EVALUATION OF ACCURACY AND COMPLETENESS OF NONFATAL INJURY AND ILLNESS REPORTING IN THE MINING INDUSTRY

FINAL REPORT (REVISED)

Prepared for:
Department of Labor

Prepared by:
Eastern Research Group, Inc.
Lexington, Massachusetts

This report was prepared for the U.S. Department of Labor (DOL), Office of the Assistant Secretary for Policy, Chief Evaluation Office, under contract number DOLF109630909. The views expressed are those of the authors and should not be attributed to DOL, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

June 11, 2013

TABLE OF CONTENTS

ES.1 OVERVIEW

Under a contract funded through the Department of Labor's Chief Evaluation Office[1] (CEO), Eastern Research Group (ERG) and its subcontractor the National Opinion Research Center at the University of Chicago (NORC), conducted a study to examine the level of accuracy and completeness of injury/illness[2] reporting in the mining industry and identify feasible improvement approaches that MSHA could implement. MSHA considers accurate data on injuries and illnesses critical to the Agency's core mission of worker protection. Such data provide the basis for understanding trends over time, using limited resources effectively, and designing improved regulatory, training, and enforcement interventions. The data reported by mine operators is used to calculate the injury and illness "severity rate."[3] Mine operators who fail to file accident and injury reports are operating mines that are, by definition, less safe than they appear. Miners working in mines that are unsafe are at higher risk of severe injuries, occupational illnesses, and fatalities.

The Statement of Work (SOW) for this project specified three priority evaluation questions to be addressed under this study:

1. To what extent is there underreporting of injuries and illnesses under Part 50? Is underreporting concentrated to specific types of mines and operators or does underreporting occur across the mining industry? Are particular types of injury more likely to be inaccurately reported?
2. How could MSHA's current Part 50 audit process be made more effective at capturing accurate injury and illness information?
3. Are there other strategies that could be implemented, in addition to the Part 50 audits, to insure more accurate and complete reporting of accidents, injuries and illnesses for the purposes of monitoring? What techniques have been used in other sectors to improve reporting of workplace injuries and illnesses?

In addition to these evaluation questions, DOL also listed set of related sub-topics that the study should address as data and resources allowed. The sub-topics are listed in Table 1 of the report. During the development of the project, a number of decisions were made on the focus of the work. First, MSHA requested that the focus of the evaluation be on estimating the extent of underreporting (i.e., priority evaluation question #1). Second, occupational illnesses were excluded from the analysis. Illnesses tend to be defined in a more subjective manner than injuries and are often reported less consistently. Finally, ERG was asked to consider OSHA's experience in injury reporting to potentially inform MSHA's approach.

To verify the accuracy of the reported injury and illness information, MSHA performs a number of activities. First, during its regular inspections of mines performed two or four times a year at each

[1] "Evaluation of Accuracy and Completeness of Nonfatal Injury and Illness Reporting in the Mining Industry," DOLF109630909.
[2] Based on discussions during development of the study work plan, DOL decided that occupational illnesses should be excluded from the analysis.
[3] The severity rate is the number of lost workdays stemming from injuries and illnesses multiplied by 200,000 and then divided by the number of hours worked at the mine.

mine,[4] MSHA will review the records submitted by mines to determine whether all forms were filed, filed on time, and whether the forms match those MSHA has on file. Second, MSHA has also conducts Part 50 audits. These audits are required when a fatality occurs, but MSHA also has the authority to conduct these audits at its discretion. The audits are designed to uncover cases of underreporting injuries and illnesses. Finally, more recently, MSHA has conducted a set of audits called Potential Pattern of Violations (PPOV) audits. These audits are performed at mines that would appear on MSHA Pattern of Violations (POV) list[5] but for a low reported injury and illness severity rate. Thus, to determine if the mine should be on the Pattern of Violations list, MSHA conducts a PPOV audit to determine the accuracy of the reported injury and illness information.

To address the three evaluation questions, ERG performed a number of analyses that are presented in the report:

(1) **Extent of underreporting**: We looked at the extent of underreporting among mines. We did this by first assessing the number of violations that MSHA finds through its activities and where (e.g., type of mine, etc.) MSHA finds those violations. As part of our review of MSHA's activities, we made a preliminary estimate of underreporting based on MSHA's findings. The more comprehensive analysis we performed, however, compared MSHA's Part 50 data (reported by mines) to workers' compensation data for Kentucky and California. It was necessary for ERG to restrict the Part 50-workers'compensation matching to two states because (a) few states were willing to provide data and (b) matching was a resource-intensive process.

