across firms. Cross-firm dispersion in the likelihood of underreporting, then, should decline as detectability increases.

Figure 2 expands this simple framework by considering the effects of a change in regulatory intensity. An escalation in regulatory intensity can take two different forms: an increase in the *frequency* of audits, and/or an increase in the *stringency* (i.e., rigor and thoroughness) of each audit. Although in many contexts the two are largely interchangeable,[53] in this model they have different properties and different expected impacts on firm behavior.

Figure 2: Expanded Theoretical Framework



[53] If the probability of detecting unreported injuries is simply a function of the amount of time that an auditor spends onsite, then a 10% increase in total audit hours would have more or less the same effect, *ceteris paribus*, regardless of whether it results from an increase in the frequency of audits or in the number of hours devoted to each audit.

To see this, consider a regime that shifts from rudimentary audits (consisting exclusively of a comparison of injury filings with detailed injury logs, workers' compensation claims, and first-incident reports) to more thorough audits involving extensive employee interviews and scrutiny of hospital records. If rudimentary audits are reasonably successful at discovering more highly detectable injuries, such as amputations, but only rarely uncover less detectable injuries, like cumulative low-back injuries, the effects of the shift will vary widely by injury type. Whereas the likelihood of discovering highly detectable injuries may change very little, if at all, the odds of detecting less detectable injuries should increase substantially. This scenario is depicted by the shift from Staircase I to Staircase II in Figure 2.

Now consider the alternative scenario in which the stringency of each audit does not change, but audits occur with greater frequency. In this scenario, the staircase will shift in a more complex fashion that depends on the marginal increase in frequency and the (initial) height and position of each step in the staircase.[54] Figure 2 illustrates a scenario in which the probability of detection shifts by roughly equal amounts for all injury types, represented by the shift from Staircase I to Staircase III.[55] Moreover, if audit frequency or stringency increases to such a degree that the probability of underreporting certain types of injuries cannot fall any further, one will observe a "kink" in the

[54] Assuming that the likelihood of reporting a given injury is a linear function of the probability that the injury will be detected by auditors, the shift in the underreporting "staircase" will depend on the marginal increase in audit frequency; the probability of detecting the injury in each audit; and the dispersion across injury types ("steps") arrayed along the x-axis (injury detectability continuum). Intuitively, this is so because the probability of detection, like the probability of getting heads when flipping a coin, does not increase in a constant (linear) fashion as the number of inspections increases. Rather, the marginal increase in probability depends on the probability of detection during any given audit and on the number of audits already conducted. More precisely, the probability of an injury being detected (D) is described by the equation $D = 1-p^n$, where p is the probability that the injury is *not* detected in each audit, and n is the number of audits. Generally speaking, the smaller the increase in audit frequency, the closer p (and the probability of detection) are to 0.5, and the smaller the dispersion across injury types in the probability of detection, the greater the relative impact of increasing audit frequency, and the more closely the shift will resemble a constant (linear) shift like that depicted by the move from Staircase I to III in Figure 2.

[55] For example, a shift from 1 to 2 inspections, in which the initial detection probabilities were .45, .55, .65, and .75 (respectively) for injury types a, b, c and d, would resemble the shift from Staircase I to Staircase III in Figure 2.

staircase where the likelihood of underreporting reaches its lower bound.[56] This scenario is illustrated by the shift from I to V in Figure 2.[57]

In reality, of course, changes in audit frequency and in audit stringency are not mutually exclusive. A scenario in which they occur simultaneously is depicted by the shift from I to IV.

Within this simple theoretical framework, one can generate a number of empirical predictions about the relationship between injury type, regulatory intensity, and underreporting. Generally speaking, the laxer an inspection regime – in other words, the smaller the probability of detection times the size of the expected penalty – the higher the staircase. Moreover, the lower the expected likelihood that a given injury will be detected (and the lower the associated penalty), the higher the step associated with that injury.

Secondly, if employers underreport most (if not all) injuries below a given level of detectability, and report nearly all injuries above that threshold, then some steps in the staircase will be much steeper than others. Consider, for example, the simplest case in which an employer underreports *all* injuries of type A or B, but *no* injuries of type C or D. Under these conditions, the entire staircase will be descended in a single step at the juncture between B and C. This is the pattern one would expect to see if employers use rules of thumb to guide reporting decisions based on readily observable information that can be easily conveyed to mid-level managers.

[56] As long as a few employers – or the agents to whom they delegate decision-making authority – are heedless of the risk of detection, perhaps because of an idiosyncratically short time horizon or even a preference for risk, then the lower bound will be greater than zero.

[57] One aspect of reporting behavior that is not incorporated into the model is the possibility of "overreporting." Overreporting occurs when an employer reports an incident that does not meet the agency's recordability criteria – for example, reporting injuries that only require first aid or reporting injuries that do not result in medical treatment (although they may have received medical attention). Both RK and NEP OSHA audits make note of overreported cases as a part of the audit protocol, but these reporting inaccuracies do not result in citations. See U.S. Department of Labor, Occupational Safety and Health Administration, *1998 Audit and Verification Program of Occupational Injury and Illness Records,* Directive Number: 00-1 (CPL 2), December 2, 1999, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=DIRECTIVES&p_id=2003, and US DOL, OSHA, Directive Number: 09-08 (CPL 02) (2009) (cited earlier). Approximately 10% of all injuries reported do not meet the recordability criteria, and approximately 5% of firms audited engaged in a significant amount of overreporting. (Note: all overreported cases have been dropped from the dataset used for the empirical analysis.)

The location of the steepest step(s) will likewise depend on the intensity of the inspection regime. In an extremely lax regime – in which employers fail to report all but the most severe injuries – the steepest step(s) will occur toward the far-right end of the *x*-axis. Conversely, in a very strict regime, in which all but the least detectable injuries are uncovered by auditors, the steepest step(s) will occur very close to the origin.

Third, the best predictors of underreporting will tend to align with the steepest step(s) in the staircase descended by the typical firm. For example, consider the simple case described above, in which the typical employer underreports *all* injuries of type A or B, and *no* injuries of type C or D. In that case, the best predictor of whether an injury is underreported will be *whether the injury is of type A or B versus type C or D.* Since the location of the steepest step(s) varies by the intensity of the inspection regime, the best empirical predictor(s) of underreporting will also vary with regulatory intensity. In a strict regime, the best predictor(s) will occur relatively close to the origin; in a lax regime, the best predictor(s) will occur much farther from the origin.

Finally, a worker protection agency's capacity to mimic the IRS – that is, to identify which employers are underreporting, merely by inspecting the injuries listed in mandatory filings – depends not only on the amount of heterogeneity across firms, but also on the intensity of the regulatory regime. To see this, consider a case in which the regulatory regime is so lax that the *only* injuries reported are those at the highest level of detectability. Under such conditions, there could still be considerable dispersion across firms in the magnitude of underreporting. For example, relatively compliant firms might report *all* injuries that are of Type D, whereas less compliant ones might only report those that fall into the top decile of the detectability continuum. Yet the composition of injuries in the two firms' injury filings would be identical. (Both would consist exclusively of Type D injuries.) In short, extreme laxity of enforcement compromises an agency's capacity to distinguish typical firms from the worst violators. As long as a regulatory regime is sufficiently robust that many firms descend the staircase at the left and middle portions of the detectability continuum, this concern does not apply. Assuming a reasonable amount of heterogeneity across employers, a robust regulatory agency should be able to develop empirical red flags or algorithms to identify which firms underreport the most injuries merely by examining firm-level attributes and the composition of injuries in regulatory filings.

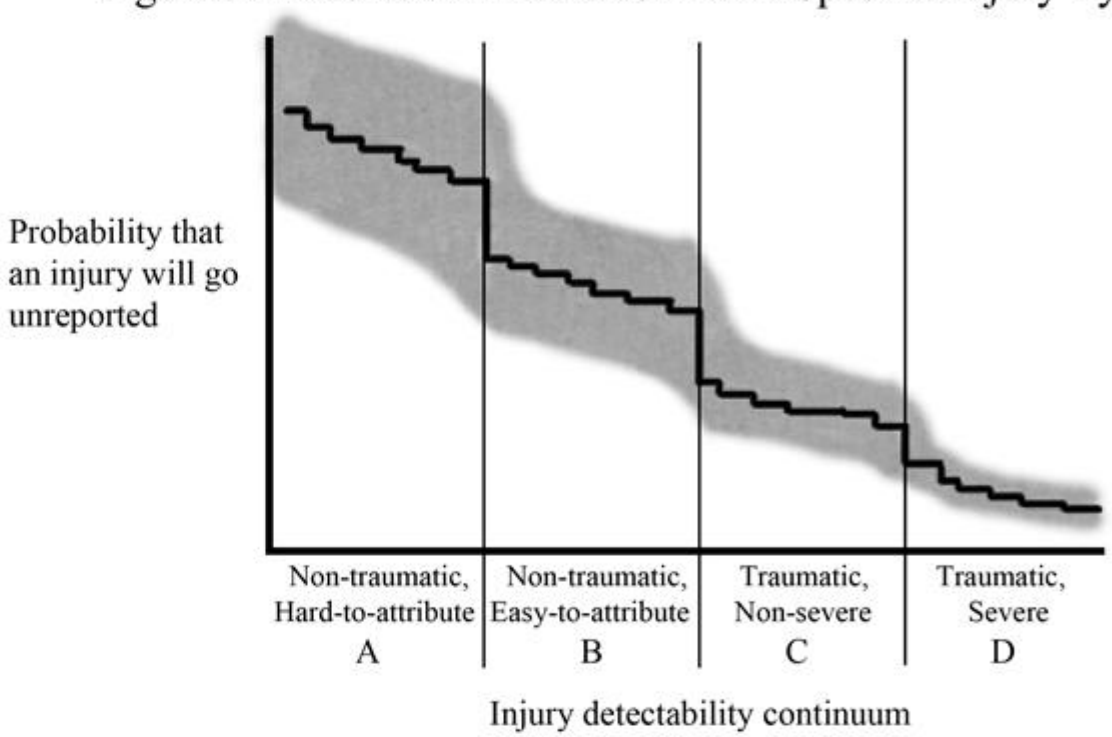
Although many of these predictions are amenable to empirical verification, further information is required to do so. Unless one can identify several discrete, observable types of injuries and place them correctly along the *x*-axis, one cannot test whether the hypotheses are consistent with available data. Drawing on expertise from the field of occupational medicine,[58] I introduce a four-part classification scheme for ordering injuries along the *x*-axis, depicted below in Figure 3. (Since the Census of Fatal Occupational Injuries and Illnesses detects virtually all occupational fatalities across all industries in the U.S., making detection a virtual certainty, I omit occupational fatalities from the empirical analysis and focus exclusively on non-fatal injuries.) There is nothing sacred about the scheme proposed. With perfect and costless information about every injury, one could calibrate the detectability of each individual injury with infinite precision, and line up all injuries along the detectability continuum in a smoothly continuous fashion. In reality, however, many aspects of an injury's detectability are not perfectly observable. Therefore, developing a categorical classification scheme reliant upon characteristics reported in injury filings is a critical simplification that helps me bridge the gap between theory and practice.

As depicted in Figure 3, I hypothesize that non-fatal injuries can be meaningfully grouped

[58] This classification scheme would not have been possible without the help of Dr. Mark Cullen. See notes 1 and 59.

into four categories of increasing detectability: nontraumatic and hard-to-attribute (A); nontraumatic and easy-to-attribute (B); traumatic but not severe (C); and traumatic and severe (D).[59] Type D (traumatic and severe) injuries are those that usually require urgent care, such as fractures, crushing, concussions and amputations. Type C (traumatic but non-severe) injuries, such as lacerations, contusions, burns, and non- eye-related abrasions, also occur instantaneously but are generally less serious and require less acute care. Type B (non-traumatic and easy-to-attribute) injuries include cumulative impairments whose underlying cause is relatively easy to detect, such as hearing loss, eye-related abrasions, hernias and heatstroke.

Figure 3: Theoretical Framework with Specific Injury Types



[59] Conversation with Mark Cullen, M.D., Professor of Medicine, Stanford University. Non-traumatic, hard-to-attribute (A) injuries include sprains and strains (mostly of the back or shoulders) as well as joint, tendon, and muscle inflammation such as tendonitis and carpel tunnel (mostly of the wrist or elbow). Non-traumatic, easy-to-attribute (B) injuries encompass sprains and strains (mostly of the knee or ankle), hearing loss, eye-related abrasions, skin-related injuries, hernias, heatstroke, and poisoning. Traumatic but not severe (C) injuries include lacerations, contusions, foreign objects in eye, burns, non-eye-related abrasions, and electric shocks. Traumatic and severe (D) injuries include fractures, crushing, dislocations, amputations, enucleation, and concussions. See companion website for more detailed description of categorization scheme, available at LINK TO WEBSITE TO BE INSERTED HERE ONCE IT IS ONLINE.

Type A, the final category, poses especially complex policy challenges. Non-traumatic and hard-to-attribute injuries encompass musculoskeletal disorders such as carpal tunnel syndrome and nonspecific strains or sprains of the back and shoulders. Although their root cause is often difficult to detect using existing medical technologies, such injuries can be severely disabling and impose enormous burdens on the labor market and health care system. One study estimated the annual treatment costs and lost wages associated with musculoskeletal diseases to be $849 billion, or 7.7% of Gross Domestic Product,[60] and another observed that "[m]ore U.S. health care dollars are spent treating back and neck pain than almost any other medical condition." [61]

Given their nontraumatic nature and the difficulty of pinning down whether they were caused by activities performed at work, such injuries are especially prone to underreporting. Not only may insured workers prefer to seek treatment from their primary care providers, but clinicians themselves have strong incentives *not* to probe the work-relatedness of the injury.[62] Indeed, the only way for regulators to uncover hard-to-attribute injuries during an audit may be to ask workers directly about their health status or pain they experience while performing job-related duties.

To test the model's predictions about how regulatory intensity affects reporting behavior, one must also discern which regulatory regimes are stricter than others. Because MSHA inspects regulated establishments much more frequently than OSHA and also collects injury data with more frequency and granularity, I take MSHA to be the stricter regime. I further subdivide my data in two ways. First, since MSHA is statutorily required to inspect underground mines twice as often as

---

[60] See "The Burden of Musculoskeletal Diseases in the United States," executive summary, *Bone and Joint Decade* (2011), http://www.boneandjointburden.org/pdfs/bmus_executive_summary_low.pdf. See also Deborah P. Lubeck, "The Costs of Musculoskeletal Disease: Health Needs Assessment and Health Economics," *Best Practice & Research Clinical Rheumatology* 17 (2003): 529–539.

[61] See Salynn Boyles, "$86 Billion Spent on Back, Neck Pain," *WebMD Health News,* February 12, 2008, http://www.webmd.com/back-pain/news/20080212/86-billion-spent-on-back-neck-pain.

[62] See, e.g., Azaroff, Levenstein, and Wegman (2002).

surface mines, I assume that underground mines are subject to greater regulatory scrutiny than surface mines.[63] Secondly, as explained above, the scope and stringency of OSHA's auditing protocols increased with the transition from the RK auditing program (1996-2006) to the short-lived NEP program (2007-09). Thus within the OSHA regime, I assume that *ceteris paribus*, the intensity of regulatory scrutiny increased with the advent of NEP audits.

IV. Methodology

The goal of the empirical analysis is to test whether the model's theoretical predictions are borne out by available data. Two real-world obstacles, however, complicate several portions of the identification strategy.

First, my estimates of the probability that an injury is not reported (or of the prevalence of underreporting) are susceptible to what I call "audit stringency bias." In addition to capturing the likelihood that an injury is actually underreported, the coefficients reflect the likelihood that an unreported injury is *brought to light during an audit, and thus included in the dataset*. Suppose, for example, that a new statute triples the amount of funding dedicated to audits. The law will have two countervailing impacts on the (observable) frequency of underreporting. If firms become aware of the change and believe it will raise the odds that unreported injuries are discovered, they will respond by underreporting fewer injuries. Yet if that the agency uses the extra resources to conduct more stringent audits, the number of unreported injuries discovered during each audit will also surely rise. The relative magnitudes of these two effects will determine the net effect of the law.

Importantly, the magnitude of audit stringency bias should generally decline as injury detectability increases. This is so because even a low-intensity audit is likely to catch highly

[63] The Mine Act requires MSHA to inspect underground mines four times per year, whereas surface mines need only be inspected twice per year.

detectable injuries like amputations, whereas only a rigorous one is likely to uncover an unreported cumulative back strain. In general, one might expect audit stringency bias to predominate in the short run, before firms can respond fully to a regulatory change. Over time, as they become more familiar with the altered incentive scheme, firms' behavioral responses to the change should mitigate audit stringency bias.

The second problem is the model's (implicit) simplifying assumption that employers are fully capable of adjusting their reporting behavior. In reality, the choice of whether to report an injury does not lie exclusively with employers; it also lies to some extent with injured workers themselves. If a worker decides that the cost of reporting an injury outweighs its benefits, she may decline to report it to her employer and simply seek treatment from her primary care physician. Instead of filing a workers' compensation claim, she may use up sick days (or vacation days) while she recuperates, or even take unpaid sick leave.[64] From an employer's perspective, encouraging employees not to report injuries in the first place may be far more appealing than omitting injuries from regulatory filings. After all, if an employee declines to report an injury, her employer can (credibly) maintain that it had no idea it even occurred. It also lessens the likelihood that the injury will ever come to light, since it is not recorded in any permanent record and may never be discovered unless the affected employee (or a coworker) brings it up during an interview.

The concern that some workers hide workplace injuries from their employers is more than theoretical. The rise of a popular risk-management philosophy known as "behavior-based safety" (BBS) has heightened concerns that such behavior is commonplace. Widely embraced by employers as a means to induce workers to play a more active role in creating a strong safety culture, BBS

[64] Under the Family and Medical Leave Act, "eligible employees of covered employers [are entitled] to take unpaid, job-protected leave for specified family and medical reasons with continuation of group health insurance coverage under the same terms and conditions as if the employee had not taken leave. " See http://www.dol.gov/whd/fmla (last visited August 1, 2014).

involves the identification of target behaviors that impact safety; the definition of such behaviors with sufficient precision to permit data collection; the tracking of targeted behaviors over time so as to chart progress toward clearly defined goals; the ongoing provision of feedback to workers; and the reinforcement of progress toward goals. Reinforcement systems commonly include the provision of tangible rewards (such as monetary bonuses or valuable prizes) to individuals or groups that achieve safety goals (such as an "injury free month"), and sometimes even include the application of penalties against individuals or groups that fail to meet specific targets. [65] Although such programs may succeed in encouraging workers to take greater care on the job, they may also discourage workers from reporting on-the-job injuries.

DOL officials have repeatedly voiced concerns about the deleterious effects of BBS-inspired incentive systems ("incentive programs") on injury reporting. On June 29, 2011, OSHA issued guidance materials for its Voluntary Protection Program discussing the danger of offering financial incentives for the non-reporting of injuries, and suggesting alternative ways to encourage safe practices without incentivizing workers to hide injuries.[66] The concern was reiterated in subsequent reports and memoranda.[67] Meanwhile, a report issued by the Department of Labor's Office of the Inspector General in early 2014 expressed nearly identical concerns about the prevalence of such programs in the mining industry.[68]

---

[65] Beth Sulzer-Azaroff and John Austin, "Does BBS Work? Behavior-Based Safety & Injury Reduction: A Survey of the Evidence," *Professional Safety* (July 2000): 19-24.

[66] U.S. Department of Labor, Occupational Safety and Health Administration, Memorandum from David Michaels to Regional Administrators, *Revised VPP Policy Memorandum #5: Further Improvements to the Voluntary Protection Programs (VPP)*, June 29, 2011, https://www.osha.gov/dcsp/vpp/policy_memo5.html.

[67] See Nancy Smith, et al., *OSHA's Voluntary Protection Programs (VPP) Review: Findings and Recommendations*, a Report Submitted to David Michaels, Assistant Secretary for OSHA, (November 2011): 21-22, https://www.osha.gov/dcsp/vpp/vpp_report_nov_2011_rev_7-11-12.pdf; and U.S. Department of Labor, Occupational Safety and Health Administration, Memorandum from Richard E. Fairfax, OSHA Deputy Assistant Secretary, to Regional Administrators, "Employer Safety Incentive and Disincentive Policies and Practices," March 12, 2012, https://www.osha.gov/as/opa/whistleblowermemo.html.

[68] See US DOL, OIG, *MSHA Has Taken Steps* (2014): 8-10.

For purposes of this study, the key point is not simply that incentive programs tend to deter the reporting of injures, but that their effects probably differ by injury type. Highly detectable injuries, especially severe ones, may be largely impervious to the influence of incentive programs. On the other hand, the reporting of less detectable injuries may be quite sensitive to their presence. In effect, even employers that *want* to fully report less-detectable injuries in the wake of an escalation in audit stringency may be unable to do so if they operate incentive programs that discourage employees from reporting injuries in the first place. Although in principle employers could suspend these programs, the benefits of doing so may be outweighed by the (perceived or actual) benefits of incentivizing workers to take greater care on the job. In short, the prevalence of incentive programs may cause some "stickiness" in the responsiveness of nontraumatic injury reporting to shifts in regulatory intensity.

To set the stage for the empirical analysis, I first manipulated the dataset to make it more amenable to statistical modeling. I dropped all fatal injuries, all incidents not resulting in an injury, all illnesses[69], all incomplete observations, and all "over-reported" injuries (i.e., those that were reported erroneously because they were not subject to OSHA reporting requirements) from the sample.[70] Using the injury-level fields that are contained in both datasets – source of injury, degree

[69] Although the theoretical model applies equally to injuries and illnesses, the difficulty of tracking illnesses in my dataset made them poorly suited for inclusion in the study.

[70] Besides fatalities, the following illnesses and injuries were categorically excluded from the OSHA data: work-related sicknesses; tuberculosis, bloodborne, and chronic obstructive pulmonary disorders; eczema; asthma; dust diseases of the lungs. The OSHA dataset also contains 98 injuries labeled "other occupational illnesses." All but 13 of these observations could be identified as more specific injuries and re-categorized using information in three text fields: injury description, injury narrative, and inspector comments. The following injuries were re-categorized: 34 instances of chipped or fractured teeth were marked as fractures, 20 instances of rashes, insect bites, scabies, and hives were marked as skin diseases, 11 respiratory conditions due to toxic agents, 7 bee stings, 6 eye injuries, 2 sprains and strains, 1 welding flash was categorized as disorders due to physical agents, 1 instance of hearing loss, 1 burn, 1 dislocation, and 1 puncture wound. The remaining 13 injuries could not be clearly identified and were dropped from the dataset. (The exclusion of these 13 observations did not substantively change any results.) The OSHA data contained 158 observations that were missing injury categories: 86 cases were categorized based on the three text fields (same as above) and 72 were dropped because they could not be categorized due to a lack of information on the injury, because they concerned pre-existing injuries, or because they were illnesses (infections, funguses, cysts, etc.) The 86 injuries

of injury, nature of injury, and affected body part – I then constructed four different indicator variables to capture injury characteristics: "severe," "traumatic," "easy-to-attribute," and "intermediate." The first three characteristics, as noted above, were developed through consultation with an expert in occupational medicine.[71] The fourth, which pertains exclusively to mining, was added because of its inclusion in a landmark historical study of mine safety.[72]

The injury characteristics upon which the empirical analysis rests – traumatic, severe, and hard-to-attribute – are defined in ways that are partially overlapping, so as to isolate different segments of the injury detectability continuum. The most fundamental distinction, traumatic/nontraumatic, divides the continuum into a pair of less-detectable categories (Types A and B) and a pair of more-detectable categories (Types C and D). The remaining categories further subdivide each respective half of the continuum. Among nontraumatic injuries, those that are hard-to-attribute (Type A) are assumed to be the least detectable of all. Meanwhile, among traumatic injuries, those that are also severe (Type D) are hypothesized to be the most detectable of all.

I also created a continuous variable ("employment") to reflect the number of workers employed (in hundreds), and four interaction terms between employment and injury type. To control for geographic variation and variation over time, I created dummy variables for the state and year in which the injury occurred. I also added a dummy indicating whether the establishment was unionized. (Although union status is available for all OSHA-regulated establishments, it is only

---

were categorized as follows: 72 sprains and strains, 6 instances of skin-related injuries, 3 eye injuries, 2 bee stings, 2 instances of hearing loss, and 1 instance of heat exhaustion. (The inclusion of these 86 observations did not significantly alter the results.) For MSHA, the following illnesses and injuries were excluded (in addition to fatalities) due to their questionable work-relatedness or their inability to be clearly categorized: pneumoconiosis/black lung; silicosis; occupational diseases, not elsewhere classified (NEC); unclassified, not determined; heart attack; cerebral hemorrhage; and other injury not elsewhere classified (NEC).

[71] See *supra* notes 1 and 59.

[72] See National Research Council, *Toward Safer Underground Coal Mines* (1982). Injuries classified as "intermediate" are those involving permanent total or permanent partial disability and those resulting from entrapment, falling/sliding/rolling materials, roof falls, haulage and machinery, electrical accidents, explosions, hoisting, impoundment, fire, or inundation.

available for coal mines.) Industry dummies were created using 2-digit SIC code for each non-mining establishment, and canvass code[73] for each mining establishment. Finally, in the MSHA-specific models, I created a dummy variable indicating whether or not the mine operates underground, and (in most models) interaction terms between underground status and injury type.