(2) **Process Analyses**: Reporting of injury and illness information can be viewed as a set of interrelated processes and breakdowns in those processes can lead inaccurate data. ERG reviewed three main processes that are involved in the reporting of injury and illness data: the process employed once an injury occurs at mine to the point at which that injury is reported to MSHA, MSHA's internal data flow processes used to track reported data, and the process used by MSHA to perform audits.

(3) **Alternative Approaches**: ERG reviewed a set of alternative approaches to verifying the accuracy of self-reported injury and illness information. These included the Occupational Safety and Health Administration's (OSHA's) approach, the use of medical records, and the use of trauma registries.

ES.2 Extent of Underreporting

The primary set of results we present in the report deal with the extent of underreporting among mine operators. As noted above, we looked at both the violations found during MSHA's activities and we performed a more comprehensive analysis to link MSHA's Part 50 data to workers' compensation data in Kentucky and California.

[4] MSHA is required to inspected all underground mines at least four times per year and all surface mines at least twice per year.

[5] MSHA maintains a number of criteria for assessing whether a mine exhibits a "Pattern of Violations" and should be placed in the POV list. A key consideration is number of significant and substantial (S&S) violations that are found at a mine, however the severity rate at the mine is also considered in whether a mine should be placed on the POV list. Details on the MSHA POV program can be found at http://www.msha.gov/pov/povsinglesource.asp.

MSHA Inspection and Audits

The report provides several tables that break down MSHA's activities by mineral (coal, metal/non-metal) and mine type (facility, surface, underground) and provides various details on the numbers of underreporting violations (30 CFR 50.20 violations) that are found by different MSHA activities.[6] Highlights of the tabulations include:

- MSHA has a significant number of mines to monitor and performs a significant number of activities at these mines. In 2012, MSHA had jurisdiction over 14,000 mines. Between 2000 and July of 2012, MSHA conducted more than 700,000 non-audit activities (e.g., inspections) at the mines under its jurisdiction and more than 12,000 Part 50 Audits. More recently, from 2010 to March 2013, MSHA conducted 82 PPOV audits.
- Only 0.7 percent of MSHA's non-audit activities found underreporting violations while seven percent of Part 50 audits found underreporting violations and 45 percent of PPOV audits found underreporting violations.[7]
- At mines where violations were found, MSHA non-audit activities resulted in 1.4 violations per action, Part 50 audits resulted in 3.1 violations per audit, and PPOV audits found almost six violations per audit.
- Overall, MSHA's non-audit activities found 6,400 violations of 30 CFR 50.20 and Part 50 audits found 2,600 violations of 30 CFR 50.20 between 2000 and July 2012. PPOV audits found 221 violations of 30 CFR 50.20 between 2010 and March 2013.
- Almost all 30 CFR 50.20 violations found during MSHA's non-audit activities (97 percent) and during Part 50 audits (also 97 percent) were associated with no lost day injuries. In contrast, 76 percent of those found during PPOV audits were for lost day injuries.

A final piece of the tabulations compared the numbers of 30 CFR 50.20(a)[8] violations found from 2006 to 2011 during MSHA non-audit activities, Part 50 audits, and PPOV audits to the number of reported injuries and illnesses for each year over that time frame. Comparing the number of violations to the number of reported injuries and illnesses results in a rough estimate of underreporting found by MSHA's own activities.[9] These comparisons show that MSHA found between five and nine percent

[6] In developing the tabulations, ERG used violations of 30 CFR 50.20 to represent underreporting violations. Violations of 30 CFR 50.20 represent underreporting of injuries, illnesses, or accidents. We did not include tabulations of timely reporting requirements (30 CFR 50.10), violations of employment or production reporting (30 CFR 50.30), or recordkeeping requirements (30 CFR 50.40). However, it should be noted that violations of the 30 CFR 50.10 and 50.30 are also considered underreporting by MSHA. The reason for focusing on 30 CFR 50.20 violations in our tabulations is that these violations are consistent with the Part 50 to workers' compensation matching that we perform and an initial scope decision for the project determined that ERG should focus on underreporting of injuries.

[7] As noted in the previous bullet point, the time frame for non-audit activities and Part 50 audits covers 2000 to July 2012 while the time frame for PPOV audits covers only 2010 to March 2013.

[8] Violations of 30 CFR 50.20(a) are specifically the *non-reporting* of an injury, illness, or accident.