The empirical analysis unfolds in three stages. In Stage One, I test the model's predictions regarding which injury characteristics, if any, significantly predict the likelihood of underreporting and how these results vary across regimes. Specifically, I fit probit models in which the unit of observation is the injury and the (binary) dependent variable indicates whether the injury was reported. I estimate four different specifications: one pertaining exclusively to OSHA-reportable injuries and three pertaining exclusively to MSHA-reportable injuries. Of the three MSHA-specific models, one includes all mines, one restricts the sample to underground mines, and the third restricts the sample to coal mines. (The latter specification is included to test the effect of union status, which is unavailable for non-coal mines.) Most of the covariates used – severe, severeXemployment, traumatic, traumaticXemployment, easy-to-attribute, easy-to-attributeXemployment, employment, state dummies, year dummies and industry dummies – are common to all models. However, three independent variables (intermediate, intermediateXemployment, and underground) apply only to mining establishments. Standard errors are clustered on audit in the OSHA models and on mine identification number in the MSHA models.

In Stage Two, I explore several more nuanced theoretical predictions. First, I probe the height and topography of the underreporting staircase descended by a typical firm in both regulatory regimes. The goal is to pin down where the steepest step(s) of the staircase occur in each regime. Although the theoretical model does not indicate where they will occur, it does predict that *they will*

[73] The five canvass codes include Coal, Sand and Gravel, Stone, Nonmetal, and Metal.

tend to occur at lower levels of detectability in the stricter regime. In other words, relative to the OSHA regime, the steep step(s) in the MSHA regime should occur closer to the origin. In the most intensely inspected environment of all, underground mines, the steep step(s) should occur closer still to the origin. Of course, given the discrete nature of the categories analyzed and the possibility that some shifts will occur within (instead of between) categories, not all such shifts may be detectable. Yet by and large, as regulatory intensity increases, the staircase should tend to "migrate" toward the origin. Finally, since underground mines are inspected twice as frequently as surface mines but otherwise are highly comparable, one would also expect the likelihood of underreporting (i.e., the height of the staircase) to be lower overall in the underground environment.[74]

To estimate the height of each step, I estimate a probit model in which the dependent variable indicates whether the injury was reported This time, however, the main covariates of interest across all models are the *marginal* declines in reporting probability associated with a change in injury type from traumatic/severe to traumatic/nonsevere (D to C); from traumatic/nonsevere to nontraumatic/easy-to-attribute (C to B); and from nontraumatic/easy-to-attribute to nontraumatic/ hard-to-attribute (B to A). In the MSHA models, additional covariates of interest are the dummy on underground mine and the interaction terms between each of the injury type dummies and underground mine. All models control for employment, employmentXinjury type, state, year, and industry. Standard errors are clustered on audit in the OSHA models and on mine ID in the MSHA models.

Stage Two of the analysis also tests several predictions of the model that pertain exclusively to OSHA. Specifically, I examine the impact of OSHA's transition from the RK program, which

[74] In principle, one would expect the likelihood of underreporting to be lower overall in the MSHA regime as compared to the OSHA regime. However, this prediction cannot be meaningfully tested because of the different content and construction of the datasets obtained from the two agencies, including the inability to collect data at the audit level in the MSHA context and the omission from the MSHA sample of all audited mines that underreported zero injuries. Therefore, I do not attempt to test the prediction that underreporting overall is lower in the MSHA environment.

audited injury logs from the years 1996-2006, to the short-lived NEP program, which spanned logs from the years 2007-2009. As noted earlier, NEP audits were significantly more stringent than the RK audits that preceded them, but they also audited a somewhat different mix of firms. Specifically, firms with DART rates below the mean for high-hazard industries (4.2 injuries and illnesses per 100 full-time employees) were initially targeted under the NEP program on the suspicion that they were most likely to be underreporting. Given the different mix of firms audited and the escalation in audit stringency bias, one would expect the overall frequency of unreported OSHA injuries to have spiked in 2007. If the model's assumptions regarding the relationship between audit stringency and injury detectability are correct, hard-to-attribute injuries should have been affected more than severe ones.[75]

To test the latter (OSHA-specific) predictions, I visually compare trends in the mean number of total, severe and hard-to-attribute injuries omitted from each year's logs. I also examine these trends more formally by estimating three negative binomial models, one for each injury category, in which the dependent variable is the number of injuries of that type not reported to OSHA and employment is used as an exposure term. The variables of interest are the year dummies. Since the only year to which valid comparisons can be made is 2007, the first year affected by the NEP program, that is the year omitted from the models. The unit of observation is the injury set, and all models include robust standard errors and control for employment, union status, state, and industry.

In Stage Three, I commence the hunt for red flags. Statistically speaking, my goal is to determine whether particular (observable) characteristics of establishments or injury sets help predict the frequency and/or percentage of unreported injuries. Although there are grounds for optimism that

[75] I cannot rule out the possibility that the high-hazard firms audited in the first year of the NEP program (i.e., those reporting a DART rate less than 4.2) under-reported different types of injuries than those audited under the RK program, and therefore that my findings regarding the varying effects of audit stringency bias on different types of injuries are spurious. However, given the substantial overlap in the industries audited (both programs audited high-hazard industries included in the OSHA Data Initiative, although the NEP program additionally audited firms in NAICS codes 311614 and 115210), this scenario seems unlikely. Even if the *overall* spike in unreported injuries is partly attributable to a change in the sample of firms selected for audits, it is hard to imagine why selection bias would also affect the composition of injuries that went unreported or the composition of injuries that were uncovered during audits.

such red flags exist, two factors may weaken my identification strategy in the MSHA context. First, unlike OSHA, no *bona fide* injury sets are available for MSHA at the firm level. As noted earlier, I created "quasi injury sets" by bundling together all injuries from a given calendar year that a given firm *untimely* reported to MSHA to all injuries in the same calendar year that were *timely* reported. Firms for which no unreported injuries ever came to light are omitted altogether. This approach may have introduced some measurement error into my estimates. Secondly, very few quasi injury sets are available for analysis. In contrast to the thousands of injury sets available for OSHA-specific analysis, the MSHA dataset contains only 342 quasi injury sets, and neither subsample contains more than 175. (Even these numbers overstate the effective variation in the sample since many mines appear in the dataset more than once.) In short, the MSHA-specific results from Stage Three should be regarded as tentative and preliminary.

For all firms reporting at least one injury, I estimate a negative binomial model in which the dependent variable is the number of unreported injuries in the (quasi) injury set, with the number of employees used as an exposure term. As always, I include employment, state, year, industry, underground (for MSHA models), and unionization (in all models except those including non-coal mines) as covariates. OSHA models include robust standard errors, and standard errors are clustered on mine ID in MSHA models. Since predicting percentages requires one to calculate a numerator as well as a denominator, targeting injury rates is generally more difficult than predicting injury counts (frequencies).[76] Nevertheless, I also estimate an ordinary least squares model with robust standard errors in which the covariates are identical but the dependent variable is the *percentage* of total injuries that were untimely reported.

[76] See Alison Morantz, "Final Project Report: Designing a Pilot Program for Strategic Mine Safety and Health Improvements through the Use of Surveillance Data to Guide Targeted Inspection Activities," NIOSH Research Contract 200-2009-28820, September 28, 2012 (available from author upon request).

In all of the models just described, the primary covariates of interest are dummy variables indicating whether the respective *percentages of reported injuries* classified as severe, traumatic, or easy-to-attribute place the employer in the top quartile of the relevant sample.[77] Upon reflection, however, it is not obvious *a priori* which sample is the most pertinent. If the agency's goal is simply to identify as many underreported injuries as possible, regardless of where they occur,[78] there are two possible methods that one might use to calculate the quartile dummies. One approach would be to generate quartile rankings for *all* employers in each regulatory regime. (For example, MSHA quartiles would be calculated across *all* mines and OSHA quartiles would be calculated across all OSHA audits.) Alternatively, one could calculate quartile dummies separately for each industry, and distinguish even further between surface and underground mines.

Which set of dummies will be the most predictive is an open empirical question. If the true underlying distribution of injuries varies significantly across industries, then opting for industry-specific dummies should minimize measurement error.[79] On the other hand, if average reporting rates also differ significantly across industries, then calculating quartile dummies across the *entire* sample should enhance their explanatory power. The latter approach also reduces sampling error stemming from the fact that some industries comprise only a few observations. Since the net effect of these factors is uncertain, I calculate three sets of dummies – one for each regulatory regime in its

---

[77] As a robustness check, I estimated models with dummies for the top 10%, 15%, and 20% instead of the top quartile. These estimates exhibited patterns quite similar to those presented here.

[78] Alternatively, an agency could decide to target firms that underreport a number/proportion of injuries that is disproportionate by the standards of their respective industries. The downside of this approach is that assuming there is some variation across industries in the prevalence of underreporting, some firms that were *not* targeted in highly noncompliant industries would actually be underreporting more injuries than firms from highly compliant industries that *were* targeted. In the analysis that follows, I assume that the agency's goal is to identify firms that underreport the most injuries, regardless of the industries to which they belong.

[79] Intuitively, using dummies that do not account for industry-specific variation could increase the likelihood of false positives and false negatives in the calculation of quartiles. For example, a relatively compliant firm in an industry with an inherently high proportion of traumatic injuries could be erroneously included in the top quartile, whereas a relatively noncompliant firm in an industry with an inherently low proportion of traumatic injuries could be erroneously excluded.

entirety, one for each industry, and one that further subdivides each mining industry into surface and underground mines – and select the one with the most predictive power.

Another definitional ambiguity involves the precise definition of "quartile." Depending on whether there are many tied values across observations – and if so, where these tied values occur across the frequency distribution – one might define "quartile" as either strictly greater than, or greater than or equal to, the frequency observed at the 75^{th} percentile. In some contexts, one of these definitions may not be feasible at all. (For example, if observed frequencies range from 0 to 2, and one third of all observations report a frequency of 2, defining quartile as strictly greater than the 75^{th} percentile would flag no observations at all.) To resolve this problem, I attempted to use a "greater than or equal to" definition in all models, but if this approach was not feasible because it flagged all observations, I used the "strictly greater than" definition instead.

The empirical analysis draws to a close by considering the case of firms that reported no injuries at all ("nonreporters"), for which no reported injury-type dummies could be calculated. The key policy question is how highly an agency should prioritize the inspection of nonreporters as compared to reporters (i.e., firms that report at least one injury). I approach this question in two ways. First, to explore the importance of observable establishment-level or mine-level characteristics for this group, I confine the sample to nonreporters and estimate a Poisson model in which the dependent variable is the number of injuries that are not reported to regulatory officials. The usual set of covariates (employment, industry, etc.) are included. Secondly, although my small sample sizes preclude me from drawing any firm conclusions, I use several descriptive techniques to compare the number of unreported injuries among reporters and nonreporters in each industry examined.

V. Results

Table 1 presents average marginal effects from the first stage of the analysis, in which my goal is to identify the most important injury-level predictors of underreporting.

For OSHA, the laxer inspection regime, the best single predictor of underreporting is the severity (or lack thereof) of the injury. The statistically significant and *positive* average marginal effect of the "easy-to-attribute" dummy, however, seems to contradict the theory. Although initially counterintuitive, this coefficient makes sense when one takes into account the presence of audit stringency bias. As noted earlier, audit stringency bias is likely to affect nontraumatic injuries, especially hard-to-attribute ones, the most. The fact that this dummy is positive and statistically significant only for OSHA, the laxer regime in which rigorous audits were only conducted for a brief period, lends credence to this hypothesis.

Similar logic could explain the statistical insignificance of the union dummy, which appears to contradict prior literature suggesting that unionized establishments underreport fewer nontraumatic injuries.[80] Although unions may increase the likelihood that injuries are initially reported, they might also increase the chances that unreported injuries are discovered during an audit (for example, by empowering workers selected for interviews to speak more candidly with auditors). In other words, the presence of a union could exacerbate the confounding influence of audit stringency bias.

Most of the other findings echo prior literature. As expected, the likelihood of underreporting declines uniformly and significantly with firm size. I also find significant variations by state, region, and industry.[81]

---

[80] See Morantz (2013) and National Research Council, *Toward Safer Underground Coal Mines* (1982).