[9] There are a number of caveats for this comparison of numbers: (1) the violations may be for years prior to the year in which they were found (a violation found in 2006 may reflect an underreported case in 2005), (2) not all violations are for injuries or illnesses, some may reflect unreported accidents, and (3) although all mines are subject to inspection each year, only a smaller subset is subject to Part 50 or PPOV audits which concentrate on finding underreporting violations.

underreporting from its own activities. That is, from its own activities, MSHA has found that between one in ten and one in five injuries have gone unreported between 2006 and 2011.

Comparison of Part 50 to State Workers' Compensation Data in Kentucky and California

ERG and its subcontractor, NORC, developed a comparative analysis of the MSHA Part 50 data to state-level workers' compensation data in Kentucky and California. For both states, the analysis involved a significant data creation stage in which both MSHA's Part 50 data and the state workers' compensation data were processed to ensure each contained comparable sets of injury cases. Data from each source from 2005 to 2009 were used in the matching process. Table ES-1 provides an overview of the results.

For Kentucky, we found that 32.5 percent of the workers' compensation cases did not find a match in the MSHA Part 50 data. We considered this an upper bound on underreporting over the time period. We calculated a lower bound by assuming that some matches were missed during the matching process (false negatives). Given the rigor of the matching process (see Section 4.1 in the report), we assumed that a 30 percent "miss rate" represent an upper bound on the amount we may have missed in the matching process. A 30 percent miss rate would imply a lower bound on underreporting of 22.8 percent. Thus, we calculated that anywhere between 23 and 33 percent of mining injury cases involving at least one day away from work in Kentucky were not reported under Part 50.

For California, we found that 45.9 percent of workers' compensation cases were not found in the MSHA Part 50 data. Using the same logic in "miss rates" as discussed for Kentucky, we calculated a lower bound for underreporting of 32.3 percent. Thus, we calculated that anywhere between 32 and 46 percent of mining injury cases involving at least one day away from work in California were not reported under Part 50.

Table ES-1 – Results for Comparing MSHA Part 50 Data to Kentucky and California Workers' Compensation Data for 2005 – 2009

Category	Kentucky	California
Number of Part 50 cases within the state used for matching	6,365 [a]	2,026 [a]
Number of workers' compensation cases used for matching	6,239 [b]	3,479 [c]
Percentage of state workers' compensation cases that were not found in the MSHA Part 50 data (upper bound on underreporting)	32.5%	45.9%
Lower bound on underreporting [d]	22.8%	32.2%

[a] These are injury cases where the worker lost more than one day of work.
[a] These are injury cases where the worker lost more than one day of work.
[b] These are MSHA-defined injury cases from the Kentucky workers' compensation data for mining operations where the injury occurred in the state of Kentucky.
[c] These are MSHA-defined injury cases from the California workers' compensation data for mining operations where the injury occurred in the state of California.
[d] This percentage assumes that some true matches were missed in the matching process. In the report we assumed that a 30 percent "miss rate" was a reasonable upper bound given the rigor of the matching process.

In addition to the summary results in Table ES-1, we also broke the matching results down by various characteristics of the cases, including number of days lost, cause, body part, union status, worker age, managerial status, and over time (year). We summarize the main results from these tabulations in Table ES-2 below.

Table ES-2 – Detailed Results for Part 50-Workers's Compensation Comparative Analysis for Kentucky and California, 2005-2009 [a]

Category	Kentucky	California
Days lost	• There is some evidence that more severe cases (more lost days) are more likely to be reported: 73 percent of workers' compensation (WC) cases with five or more lost days matched to MSHA data while 57 percent of WC cases with less than five days matched to MSHA data.	• No information was available.
Cause	• The cases with the three least-common causes of injury, "Burn or scald", "Cut, puncture, or scrape," and "Other cause of injury," along with "Motor vehicle," which has a slightly higher frequency, are also the least likely to find a match in the MSHA data.	• The cases least likely to find a match in the MSHA data were: "Burn or scald," the motor vehicle category, and incidents involving a fall, trip, or slip.
Body Part	• The lowest match rate (61 percent) was observed for injuries in the "Other/unknown" category. The second lowest match rate (67 percent) was observed for injuries of the trunk & internal organs. This category is the one with the greatest frequency of occurrence, so it accounts for the largest number of unreported cases during this 5-year period.	• The cases least likely to find a match observed for "Multiple and unclassified" injuries (41 percent). The second lowest match rate (48 percent) was observed for injuries of the Head, Neck, and Spinal Cord. The next two lowest match rates corresponded to the categories for upper and lower extremities.
Union	• The match rate was higher in the union mines (81 percent) than in non-union mines (77 percent), though this observation must be interpreted cautiously because (1) we were not able to determine union status for all cases, (2) the relatively small number of union mines in the sample and (3) Kentucky WC data did not permit us to resolve an injured worker's employment to the level of the individual mine, but only to the level of the mining company.	• No information was available.
Age	• There was no discernible trend across age groups.	• There was no discernible trend across age groups.
Managerial/Non-Managerial Status	• The match rate for non-managers was slightly lower than that for managers (67 percent compared to 71 percent). Nevertheless, these results should be interpreted with caution because of the small number of manager cases involved	• No information was available.
Over Time	• The results show little improvement in match rates over time, though the rates do increase steadily from 64 percent in 2005 to 70 percent in 2009.	• The results show little change in match rates over time with a low of 53 percent in 2009 and a high of 58 percent in 2008 and most rates in the 53 to 55 percent range.