[81] Notably, injuries taking place in New York establishments are significantly more likely to go unreported across the different models (except Model 4). In Model 1, injuries taking place in Arizona and Montana establishments are significantly more likely to go unreported while those taking place in California, Kentucky, Missouri, Mississippi, New

Although injury severity is also a significant predictor of underreporting in the main MSHA model (Model 2), the other two characteristics examined – whether the injury is traumatic and whether it is easy to attribute – also have significant predictive value. In the subsample that is subject to the greatest regulatory scrutiny of all, underground mines, the pattern is similar but the characteristic that lies farthest from the origin (injury severity) loses statistical significance. The coal-only model displays identical patterns to the underground-mine model, which is not surprising since most of the sample (about 91% of the observations and 70% of the clusters) consists of underground mines. In all of these regards, the MSHA findings are consistent with the model's predictions. Also as expected, the likelihood of underreporting also declines with firm size and I detect significant variations by state, region, and industry.[82]

Jersey, Nevada, South Dakota, and Virginia establishments are significantly less likely to go unreported. In the three MSHA models, injuries at West Virginia and Kentucky establishments are significantly more likely to go unreported. Pennsylvania is excluded (used as the basis for comparison) for all models. OSHA-regulated industries associated with an increased probability of underreporting include *Building Construction General Contractors and Operative Builders (15), Chemicals and Allied Products (28), Electronic and Other Electrical Equipment and Components, Except Computer Equipment (36),* and *Local and Suburban Transit and Interurban Highway Passenger Transportation (41)*, while *Construction Special Trade Contractors (17)* is associated with a decreased probability of underreporting. SIC Major Group 80 (Health Services) is excluded from Model 1 because it is the factor level appearing most frequently in the OSHA injury set level data. For OSHA, one time trend emerges from the injury-level analysis: the year 2006 is associated with a significant decrease in the probability that an injury went unreported, while 2007 is associated with a significant increase in the likelihood of underreporting. Interestingly, these two years are the last audited log-year of the RK program and the first audited log-year of the NEP program, respectively. Across MSHA models, 2001 is associated with a significant decrease in the likelihood of underreporting. The year 1996 is excluded as the baseline in all models.

[82] Notably, injuries taking place in New York establishments are significantly more likely to go unreported (0.05 to 0.19 higher probability) across the different models (except Model 4). In Model 1, injuries taking place in Arizona and Montana establishments are significantly more likely to go unreported while those taking place in California, Kentucky, Missouri, Mississippi, Nevada, South Dakota, and Virginia establishments are significantly less likely to go unreported. In the three MSHA models, injuries at Kentucky and West Virginia establishments are significantly more likely to go unreported. Pennsylvania is excluded (used as the basis for comparison) for all models. OSHA-regulated industries associated with an increased probability of underreporting include *Building Construction General Contractors and Operative Builders (15), Chemicals and Allied Products (28), Electronic and Other Electrical Equipment and Components, Except Computer Equipment (36),* and *Local and Suburban Transit and Interurban Highway Passenger Transportation (41)*. SIC Major Group 80 (Health Services) is excluded from Model 1 because it is the factor level appearing most frequently in the OSHA injury set level data. According to Models 2 and 3, injuries at stone mines are associated with a 0.05 to 0.08 increased probability of underreporting. Coal is the excluded industry for MSHA. For OSHA, one time trend emerges from the injury-level analysis: the year 2006 is associated with a significant decrease in the probability that an injury went unreported, while 2007 and 2008 are associated with a significant increase in the likelihood of underreporting. Interestingly, these years are the last audited log-year of the RK program and the first two audited log-year of the NEP program, respectively. Across MSHA models, there is not notable time trend. The year 1996 is excluded as the baseline in all models.

Interestingly, the marginal effect of the "intermediate" dummy is statistically insignificant in the first two MSHA models but significant and *positive* in the coal-only model. Given its failure to predict underreporting in any other model, combined with the absence of any prior literature confirming its hypothesized negative correlation with underreporting, the "intermediate" dummy is excluded from all subsequent models.[83]

The MSHA results do, however, exhibit two perplexing patterns. First, since underground mine are inspected twice as often as surface mines, the statistical insignificance of the "underground" dummy does not support the expectation that the *overall* likelihood of underreporting (i.e., the height of the staircase) declines in the stricter regime. This puzzling finding could indicate the combined effects of audit stringency bias and incentive programs. As noted earlier, the most stringent audits that MSHA conducts, its "PPOV" audits, have taken place almost exclusively at underground mines. Thus one would expect audit stringency bias to be the strongest in underground mines. The fact that underground mine operators have not responded to this increase in regulatory scrutiny by reporting more nontraumatic injuries (i.e., lowering the height of the staircase) could be due to the existence of incentive programs that constrain employers' capacity to report less-detectable injuries. In apparent corroboration of this hypothesis, the negative marginal effect of underground mines attains statistical significance if one drops all observations from 2006 (the first year of injury logs audited under the PPOV program) and after.

Stage Two of the analysis tests more detailed predictions of the model. Although the probit models presented in Table 2 superficially resemble those presented in Table 1, the injury-type variables are constructed differently. This time, each injury of a specified type and each injury whose detectability equals or is less than that of the specified type are coded as 1, while all injuries

[83] See note 10 and accompanying text for an explanation for why the variable was included.

of higher detectability than X are coded as 0.[84] In effect, this coding scheme enables one to examine *marginal differences between* adjacent injury categories.

As expected, the steepest step in the OSHA model lies at the juncture between traumatic nonsevere and severe injuries (Types C and D). Once again, the negative and significant marginal effect of Type A (hard-to-attribute) injuries likely reflects the predominance of audit stringency bias, i.e., the fact that unreported hard-to-attribute injuries are only rarely being uncovered by OSHA auditors.

The MSHA models in Table 2 bring to light additional subtleties. Once again, I estimate three specifications: one for the entire sample that differentiates between surface and underground mines, one for underground mines, and one for coal mines that differentiates between surface and underground mines and also includes a union dummy. As expected, the step between nontraumatic, easy-to-attribute and traumatic, nonsevere injures (C→B) is large and significant for all mines across all models. However, the step between nontraumatic, easy-to-attribute injuries and nontraumatic, hard-to-attribute ones (B→A) is also positive and significant for underground mines, although not among surface mines. In other words, there is *an additional step* between Type A and Type B injuries, but only for underground mines. All of these findings accord with the model's core predictions. As before, the union dummy is uniformly insignificant, and the data exhibit significant variation by year, state and industry.[85]

---

[84] In Table 2, the "specified type" in question refers to the type to which the arrow is pointing. For example, in the first and sixth lines of Table 2, the reference to "Type D→C" implies that each injury of Type C as well as each injury of Types A and B is coded as 1, while each injury of Type D is coded as 0.

[85] Again, injuries taking place in New York establishments are significantly more likely to go unreported across the different models (except Model 4). In Model 1 (OSHA), injuries occurring in Arizona and Montana are significantly more likely to go unreported, whereas injuries occurring in California, Kentucky, Missouri, Mississippi, Nevada, South Dakota, and Virginia are significantly less likely to go unreported. In Models 2-4 (MSHA), injuries occurring in Kentucky and West Virginia are significantly more likely to go unreported. For OSHA, several industries – *Building Construction General Contractors And Operative Builders (15); Chemicals And Allied Products (28); Electronic And Other Electrical Equipment And Components, Except Computer Equipment (36); and Local, Suburban Transit And*

Turning to changes over time, Figure 4 explores OSHA's transition from the RK program to the NEP auditing program. The trend lines show annual means averaged across all injury sets. The top line represents total unreported injuries; the middle line represents unreported hard-to-attribute injuries; and the bottom line depicts unreported severe injuries. As noted earlier, the rigor of each audit increased markedly under the NEP program, although the average annual frequency of audits slightly declined. A shift of this type not only should have increased the overall frequency of unreported injuries because of audit stringency bias, but also should have had a larger impact on hard-to-attribute injuries than severe ones. The trends displayed in the figure align well with the model's predictions. Although the mean number of hard-to-attribute (and total) injuries that went unreported rose dramatically in 2007, the rise in unreported severe injuries was negligible. Table A (included as an appendix) formalizes these comparisons by estimating negative binomial models in which the unit of the analysis is the injury set and the dependent variable is the frequency of (severe, hard-to-attribute, and total) unreported injuries, respectively.[86] The models confirm that relative to prior years, the transition to the NEP program in 2007 had a statistically significant impact on the frequency of unreported hard-to-attribute (and total) injuries, but not on the frequency of unreported severe injuries.

In Stage Three, the key question becomes whether scrutinizing establishment-level characteristics and the composition of reported injuries could help regulators identify which firms underreport the most. This portion of the analysis relies exclusively on the injury set as the unit of analysis.

Interurban Highway Passenger Transport (41) – predict a significantly increased probability of underreporting. The MSHA Models 2 and 3 show that injuries occurring at stone mines are significantly more likely to be underreported. The time trends in Table 2 mirror those in Table 1. The year 1996 is excluded as the baseline in all models. Pennsylvania is excluded from the state dummies; coal and SIC Major Group 80 - Health Services are the excluded industry dummies.

[86] All models include robust standard errors, and the number of employees is used as an exposure term. Other model covariates include the number of employees (in hundreds), an indicator of union status, and state and industry dummies.

Based on the results presented in Table 3, the answer is a tentative yes. Having a relatively high percentage (within the top quartile) of reported injuries that are highly detectable is a statistically significant predictor of underreporting in both regimes. As one would expect given the findings from Stages 1 and 2, the precise nature of the red flag varies across models. In the OSHA regime, a reported percentage of *severe* injuries that places the employer in the top quartile of the sample predicts a large (roughly 57%) and statistically significant increase in unreported injuries. Meanwhile, in the full MSHA sample, a reported percentage of *easy-to-attribute* injuries that places the employer in the top quartile of the sample predicts a sizable (roughly 54%) increase in unreported injuries.[87] The top-quartile easy-to-attribute dummy loses statistical significance in both MSHA subsamples, however, presumably due to the small number of observations. Employment retains its robust negative association with the frequency of unreported injuries across all models. As before, there is significant variation across states, industries, and years.[88]

The union dummy displays an erratic pattern: it is significant and *greater* than 1 in the OSHA model, but statistically insignificant in the MSHA coal model. Given its uniform statistical

---

[87] These findings are consistent regardless of whether one calculates the top-quartile dummies across the whole sample or separately for each industry, although the coefficient estimates fluctuate slightly. (The coefficients presented in Table 3 define top-quartile dummies across the entire sample.) The precise definition of quartile also varies slightly across samples. In most models, an observation is flagged as being in the top quartile if the fraction of reported injuries is *greater than or equal to* that of the 75th percentile. Defining the dummy in this manner (as opposed to defining it as strictly greater than the 75th percentile) was the only way to avoid flagging zero injuries. The sole exception to this rule is the top-quartile dummy for severe injuries in Model 1, which was defined to include only observations that are *strictly greater than* the 75th percentile so as to avoid flagging all observations. At least 20 percent and no more than 30 percent of all observations were flagged in every model.

[88] Establishments in Missouri (except in Model 4) and establishments in Mississippi (except in Model 3) underreport a significantly smaller number of injuries. The MSHA models show that mines in Indiana underreport a significantly smaller number of injuries. In Model 1, several industries – *Building Construction General Contractors And Operative Builders (15); Food and Kindred Products (20); Furniture And Fixtures (25); Chemicals And Allied Products (28); Primary Metal Industries (33); Fabricated Metal Products (34); Industrial And Commercial Machinery And Computer Equipment (35); Electronic, Electrical Equipment And Components, Except Computer Equipment (36); Transit And Interurban Highway Passenger Transportation (41)* – predict a significantly greater number of unreported injuries. Models 2 and 3 predict that stone mines underreport a significantly greater number of injuries. In Model 1, the 2006 coefficient is less than one and significant at the .1% level, and the 2007 and 2008 coefficients are greater than one and significant at the .1% level. No clear time trends are apparent in the MSHA models. As before, Pennsylvania is excluded from the state dummies; 1996 is excluded from the year dummies; and coal and SIC Major Group 80 - Health Services are the excluded industry dummies.

insignificance in Stages 1 and 2 of the analysis, the surprise is that the dummy gains significance (and carries a positive sign) in the OSHA model. It is important to bear in mind that although the prior models compare all injuries to one another, Table 3 makes comparison across injury *sets*. Perhaps in the OSHA-regulated economy, unions play a vitally important role in enhancing the scope and rigor of audits, so that *ceteris paribus*, auditors find more violations at unionized establishments. If the increase in audit intensity bias is sufficiently large, it could outweigh unions' (presumed) salutary effect on the overall reporting of workplace injuries. Alternatively, perhaps incentive programs are more prevalent at unionized establishments.

Table 4 presents results from an OLS model in which the dependent variable is the *percentage* of total injuries that went unreported among establishments that reported at least one injury. This time, none of the covariates in the OSHA model has any predictive value. In the MSHA model, on the other hand, the coefficient on the easy-to-attribute top-quartile dummy remains sizable and statistically significant in the main model and in the coal subsample. For the first time, the coefficient on underground mine also attains statistical significance. Employment remains negative and significant in all three MSHA models, and the union dummy is uniformly insignificant. Although a few of the state, year, and industry dummies attain statistical significance within models, none is robust across all models.[89]

[89] No clear state trends are apparent across models although there is significant variation across states *within* each of Models 1-4. In Model 1, two industries – *Chemicals And Allied Products (28); Local, Suburban Transit And Interurban Highway Passenger Transport (41)* – predict a significantly higher percentage of unreported injuries, whereas several other industries – *Agricultural Production Crops (1); Agricultural Production Livestock And Animal Specialties (2); Agricultural Services (7); Water Transportation (44); Transportation Services (47); Communications (48)* – predict a significantly lower percentage of unreported injuries. Models 2 and 3 predict that stone mines underreport a significantly higher percentage of injuries. Similar to previous OSHA models, the 2006 coefficient is negative and significant at the 1% level in Model 1 while the 2007 coefficient is positive and significant at the 5% level. The MSHA models do not show clear time trends. As before, Pennsylvania is excluded from the state dummies; 1996 is excluded from the year dummies; and coal and SIC Major Group 80 - Health Services are the excluded industry dummies.