[a] The main body of the reports presents the results in terms of "non-match" rates. For ease of discussion we have presented this discussion in terms of match rates.

ES.3 PROCESS ANALYSIS

The second set of results we developed involved three processes that comprise the overall the data-reporting process under Part 50: (1) the reporting process by mine operators, (2) MSHA's data processing steps, and (3) MSHA's audit processes to ensure the accuracy of data. The main body of the

report contains details on each process, including process maps for each. In this Executive Summary we focus on just the main results from each process review.

Mine operator injury reporting process

Our review of the mine operator process involved identifying sources of underreporting by operators. Naturally, it is the responsibility of the operators to ensure that injury and illness data are reported accurately and in a timely fashion to MSHA. Failure to do so constitutes a violation of the law. There are many steps in the process that can result in violations if operators do not follow the requirements of 30 CFR 50. There are three sources of underreporting that we highlighted in the report:

- *Employee fear of reprisal* – A miner may be injured and it may be a reportable injury under Part 50, but the miner may not report the injury to the operator due to a fear of reprisal.
- *Incentive and disincentive programs* – Some operators offer incentive programs that discourage injury and illness reporting. For example, an operator may offer a bonus to its miners for having no injuries over a certain time frame. Thus, if a miner reports an injury, then he and other miners at his mine may lose bonus money. This type of program could discourage reporting of injuries.
- *Gray areas related to first aid and medical treatment* – A final area that may contribute to underreporting has to do with "gray" areas in what types of injuries are reportable. Under Part 50, only injuries that require medical treatment need to be reported and injuries only require first aid do not. The text of 30 CFR 50 provides details on what must be reported and MSHA's "Yellow Jacket" report[10] provides guidance on distinguishing between the medical treatment and first aid. Thus, MSHA staff indicated to ERG that there should be little or no uncertainty as to what is and what is not medical treatment (and hence, reportable under Part 50). However, other MSHA staff commented that the nuances between first aid and medical treatment may create a perceived "gray area" for some mine operators that leads to underreporting. Nevertheless, it is the responsibility of the mine operators to know and understand the law and the text of 30 CFR 50 and to accurately report injuries.

Data flow analysis

ERG reviewed the steps taken by MSHA in processing the data that it receives from operators. Data are primarily processed by MSHA's Denver office. This review led to identifying two areas where inaccuracies could arise in the final reported data by MSHA:

- *Potential data coding ambiguities in some cases.* There may be some injuries which operators have trouble determining the coding for. In these cases, operators can request guidance from MSHA's Denver data process staff or their District Manager (DM). MSHA has protocols in place for determining how injuries are coded and these are followed by the MSHA Denver staff and by DMs. Nevertheless, it is possible for the same type of injury at two different mines to be coded differently based on differing guidance provided by DMs. ERG expects, however, this is rare and that most injuries are coded consistently.

[10] http://www.msha.gov/stats/part50/rptonpart50.pdf.

- *Closing out the data before all information on return to work is received.* MSHA closes its data for a calendar year on June 30 of the following year. For most injuries this does not pose an issue. However, for injuries where the miner has not returned to work by June 30 in the following year, MSHA truncates the days away from work measure using June 30 as a cut off. For example, if a miner was injured on December 2, 2011 and had not returned to work by June 30, 2012, then MSHA would record the worker as being away from work for 150 days.[11] Thus, for injuries where MSHA does not receive a 7000-1(d) form before June 30 of the following year, the data on return to work will not be accurate. For injuries where the worker has truly not returned by June 30 of the following year, the data underestimate return to work. For cases where the operator has neglected to send in a 7000-1(d) form (a violation Part 50 requirements) for a worker who returned before June 30 of the following year, the data will overestimate days away. Despite this, however, this only applies in cases of long-term injuries (i.e., the worker is out at least 6 months) or where the operator fails to return the 7000-1(d) form.