The final models presented in Table 5 again estimate the number of unreported injuries per (quasi) injury set, but this time only among "nonreporters," i.e., establishments that timely reported zero injuries. As expected, the frequency of unreported injuries declines uniformly with establishment size. There is some robust variation by industry, although the findings do not vary consistently by state or year.[90] The more policy-relevant question, however, is how the prevalence of underreporting among nonreporters compares to the prevalence of underreporting among firms that reported at least one injury ("reporters"). I approach this question in two ways. First, I compare the respective percentages of firms that were highly noncompliant, defined as failing to report more than one injury. Secondly, I compare the mean number of unreported injuries across the two groups. In making these comparisons, I divide each group into small, medium and large firms, and calculate mean values for each subgroup. The results of these analyses, presented in Table 6, suggest that nonreporters are generally no more likely than reporters to engage in underreporting. Indeed, by many measures, nonreporters in the OSHA sample actually underreport *fewer* injuries than reporters. In the MSHA sample, there are no statistically significant distinctions at all between the two groups.

Notwithstanding these broad trends, however, there are a few industries in which nonreporters *do* appear less compliant than their peers. In the OSHA setting, for example, non-unionized, medium-sized nonreporters in the Transportation, Communications, Electric, Gas and Sanitary Services industry (SIC Major Groups 41-45, 47-49) and the Wholesale Trade Industry (SIC Major Groups 50-51) performed worse than other firms in the same industries. Meanwhile, in the MSHA sample, nonreporters in the nonmetal mining sector appear to be less compliant than other

[90] Again, no clear state trends are apparent across models although there is significant variation across states *within* each of Models 1-4. In Model 1, several industries – *Heavy Construction, Except Building Construction – Contractors (16); Food And Kindred Products (20); Lumber And Wood Products, Except Furniture (24); Furniture And Fixtures (25); Motor Freight Transportation (42); Transportation By Air (45)* – predict a significantly greater number of unreported injuries. The MSHA models do no display clear industry trends. No clear time trends emerged in the results. As before, Pennsylvania is excluded from the state dummies; 1996 is excluded from the year dummies; and coal and SIC Major Group 80 - Health Services are the excluded industry dummies.

mines. Yet these findings should be viewed with caution. Given the small sample sizes, they could merely be statistical artifacts. Moreover, since the MSHA sample includes only those mines that untimely reported at least one injury, all of the MSHA comparisons should be construed as "worst-case depictions" of the relative prevalence of underreporting among nonreporters.[91]

VI. Practical Challenges Posed by Policy Implementation

My empirical findings suggest that regulatory intensity, injury detectability, and underreporting relate to one another in systematic and predictable ways that worker protection agencies could profitably explore. Because of the confounding effects of audit stringency bias and the "stickiness" of firms' reporting behavior due to incentive programs, testing the predictions of the theoretical model is not entirely straightforward. Nevertheless, my findings generally bear out its core predictions. More-detectable injuries are generally less likely to be underreported than less-detectable injuries, and the single best predictor of underreporting relates in coherent ways to the overall rigor of the inspection regime. The reporting of less-detectable injuries is likewise far more responsive to fluctuations in regulatory stringency. Stage Three of the analysis further suggests that analyzing the *composition* of injuries in mandated filings could help regulators anticipate which employers are most likely to underreport.

Yet my theoretical model is relatively static in the sense that regulatory behavior is treated as largely exogenous, and firms are assumed to respond in straightforward, predictable ways to changes in the likelihood of detection. In other words, the model assumes that the interaction between agency and employer closely resembles a one-shot game. If one were to relax this assumption and assume that *both* sides in the enforcement game behave in an iterative and strategic fashion, how might this

[91] This is so because the sample excludes mines that actually had, and likewise reported, zero injuries during a given fiscal year.

affect the model's predictions? For example, if firms learned of the precise algorithm being used, could they respond in ways that would thwart the agency's capacity to implement this approach?

Although development of a full game-theoretic model is beyond the scope of this paper, several practical considerations suggest that employers could not easily strip the agency of its targeting capacity. Suppose, for example, that an employer who engages in extensive underreporting, and therefore would ordinarily fall into the top quartile of reported traumatic injuries, has just learned of the new targeting strategy and wishes to avoid being targeted. Realistically, there are only three strategies at its disposal. First, the employer could start to fully report all nontraumatic injuries. If this occurs, the program will achieve precisely its intended effect of deterring underreporting. The more important question, therefore, is whether an employer can evade detection in ways that thwart the policy's intended goals. Two possibilities come to mind. First, an employer could choose to report even fewer *traumatic* injuries. Second, an employer could report "extra" nontraumatic injuries, such as those that do occur but are not reportable.[92] Either approach would lower the percentage of nontraumatic injuries reported, and thereby lessen the likelihood that an employer will be audited.

If employers responded in this way, however, the agency could develop counter-measures to deter such behavior. For example, if employers responded by reporting even fewer traumatic injuries, the agency could also target employees whose total injury count fell below industry norms[93], or increase the penalties for underreporting traumatic injuries. If some employers responded by reporting injuries it knew were not reportable, the agency could respond by penalizing

[92] See note 57 for a discussion of "overreporting" in the OSHA dataset.

[93] This was precisely the targeting criteria used during much of OSHA's NEP program, which targeted establishments in high-hazard industries that participated in the OSHA Data Initiative (ODI and reported a DART (days away, restricted, or transferred) injury rate between 0 and 4.2. See US DOL, OSHA, Directive Number: 09-08 (CPL 02) (2009), and US DOL, OSHA, Directive Number: 10-07 (CPL 02) (2010).

persistent over-reporting. In short, relaxing the assumption of exogeneity does not automatically doom a targeting strategy such as that outlined here. It simply means that worker protection agencies, like the IRS, would have to continuously adjust their targeting criteria (and perhaps also their penalty structure) in response to observed changes in industry behavior.

Another important issue that bears on policy implementation is the likelihood that the *social cost* of underreporting varies by injury type. For example, the choice of whether to devote the marginal auditing dollar to increasing the frequency of inspections, or to increasing the rigor of inspections, depends in part on the *social cost* associated with undercounting different types of injuries. Since musculoskeletal diseases are the costliest injuries to treat, an agency might reasonably devote its marginal inspection dollar to increasing inspection rigor and thereby improving the reporting of hard-to-attribute (Type A) injuries. Alternatively, the agency might choose to conduct more frequent inspections to improve the measurement of traumatic yet nonsevere (Type C) injuries, whose work-relatedness is usually easier to ascertain. The optimal balance between inspection frequency and rigor cannot be determined in a vacuum, but only with reference to such broader policy considerations.

VII. Conclusion

Like the Internal Revenue Service, several of the Department of Labor's worker protection agencies receive self-reported filings from U.S. establishments on the quantity and magnitude of regulated activities.[94] Each agency relies, at least in part, on this self-reported information to calibrate the frequency and intensity of its regulatory inspections and audits. Yet self-reported injury filings – like income tax filings – are prone to underreporting. Unlike the IRS, which has developed

[94] It should be noted that in contrast to the IRS and MSHA, which receive information at least once per year from all regulated establishments, OSHA only receives such information from a sizable subset of establishments in high-hazard industries. Nevertheless, the same essential principles apply.

proprietary algorithms to detect telltale patterns or red flags of underreporting in income tax filings, neither OSHA nor MSHA has developed robust statistical techniques to ferret out the underreporting of occupational injuries and illnesses. An important first step toward this goal is determining which *types* of injuries are most (or least) likely to evade detection.

I take up this challenge by completing four interrelated tasks. First, I develop an informal theoretical model of the relationship between regulatory intensity, injury type, and underreporting. The model yields concrete predictions about how injury reporting behavior will respond to changes in the frequency and/or stringency of auditing. Secondly, I propose a scheme whereby different types of injuries can be classified according to their relative detectability. Third, using a dataset that encompasses thirteen years of granular audit data obtained from OSHA and MSHA, I test the model's predictions regarding which types of injuries will be underreported the most, both across regimes and over time. Finally, I explore whether any readily observable, establishment-level covariates – such as the percentage of reported injuries that are highly detectable – could be used by labor regulators to identify which employers are underreporting the most injuries.

By and large, my findings provide considerable grounds for optimism that the model reasonably approximates reality. The best predictors of underreporting in less rigorous regimes differ from the best predictors of underreporting in more rigorous regimes in precisely the ways that theory predicts. The model's more nuanced predictions are sometimes difficult to test because of the confounding influence of audit stringency bias, the simultaneous adoption of multiple regulatory reforms, and the prevalence of incentive programs that induce workers to hide their injuries. Nevertheless, my findings still largely bear out the model's theoretical predictions. For example, OSHA's transition to a more rigorous auditing regime around 2007 had a dramatic effect on the frequency of unreported hard-to-attribute injuries, but no statistically significant effect on the frequency of unreported severe injuries. Most importantly, at least one formulation of the red flag explored – a reported percentage of highly-detectable injuries that places a firm within the top quartile of its respective sample – does have significant predictive value. Specifically, having a high percentage of reported injuries that are *severe* is a significant predictor of the number of underreported injuries in the OSHA regime, whereas having a high percentage of reported injuries that are *easy to attribute* is a significant predictor of both the frequency and percentage of underreported injuries in the MSHA regime. Overall, my findings provide grounds for optimism that DOL's worker protection agencies could use data-mining techniques to combat underreporting in a manner similar to their counterparts at the IRS.

The paper also briefly discusses several challenges associated with making the leap from theory to practice. I suggest that even if one relaxes the assumption of exogeneity, so that both regulators and firms behave in an iterative and strategic manner, the use of targeting techniques to identify the most likely violators could significantly improve upon the status quo. I also argue that to find the optimal tradeoff between the frequency and stringency of audits, regulators should weigh the respective social costs associated with underreporting different types of injuries.

My findings point to several promising areas for future research. First, the theoretical model proposed here could be tested in other settings to explore whether it is generalizable to other time periods and enforcement regimes. Finding new datasets on which to test its core predictions is particularly vital, since the datasets used here were prone to both measurement error and selection bias. Conducting audits on a stratified random sample of *all* establishments within a particular industry would yield the most valuable insights.[95] Secondly, the injury classification scheme proposed, including the categorization of various injuries as severe, traumatic, and/or hard-to-

[95] In a similar vein, a report issued on 2014, the Department of Labor's Office of the Inspector General emphasized the importance of "deriving betters estimates of [underreporting's] overall occurrence" by performing random audits from one or more sectors of the mining industry. See US DOL, OIG, *MSHA Has Taken Steps* (2014).

attribute, could potentially be refined in ways that would enhance its predictive value. Finally and most importantly, it remains to be seen whether red flags similar to those developed here – such as the respective percentages of severe or easy-to-attribute injuries reported to regulatory officials – could become tools of practical value to DOL regulators. Using field experiments and other experimental techniques to explore their predictive accuracy would give inspectors new leverage in their efforts to combat underreporting, and stimulate the development of more sophisticated targeting methods.

Table 1: Average Marginal Effect of Injury⊕ and Establishment Characteristics on the Probability that an Injury is Not Reported: Probit Models Testing Basic Injury Characteristics

	1. OSHA	**2. MSHA**	**3. MSHA** (Underground Only)	**4. MSHA** (Coal Only)
Severe Inj. (Type D)	-0.066***	-0.020*	-0.018	-0.016
	(0.01)	(0.01)	(0.01)	(0.01)
Traumatic Inj. (Types C, D)	-0.016	-0.073***	-0.070***	-0.091***
	(0.01)	(0.01)	(0.02)	(0.02)
Easy-to-Attr. Inj. (Types B, C, D)	0.036**	-0.024*	-0.037**	-0.035**
	(0.01)	(0.01)	(0.01)	(0.01)
Intermediate Inj.		0.016	0.016	0.024*
		(0.01)	(0.01)	(0.01)
Employment	-0.000***	-0.017**	-0.022***	-0.024***
	(0.00)	(0.01)	(0.00)	(0.00)
Underground Mine		-0.025		-0.021
		(0.02)		(0.02)
Union	0.009			-0.017
	(0.01)			(0.01)
Injury Char.×Employment Interactions	Yes	Yes	Yes	Yes
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	No
Number of Injuries	12093	6415	3995	3666
Number of Clusters	2523	291	104	117
Mean of Dependent Variable	0.127	0.096	0.070	0.080

⊕ Injury types are as follows: Type A is nontraumatic and hard-to-attribute; Type B is nontraumatic and easy-to-attribute; Type C is traumatic and nonsevere; and Type D is traumatic and severe.

Model: The model is a probit regression. Average marginal effects are presented. Standard errors are clustered on the audit identification number (OSHA) or on the mine identification number (MSHA). Significance levels: *** .1%, ** 1%, * 5%.

Dependent Variable: The dependent variable is set to 1 if an injury was not recorded on the audited OSHA 200 or 300 Log or if an injury was reported to MSHA after the year close for the calendar year of the injury, and 0 otherwise.

Covariates: All models include indicators of whether an injury is classified as severe, traumatic, or easy-to-attribute; the number of employees at the establishment (in hundreds); interactions between each injury type indicator and employment; and state and year dummies. Models 1-3 include controls for industry. For MSHA models, an indicator of whether an injury is intermediate is included, and industries are the canvas codes (coal, sand and gravel, stone, nonmetal, and metal). Models 1 and 4 include an indicator of union status at the injury site for the calendar year of the injury. Models 2 and 4 include an indicator of whether the injury occurred at an underground mine.

Unit of Observation: The unit of observation is the injury.

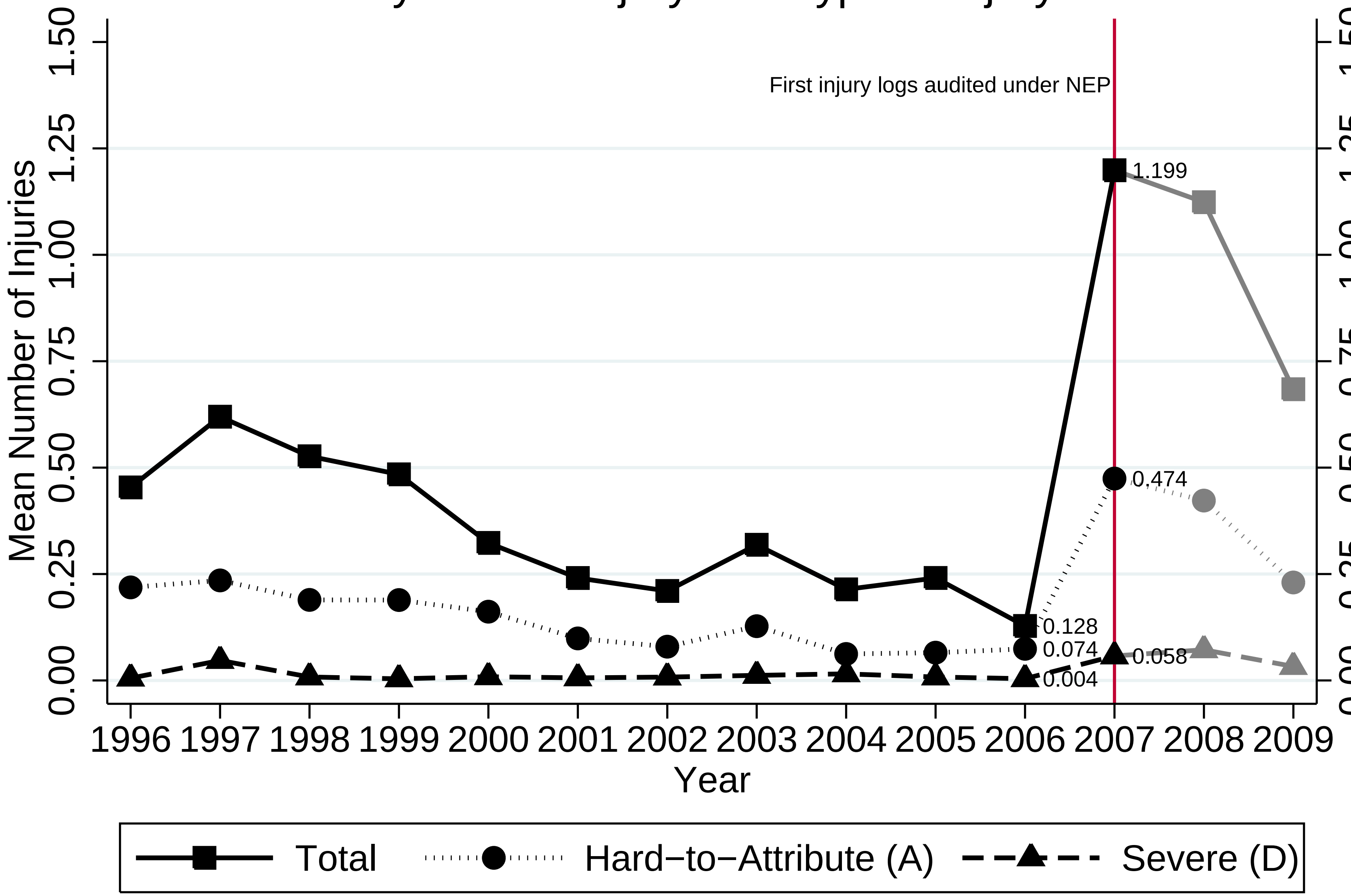
Sample: In Model 1, the sample includes reported and unreported injuries (excluding occupational illnesses) collected during audits conducted under the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009), covering injury logs from the years 1996-2006, and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012), covering injury logs from the years 2007-2009. In Model 2, the sample includes MSHA reported and late-reported injuries (excluding occupational illnesses, fatalities, and accidents not resulting in injuries) for any mine-year in which at least one injury was reported after the year close. Model 2 includes injuries from 1992, 1994, 1996-1998, and 2000-2012. Before 2000, the data only includes one mine for each year. Model 3 is limited to injuries in the MSHA data that occurred at underground mines and includes injuries from 2000-2012. Model 4 is limited to injuries in the MSHA data that occurred at coal mines and includes injuries from 2000-2001 and 2003-2012. For details on sample construction and omitted observations, see footnote 70. In Models 1-3, some observations were dropped due to their having attributes that perfectly predict success or failure. Following is the number of observations dropped in each model; Model 1: 135 obs, Model 2: 7 obs, Model 3: 3 obs.

Table 2: Average Marginal Effect of Injury Type$^{\oplus}$ and Establishment Characteristics on the Probability that an Injury is Not Reported: Probit Models Testing Marginal Differences Between Injury Types

	1. OSHA	**2. MSHA**	**3. MSHA** (Underground Only)	**4. MSHA** (Coal Only)
Type D→C [at Surface Mine in Models 2 & 4]†	0.069***	0.029		0.005
	(0.01)	(0.02)		(0.02)
Type C→B [at Surface Mine in Models 2 & 4]†	0.018	0.051*		0.146***
	(0.02)	(0.02)		(0.04)
Type B→A [at Surface Mine in Models 2 & 4]†	-0.038*	0.027		0.061
	(0.02)	(0.02)		(0.06)
Employment	-0.000***	-0.017**	-0.022***	-0.024***
	(0.00)	(0.01)	(0.00)	(0.00)
Underground Mine		[-0.046*]^		[0.010]^
		[(0.02)]		[(0.02)]
Type D→C at Underground Mine†		0.002	0.007	0.005
		(0.01)	(0.01)	(0.01)
Type C→B at Underground Mine†		0.076***	0.060***	0.066***
		(0.02)	(0.02)	(0.02)
Type B→A at Underground Mine†		0.042**	0.056***	0.061***
		(0.02)	(0.01)	(0.02)
Union	0.009			-0.017
	(0.01)			(0.01)
Injury Type×Employment Interactions	Yes	Yes	Yes	Yes
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	No
Number of Injuries	12093	6415	3995	3666
Number of Clusters	2523	291	104	117
Mean of Dependent Variable	0.127	0.096	0.070	0.080

⊕ Injury types are as follows: Type A is nontraumatic and hard-to-attribute; Type B is nontraumatic and easy-to-attribute; Type C is traumatic and nonsevere; and Type D is traumatic and severe.
† These are not the covariates included in the model. See Covariates for a full list of the covariates used in each probit model.
∧ The marginal effect of underground at Type C is presented. The marginal effects of underground at other injury types is not shown and were not significant at the 5% level for either model.
Model: The model is a probit regression. Average marginal effects are presented. Standard errors are clustered on the audit identification number (OSHA) or on the mine identification number (MSHA). Significance levels: *** .1%, ** 1%, * 5%.
Dependent Variable: The dependent variable is 1 if an injury was not recorded on the audited OSHA 200 or 300 Log or if an injury was reported to MSHA after the year close for the calendar year of the injury, and 0 otherwise.
Covariates: While the table above presents marginal effects, the probit models include the following covariates. Each model includes injury type indicators for injury types C, B, and A, which are 1 for that injury type and also for injury types of lesser detectability (see Figure 3). Also included are the number of employees at the establishment (in hundreds), interactions between each injury type indicator and employment, and state and year dummies. Models 1-3 include controls for industry. For MSHA, industries are the canvass codes (coal, sand and gravel, stone, nonmetal, and metal). Models 1 and 4 include union status at the establishment for the calendar year of the injury. Models 2 and 4 include an indicator of whether the injury occurred at an underground mine, and interaction terms between each injury type indicator and the underground dummy.
Unit of Observation: The unit of observation is the injury.
Sample: In Model 1, the sample includes reported and unreported injuries (excluding occupational illnesses) collected during audits conducted under the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009), covering injury logs from the years 1996-2006, and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012), covering injury logs from the years 2007-2009. In Models 2-4, the sample includes MSHA reported and late-reported injuries (excluding occupational illnesses, fatalities, and accidents not resulting in injuries) for any mine-year in which at least one injury was reported after the year close. Model 2 includes injuries from 1992, 1994, 1996-1998, and 2000-2012. Model 3 is limited to underground mine injuries and includes injuries from 2000-2012. Model 4 is limited to coal mine injuries and includes injuries from 2000-2001 and 2003-2012. For details on sample construction and omitted observations, see footnote 70. In Models 1-3, some observations were dropped due to their having attributes that perfectly predict success or failure. Following is the number of observations dropped in each model; Model 1: 135 obs, Model 2: 7 obs, Model 3: 3 obs.

Figure 4: Mean Number of Unreported Injuries per Audit, OSHA, by Year of Injury and Type of Injury



Notes: Each point represents the mean across injury sets (i.e. audits) for each year. Each year represents the year of the injury logs that were audited (not the year in which the audit took place). NEP stands for National Emphasis Program. The years 2008 and 2009 are grayed out because the selection criteria for establishments and industries changed under NEP2 (2007 and 2008) and NEP3 (2008 and 2009); thus, these years are not direclty comparable to prior years. See Section II: Description of Data in the paper for more details.

Unit of Observation: The unit of observation is the injury set (i.e. audit).

Sample: The sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (1996-2009), collectively covering injury logs from the years 1996-2009. For more details on sample construction and omitted observations, see footnote 70.

Table 3: Effect of Reported Injury Distribution and Establishment Characteristics on Number of Unreported Injuries (Among Establishments and Mines Reporting at Least One Injury)

	1. OSHA	**2. MSHA**	**3. MSHA** (Underground Only)	**4. MSHA** (Coal Only)
Top Quartile, % Severe Inj.	1.569***	1.026	0.982	1.026
	(0.18)	(0.13)	(0.16)	(0.17)
Top Quartile, % Traum. Inj.	0.854	1.041	1.064	0.974
	(0.13)	(0.18)	(0.26)	(0.22)
Top Quartile, % Easy-to-Attr. Inj.	0.875	1.544**	1.154	1.432
	(0.14)	(0.20)	(0.22)	(0.27)
Employment	0.898***	0.786**	0.712***	0.709***
	(0.01)	(0.07)	(0.03)	(0.03)
Underground Mine		0.926		1.056
		(0.16)		(0.20)
Union	1.440**			0.879
	(0.17)			(0.15)
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	No
Number of (Quasi) Injury Sets	2556	342	169	175
Number of Clusters		233	101	110
Mean of Dependent Variable (Raw)	0.523	1.541	1.627	1.640

Model: The model is a Negative Binomial, selected because the sample variance of the dependent variable far exceeds the sample mean. Coefficients are presented as Incident Rate Ratios (IRR). The number of employees is used as an exposure term. Robust standard errors are used for OSHA, and standard errors are clustered on the mine identification number for MSHA. Significance Levels: *** .1%, ** 1%, * 5%.

Important Caveat: The sample used for the OSHA model includes some establishments which do not underreport at all. This is not true of the MSHA samples used in Models 2, 3, and 4; all mines in these samples underreport at least one injury.

Dependent Variable: The dependent variable is the number of unreported injuries in the (quasi) injury set.

Covariates: Top-quartile dummies in each model indicate whether or not the reported fractions of severe, traumatic, and easy-to-attribute injuries in the firm's (quasi) injury set place it within the top quartile of the sample analyzed. Covariates invariably include the number of employees at the establishment (in hundreds), and state, year and industry dummies. In most models, an observation is flagged as being in the top quartile if the fraction of reported injuries is *greater than or equal to* that of the 75th percentile. Defining the dummy in this manner (as opposed to defining it as strictly greater than the 75th percentile) was the only way to avoid flagging zero injuries. The sole exception to this rule is the top-quartile dummy for severe injuries in Model 1, which was defined to include only observations that are *strictly greater than* the 75th percentile to avoid flagging all observations. In each model, at least 20 percent and no more than 30 percent of all observations are flagged. Models 1 and 4 include a union status indicator, and models 2 and 4 include an indicator for whether the injuries in the quasi injury set occured in an underground mine.

Unit of Observation: The unit of observation is the (quasi) injury set.

Sample: In Model 1, the sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009 that included at least one reported injury. In Models 2, 3, and 4, the sample includes MSHA quasi injury sets for mine-years with at least one reported injury and one injury reported after the year close. Model 2 includes quasi injury sets from 1992, 1994, 1996-1998, and 2000-2012. Before 2000, the data only includes one quasi injury set for each year. Model 3 excludes all quasi injury sets that are not from an underground mine and includes quasi injury sets from 2000-2012. Model 4 excludes all quasi injury sets that are not from a coal mine and includes quasi injury sets from 2000-2001 and 2003-2012. For more details on sample construction and omitted observations, see footnote 70.

Table 4: Effect of Reported Injury Distribution and Establishment Characteristics on Percentage Unreported Injuries (Among Establishments and Mines Reporting at Least One Injury)

	1. OSHA	2. MSHA	3. MSHA (Underground Only)	4. MSHA (Coal Only)
Top Quartile, % Severe Inj.	0.008	0.018	0.018	0.009
	(0.01)	(0.02)	(0.02)	(0.03)
Top Quartile, % Traum. Inj.	-0.007	0.003	-0.011	0.007
	(0.01)	(0.02)	(0.03)	(0.03)
Top Quartile, % Easy-to-Attr. Inj.	0.003	0.127***	0.050	0.094**
	(0.01)	(0.02)	(0.03)	(0.03)
Employment	0.000	-0.015*	-0.028***	-0.026***
	(0.00)	(0.01)	(0.00)	(0.01)
Underground Mine		-0.061*		-0.100**
		(0.02)		(0.03)
Union	0.017			-0.047
	(0.01)			(0.02)
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	No
Number of (Quasi) Injury Sets	2556	342	169	175
Number of Clusters		233	101	110
Mean of Dependent Variable (Raw)	0.072	0.190	0.138	0.162

Model: The model is OLS. Robust standard errors are used for OSHA, and standard errors are clustered on the mine identification number for MSHA. Significance Levels: *** .1%, ** 1%, * 5%.

Important Caveat: The sample used for the OSHA model includes some establishments which do not underreport at all. This is not true of the MSHA samples used in Models 2, 3, and 4; all mines in these samples underreport at least one injury.

Dependent Variable: The dependent variable is the percentage of unreported injuries in the (quasi) injury set.

Covariates: Top-quartile dummies in each model indicate whether or not the reported fractions of severe, traumatic, and easy-to-attribute injuries in the firm's injury set place it within the top quartile of the sample analyzed. Covariates invariably include the number of employees at the establishment (in hundreds), and state, year and industry dummies. In most models, an observation is flagged as being in the top quartile if the fraction of reported injuries is *greater than or equal to* that of the 75th percentile. Defining the dummy in this manner (as opposed to defining it as strictly greater than the 75th percentile) was the only way to avoid flagging zero injuries. The sole exception to this rule is the top-quartile dummy for severe injuries in Model 1, which was defined to include only observations that are *strictly greater than* the 75th percentile to avoid flagging all observations. In each model, at least 20 percent and no more than 30 percent of all observations are flagged. Models 1 and 4 include a union status indicator, and models 2 and 4 include an indicator for whether the injuries in the quasi injury set occured in an underground mine.

Unit of Observation: The unit of observation is the (quasi) injury set.

Sample: In Model 1, the sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009 that included at least one reported injury. In Models 2, 3, and 4, the sample includes MSHA quasi injury sets for mine-years with at least one reported injury and one injury reported after the year close. Model 2 includes quasi injury sets from 1992, 1994, 1996-1998, and 2000-2012. Before 2000, the data only includes one quasi injury set for each year. Model 3 excludes all quasi injury sets that are not from an underground mine and includes quasi injury sets from 2000-2012. Model 4 excludes all quasi injury sets that are not from a coal mine and includes quasi injury sets from 2000-2001 and 2003-2012. For more details on sample construction and omitted observations, see footnote 70.

Table 5: Effect of Establishment Characteristics on Number of Unreported Injuries (Among Establishments and Mines Reporting Zero Injuries)

	1. OSHA	**2. MSHA**	**3. MSHA** (Underground Only)	**4. MSHA** (Coal Only)
Employment	0.624***	0.242***	0.004***	0.004***
	(0.08)	(0.05)	(0.00)	(0.00)
Underground Mine		0.537***		
		(0.10)		
Union	1.255			
	(0.47)			
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	No	No
Number of (Quasi) Injury Sets	755	72	7	7
Number of Clusters		66	7	7
Mean of Dependent Variable (Raw)	0.253	1.292	1.286	1.000

Model: The model is a Poisson regression. Coefficients are presented as Incident Rate Ratios (IRR). The number of employees is used as an exposure term. Robust standard errors are used for OSHA, and standard errors are clustered on the mine identification number for MSHA. Significance Levels: *** .1%, ** 1%, * 5%.

Important Caveat: The sample used for the OSHA model includes some establishments which do not underreport at all. This is not true of the MSHA samples used in Models 2, 3, and 4; all mines in these samples underreport at least one injury.

Dependent Variable: The dependent variable is the number of unreported injuries in the (quasi) injury set.

Covariates: All models include the number of employees at the establishment (in hundreds), and state and year dummies. Models 1-3 include controls for industry. For MSHA, the industry controls are the canvas codes (coal, sand and gravel, stone, nonmetal, and metal). Models 1 and 4 include an indicator of union status at the audited establishment or mine. Models 2 and 4 include an indicator of whether the injuries in the quasi injury set occurred in an underground mine.

Unit of Observation: The unit of observation is the (quasi) injury set.

Sample: In Model 1, the sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009 that reported zero injuries for the audited year. In Models 2, 3, and 4, the sample includes MSHA quasi injury sets for mine-years with zero reported injuries and at least one injury reported after the year close. Model 2 includes quasi injury sets from 2001-2012. Model 3 excludes quasi injury sets that are not from underground mines and includes quasi injury sets from 2003, 2005-2006, 2008, and 2011. Model 4 excludes quasi injury sets that are not from coal mines and includes quasi injury sets from 2003-2004 and 2006-2007. For more details on sample construction and omitted observations, see footnote 70.

Table 6: Relationship Between Nonreporter Status on the Number of Unreported Injuries by Regime, Industry, and Size

OSHA (Sample Includes All Audited Firms)

Industry, Size	Number of Injury Sets in Sample	Number of Nonreporters	Percentage of Injury Sets with Multiple Unreported Injuries		Mean Number of Unreported Injuries	
			Nonreporter	Reporter	Nonreporter	Reporter
All Industries, Small	1112	386	4.15%	4.55%	0.22	0.19
All Industries, Medium	1101	229	4.80%*	9.06%	0.23*	0.40
All Industries, Large	1100	142	4.23%***	15.14%	0.39^	0.88
Transportation, Communications, etc., Medium (Non-Union)	79	20	10.00%*	0%	0.45*	0.10
Wholesale Trade, Medium (Non-Union)	51	10	10.00%*	0%	0.50**	0.07

MSHA (Sample Only Includes Firms That Reported At Least One Late Injury)

Industry, Size	Number of Quasi Injury Sets in Sample	Number of Nonreporters	Percentage of Quasi Injury Sets with Multiple Unreported Injuries		Mean Number of Unreported Injuries	
			Nonreporter	Reporter	Nonreporter	Reporter
All Industries, Small	136	68	14.71%	16.18%	1.28	1.22
All Industries, Medium	140	3	33.33%	23.36%	1.67	1.59
All Industries, Large	138	1	0%	26.28%	1.00	1.65
Nonmetal, All	23	6	66.67%**	5.88%	2.50**	1.06

∧ While the p-value for a two-tailed t-test is 0.062, the p-value for a one-tailed t-test is .031. Thus, we can reject the hypothesis that large OSHA reporters do not have more unreported injuries on average than non-reporters.

Significance: For percentage of injury sets with multiple unreported injuries, I compare the respective proportions among nonreporters and reporters using a two-tailed two-sample z-test of proportions. For mean number of unreported injuries, I compare the respective means among nonreporters and reporters using a two-tailed two-sample t-test. Significance Levels: *** .1%, ** 1%, * 5%.

Nonreporter: A nonreporter is defined as an establishment-year without any reported injuries. An establishment may be audited in multiple years, and its status as a reporter or nonreporter may change between years.

Size: Size is defined by the number of employees: "small" represents the bottom third of the distribution, "medium" represents the middle third of the distribution, and "large" represents the top third of the distribution. For OSHA, size is defined separately based on industry groupings (the data include the following seven of the ten OSHA "divisions": Agriculture, Forestry, and Fishing (SIC Major Groups 01, 02, 07); Construction (SIC Major Groups 15-17); Manufacturing (SIC Major Groups 20, 22-39); Transportation, Communications, Electric, Gas, and Sanitary Services (SIC Major Groups 41-45, 47-49); Wholesale Trade (SIC Major Groups 50-51); Retail Trade (SIC Major Groups 52-54, 59); and Services (SIC Major Groups 76, 80, 87)). For OSHA, statistics are first calculated separately for each industry grouping and size, and then averaged across all industry groupings. For MSHA, size and statistics are calculated across all mines.

Unit of Observation: The unit of observation is the (quasi) injury set.

Sample: The OSHA sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009. The MSHA sample includes quasi injury sets for mine-years with at least one injury reported after the year close from the years 1992, 1994, 1996-1998, 2000-2012. Before 2000, the data only includes one quasi injury set for each year. For more details on sample construction and omitted observations, see footnote 70.

Appendix

Table A: Effect of Year of Injury on the Number of Unreported Injuries, by Type of Injury (Among All OSHA Establishments)

	1. Total Num. Unreported Inj.	2. Num. Unreported Severe Inj.	3. Num. Unreported Hard-to-Attribute Inj.
1996 Dummy	0.457**	0.180	0.602
	(0.11)	(0.18)	(0.17)
1997 Dummy	0.625	1.107	0.742
	(0.16)	(0.61)	(0.22)
1998 Dummy	0.494**	0.160*	0.523*
	(0.12)	(0.14)	(0.16)
1999 Dummy	0.516**	0.111*	0.565
	(0.13)	(0.12)	(0.17)
2000 Dummy	0.265***	0.181	0.385**
	(0.07)	(0.16)	(0.12)
2001 Dummy	0.270***	0.152	0.309***
	(0.07)	(0.18)	(0.09)
2002 Dummy	0.263***	0.200**	0.251***
	(0.07)	(0.12)	(0.08)
2003 Dummy	0.298***	0.292*	0.327***
	(0.07)	(0.17)	(0.10)
2004 Dummy	0.248***	0.513	0.202***
	(0.07)	(0.35)	(0.07)
2005 Dummy	0.257***	0.189	0.218***
	(0.07)	(0.16)	(0.08)
2006 Dummy	0.134***	0.131	0.210***
	(0.04)	(0.16)	(0.07)
2008 Dummy	0.831	1.229	0.942
	(0.20)	(0.62)	(0.25)
2009 Dummy	0.468**	0.480	0.451**
	(0.12)	(0.31)	(0.14)
Employment	0.896***	0.907**	0.904***
	(0.01)	(0.03)	(0.01)
Union	1.451**	0.566	1.545**
	(0.17)	(0.19)	(0.22)
State and Industry Dummies	Yes	Yes	Yes
Number of Injury Sets	3311	3311	3311
Mean of Dependent Variable (Raw)	0.461	0.020	0.179

Model: The model is a Negative Binomial, selected because the sample variance of the dependent variable far exceeds the sample mean. Coefficients are presented as Incident Rate Ratios (IRR). The number of employees is used as an exposure term. The year dummy for 2007 was excluded. Each year represents the year of the injury logs that were audited (not the year in which the audit took place). The first year that injury logs were audited under the National Emphasis Program (NEP) is 2007. All models use robust standard errors. Significance Levels: *** .1%, ** 1%, * 5%.

Dependent Variable: For Model 1, the dependent variable is the total number of unreported injuries in the injury set. For Model 2, the dependent variable is the number of unreported severe injuries in the injury set. For Model 3, the dependent variable is the number of unreported hard-to-attribute injuries in the injury set.

Covariates: Each model includes the number of employees at the establishment in hundreds; an indicator of union status at the audited establishment; and state, year, and industry dummies.

Unit of Observation: The unit of observation is the injury set.

Sample: The sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009. For more details on sample construction and omitted observations, see footnote 70.

References

Azaroff, Lenore S., Charles Levenstein, and David H. Wegman. "Occupational Injury and Illness Surveillance: Conceptual Filters Explain Underreporting." *American Journal of Public Health* 92 (2002): 1421-1429.

Barling, Julian, E. Kevin Kelloway, and Roderick D. Iverson. "Accidental Outcomes: Attitudinal Consequences of Workplace Injuries." *Journal of Occupational Health Psychology* 8 (2003): 74-85.

Boden, Leslie, and Alexander Ozonoff. "Capture-Recapture Estimates of Nonfatal Workplace Injuries and Illnesses." *Annals of Epidemiology* 18 (2008): 500-506.

Boyles, Salynn. "$86 Billion Spent on Back, Neck Pain." *WebMD Health News*, February 12, 2008, http://www.webmd.com/back-pain/news/20080212/86-billion-spent-on-back-neck-pain.

California Department of Human Resources. *Workers' Compensation Preview.* January 2014, http://www.calhr.ca.gov/Documents/workers-compensation-preview.pdf.

Daniels, Christine, and Peter Marlow. "Literature Review on the Reporting of Workplace Injury Trends." *Health and Safety Laboratory* 36 (2005).

Dionne, Georges, and Pierre St-Michel. "Workers' Compensation and Moral Hazard." *The Review of Economics and Statistics* 73 (May 1991): 236-244.

Dong, Xiuwen S., Alissa Fujimoto, Knut Ringen, Erich Stafford, James W. Platner, Janie L. Gittleman, and Xuanwen Wang. "Injury Underreporting Among Small Establishments in the Construction Industry." *American Journal of Industrial Medicine* 54 (2011): 339-349.

Eastern Research Group. *Final Report (Revised): Evaluation of the Accuracy and Completeness* *of Nonfatal Injury and Illness Reporting in the Mining Industry.* Lexington, MA: Eastern Research Group, June 11, 2013, http://www.dol.gov/asp/evaluation/reports/MSHA-Part50-Underreporting.pdf.

Fan, Z. Joyce, David K. Bonauto, Michael P. Foley, and Barbara A. Silverstein. "Underreporting of Work-Related Injury or Illness to Workers' Compensation: Individual and Industry Factors." *Journal of Occupational and Environmental Medicine* 48 (2006): 914-922.

Galizzi, Monica, Petra Miesmaa, Laura Punnett, Craig Slatin, and The Phase in Healthcare Research Team. "Injured Workers' Underreporting in the Health Care Industry: An Analysis Using Quantitative, Qualitative, and Observational Data." *Industrial Relations* 49 (2010): 22-43.

Gray, Wayne B., and John T. Scholz. "Does Regulatory Enforcement Work? A Panel Analysis of OSHA Enforcement." *Law and Society Review* 27 (1993): 177-214.

Greenhouse, Stephen. "Work-Related Injuries Underreported." *The New York Times.* November 16, 2009, http://www.nytimes.com/2009/11/17/us/17osha.html.

Harcourt, M., H. Lam, and S. Harcourt. "The Impact of Workers' Compensation Experience-Rating on Discriminatory Hiring Practices." *Journal of Economic Issues* 41 (2007): 681-699.

Hunting, Katherine L., and James L. Weeks. "Transport Injuries in Small Coal Mines: An Exploratory Analysis." *American Journal of Industrial Medicine* 23 (1993): 391-406.

Increasing Access to Workers' Compensation Medical Benefits for Low-Income and Immigrant Workers. NH Coalition for Occupational Safety and Health Research Report. March 31, 2008, http://www.nhcosh.org/pdfs/NHCOSH_WC_Report_3-31-08.pdf.

"Improving Construction Safety Performance." Report A-3, *The Business Roundtable* (January

1982).

Kniesner, Thomas J., and John D. Leeth. “Compensating Wage Differentials for Fatal Injury Risk in Australia, Japan, and the United States.” *Journal of Risk and Uncertainty* 4 (1991): 75-90.

Kralj, Boris. “Employer Responses to Workers’ Compensation Insurance Experience Rating.” *Industrial Relations* 49 (Winter 1994): 41-61.

La Duke, Phil. “What Every Entrepreneur Should Know about Worker Safety.” *Entrepreneur*. May 30, 2014, http://www.entrepreneur.com/article/234305.

Laffont, Jean-Jacques and David Martimort. *The Theory of Incentives: The Principal-Agent Model.* Princeton, NJ: Princeton University Press, 2002.

Lewis-Beck, S., and John R. Alford. “Can Government Regulate Safety? The Coal Mine Example.” *The American Political Science Review* 74 (1980): 745-756.

Lubeck, Deborah P. “The Costs of Musculoskeletal Disease: Health Needs Assessment and Health Economics.” *Best Practice & Research Clinical Rheumatology* 17 (2003): 529–539.

Machlup, Fritz. “Theories of the Firm: Marginalist, Behavioral, Managerial.” *The American Economic Review* 57 (1967): 1-33.

McCurdy, S. A., M. B. Schenker, and S. J. Samuels. “Reporting of Occupational Injury and Illness in the Semiconductor Manufacturing Industry.” *American Journal of Public Health* 81 (1991): 85-89.

Melfi, Patrick. “New OSHA Initiative Targets Underreporting of Workplace Injuries.” *New York*

Labor and Employment Law Report. February 22, 2010, http://www.nylaborandemploymentlawreport.com/2010/02/articles/occupational-safety-and-health/new-osha-initiative-targets-underreporting-of-workplace-injuries.

Morantz, Alison. "Final Project Report: Designing a Pilot Program for Strategic Mine Safety and Health Improvements through the Use of Surveillance Data to Guide Targeted Inspection Activities." NIOSH Research Contract 200-2009-28820, September 28, 2012 (available from author upon request).

Morantz, Alison. "Coal Mine Safety: Do Unions Make a Difference?" *Industrial and Labor Relations Review* 66 (2013): 88-116.

Morantz, Alison and Alexandre Mas. "Does Post-Accident Drug Testing Reduce Injuries? Evidence from a Large Retail Chain." *American Law and Economics Review* 10 (2008): 246-302.

National Research Council. *Toward Safer Underground Coal Mines.* Washington, D.C.: National Academy Press, 1982.

Oleinick, A., J. V. Gluck, and K. E. Guire. "Establishment Size and Risk of Occupational Injury." *American Journal of Industrial Medicine* 28 (1995): 1-21.

Parker, David L., William R. Carl, L. Ronald French, and Frank B. Martin. "Characteristics of Adolescent Work Injuries Reported to the Minnesota Department of Labor and Industry." *American Journal of Public Health* 84 (1994): 606-611.

Paul, Leigh J., James P. Marcin, and Ted R. Miller. "An Estimate of the US Government's Undercount of Nonfatal Occupational Injuries." *Journal of Occupational and Environmental Medicine* 46 (2004): 10-18.

Polich, Steven. "Do Not Underestimate the Importance of OSHA Incidence Rates." *Lockton*

Companies. December 2012, http://www.lockton.com/Resource_/PageResource/MKT/Polich_OSHA%20incidence%20rates_Dec%2012%20update.pdf.

Probst, Tahira M., Ty L. Brubaker, and Anthony Barsotti. "Organizational Injury Rate Underreporting: The Moderating Effect of Organizational Safety Climate." *Journal of Applied Psychology* 93 (2008): 1147-54.

Robertson, Leon S., and J. Philip Keeve. "Worker Injuries: The Effects of Workers' Compensation and OSHA Inspections." *Journal of Health Politics, Policy and Law* 8 (1983): 581-597.

Rosenman, K. D., J. C. Gardiner, J. Wang, J. Biddle, A. Hogan, M. J. Reilly, K. Roberts, and E. Welch. "Why Most Workers with Occupational Repetitive Trauma Do Not File for Workers' Compensation." *Journal of Occupational and Environmental Medicine* 42 (2000): 25–34.

Rosenman, Kenneth D., Alice Kalush, Mary Jo Reilly, Joseph C. Gardiner, Mathew Reeves, and Zhewui Luo. "How Much Work-Related Injury and Illness Is Missed By the Current National Surveillance System?" *Journal of Environmental Medicine* 48 (2006): 357-365.

Ruser, John W. "Examining Evidence on Whether BLS Undercounts Workplace Injuries and Illnesses." *Monthly Labor Review* (August 2008): 20-32.

Shabecoff, Philip. "OSHA Seeks $2.59 Million Fine for Meatpacker's Injury Reports." *The New York Times.* July 22, 1987, http://www.nytimes.com/1987/07/22/us/osha-seeks-2.59-million-fine-for-meatpacker-s-injury-reports.html.

Shapiro, Sidney A. "Occupational Safety and Health Regulation." *Encyclopedia of Law and Economics, Cheltenham, Edward Elgar* 5540 (2000): 596-625.

Siddharthan, Kris, Michael Hodgson, Deborah Rosenberg, Donna Haiduven, and Audrey Nelson. "Under-Reporting of Work-Related Musculoskeletal Disorders in the Veterans Administration." *International Journal of Health Care Quality Assurance* 19 (2006): 463-476.

Smith, Nancy, Christi Griffin, John Newquist, Kris Hoffman, Mark Briggs, and Diane Price. *OSHA's Voluntary Protection Programs (VPP) Review: Findings and Recommendations*, a Report Submitted to David Michaels, Assistant Secretary for OSHA, (November 2011): 21-22, https://www.osha.gov/dcsp/vpp/vpp_report_nov_2011_rev_7-11-12.pdf.

Stevens, Arthur W. "New Comp Law Defines 'Preexisting Conditions'." *Black Chapman Webber & Stevens Attorneys* (blog). June 8, 2010, http://www.blackchapman.com/new-comp-law-defines-preexisting-conditions.

Sulzer-Azaroff, Beth and John Austin. "Does BBS Work? Behavior-Based Safety & Injury Reduction: A Survey of the Evidence." *Professional Safety* (July 2000): 19-24.

"The Burden of Musculoskeletal Diseases in the United States." Executive Summary. *Bone and Joint Decade* (2011), http://www.boneandjointburden.org/pdfs/bmus_executive_summary_low.pdf.

U.S. Department of Labor, Occupational Health and Safety Administration. *Injury and Illness Recordkeeping National Emphasis Program (RK NEP)*. Directive Number: 10-07 (CPL 02), September 28, 2010, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_10-07.pdf.

U.S. Department of Labor, Occupational Safety and Health Administration. *Audit and Verification Program of Occupational Injury and Illness Records.* Directive Number: CPL 02-00-138, January 12, 2006, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=DIRECTIVES&p_id=3329.

U.S. Department of Labor, Occupational Safety and Health Administration. *Injury and Illness Recordkeeping National Emphasis Program (RK NEP).* Directive Number: 09-08 (CPL 02), September 30, 2009, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_09-08.pdf.

U.S. Department of Labor, Occupational Safety and Health Administration, Memorandum from David Michaels to Regional Administrators. *Revised VPP Policy Memorandum #5: Further Improvements to the Voluntary Protection Programs (VPP).* June 29, 2011, https://www.osha.gov/dcsp/vpp/policy_memo5.html.

U.S. Department of Labor, Occupational Safety and Health Administration, Memorandum from Richard E. Fairfax, OSHA Deputy Assistant Secretary, to Regional Administrators. *Employer Safety Incentive and Disincentive Policies and Practices*. March 12, 2012, https://www.osha.gov/as/opa/whistleblowermemo.html.

U.S. Department of Labor, Occupational Safety and Health Administration. *Site-Specific Targeting 2014 (SST-14).* Directive Number: 14-01 (CPL 02), March 6, 2014, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02-14-01.pdf.

U.S. Department of Labor, Occupational Safety and Health Administration. *US Department of Labor's OSHA Cites Houston Manufacturing Company for Hiding Work-Related Injuries and Illnesses; Fines Exceed $1.2 Million.* Release Number: 10-1179-DAL, September 1, 2010, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=NEWS_RELEASES&p_id=18261.

U.S. Department of Labor, Occupational Safety and Health Administration. *1998 Audit and Verification Program of Occupational Injury and Illness Records.* Directive Number: 00-1 (CPL 2), December 2, 1999, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=DIRECTIVES&p_id=2003.

U.S. Department of Labor, Office of the Inspector General (OIG). *MSHA Has Taken Steps to Detect and Deter Underreporting of Accidents and Occupational Injuries and Illnesses, But More Action Is Still Needed.* March 31, 2014, http://www.oig.dol.gov/public/reports/oa/2014/05-14-001-06-001.pdf.

Weil, David. "If OSHA Is So Bad, Why Is Compliance So Good?" *The Rand Journal of Economics* 27 (1996): 618-640.

Welch, Laura, and Katherine Hunting. "Injury Surveillance in Construction: What Is an 'Injury', Anyway?" *American Journal of Industrial Medicine* 44 (2003): 191-196.

Wiatrowski, William J. "Examining the Completeness of Occupational Injury and Illness Data: An Update on Current Research." *Monthly Labor Review* (June 2014).

Wiatrowski, William J. "Using Workplace Safety and Health Data for Injury Prevention." *Monthly Labor Review* (October 2013).

Williams, Robert J., and J. Douglas Barrett. "Corporate Philanthropy, Criminal Activity, and Firm Reputation: Is There a Link?" *Journal of Business Ethics* 26 (2000): 341-350.