MSHA audit processes

ERG also reviewed the process used by MSHA, including data sources, for conducting audits with a focus on the Part 50 and PPOV audits. As noted by MSHA to ERG, the PPOV audits are more comprehensive and rely on a large number of data sources to identify violations. Data made available by MSHA to ERG indicate that Part 50 audits involved 0.28 MSHA staff hours per mine employee. PPOV audits during 2012 and 2013 involved 2.0 – 2.5 MSHA staff hours per mine employee. PPOV audits that found 30 CFR 50.20 violations found close to six violations per audit and Part 50 audits that found violations found three violations per audit.

ES.4 ALTERNATIVE STRATEGIES

ERG reviewed three alternative strategies related to ensuring that data are accurately reported. These included OSHA's approach, use of medical records, and use of trauma registries. Our findings can be summarized as follows:

- *OSHA's approach* – ERG provided details on OSHA's approach to verifying the accuracy of reported injury and illness information. Our assessment of OSHA's approaches did not reveal any specific practices that OSHA uses that would substantially improve on MSHA's approach to verifying the accuracy of injury and illness reporting. For one, at the time we performed our research, OSHA was itself undergoing a review of its procedures. Nevertheless, ERG has provided detailed information on OSHA which MSHA should consider.
- *Medical records* – We also considered whether reviewing medical records would be useful to assessing underreporting. We determined that medical records may be useful during specific audits (depending on time constraints), but would not be useful for a larger program of verifying the accuracy of reported information across several mines. One significant

[11] There are 30 weeks between December 2, 2011 and June 30, 2012. Assuming the worker works 5 days per week, then the total number of lost work days would be 30 × 5 = 150.

drawback is that medical records usually do not specify whether an injury was work related or the type of establishment the injury occurred in.

- *Trauma Registries* – ERG reviewed trauma registry information for six mining states and determined these registries would also not be useful on a large scale. As with medical records, the accuracy and reliability of coded information on work-relatedness and job type is questionable.

ES.5 Conclusions and Recommendations

Based on the work we performed under this project, we developed a set of conclusions and recommendations organized around the three priority evaluation questions.

Priority Evaluation Question #1: Extent of Underreporting

Priority Evaluation Question #1: *To what extent is there underreporting of injuries and illnesses under Part 50? Is underreporting concentrated to specific types of mines and operators or does underreporting occur across the mining industry? Are particular types of injury more likely to be inaccurately reported?*

Priority Evaluation Question #1 is primarily concerned with the extent of underreporting and aspects of and underreporting (e.g., characteristics of mines, trends, usefulness of Part 50 for accurately reporting injury rate). ERG reviewed data from MSHA's own inspection and Part 50 audit activities and performed an extensive matching exercise between Kentucky and California workers' compensation data and MSHA Part 50 data. In reviewing MSHA's inspection and Part 50 audit activities, we found that those activities alone indicate underreporting on the 5 – 9 percent range. The workers' compensation to Part 50 matching, however, shows more extensive underreporting. For Kentucky, we concluded that underreporting could be in the 23 – 33 percent range and for California in the 32 – 46 percent range.

The data we reviewed also indicate that MSHA's non-audit activities appear to find close to the same rate of violations across mine and mineral types. Part 50 audits, however, appear to find different rates among mine and mineral types. Based on the results from Part 50 audits, coal mines appear to have larger violation rates per 1,000 workers compared to metal non-metal mines. Furthermore, facilities have higher violation rates per 1,000 workers compared to either underground or surface mines and underground mines have a higher rate compared to surface mines.

There are two ways to view injury types: causes and body parts affected. For the Kentucky analysis, we found that the causes that were the least common in the data were least likely to find matches in the Part 50 data; this included burns/scalds, cuts/punctures/scrapes, and an "other" category.[12] For the California analysis we found that burns/scalds, motor vehicle-related incidents,[13] and falls/slips/trips were the most commonly underreported causes. In terms of body part affected, in Kentucky we found that trunks and internal organs, head and neck, and other/unknown injuries were the most commonly underreported. In California we found that head and neck, multiple and unclassified, and trunk and internal organ injuries were the least likely to find a match.

[12] The "other" category included such causes as Absorption Ingestion or Inhalation Not Otherwise Classified, Foreign Body in Eye, and Other Injury (Not Otherwise Classified).
[13] Motor vehicle incidents, however, may occur off the mine site and may not be subject to reporting under Part 50.

Priority Evaluation Question #2: Improving MSHA Part 50 Audit Processes

The second priority evaluation question focuses on ways in which Part 50 audits can be improved to be more effective at capturing accurate information. We developed a number of recommendations under this question. These recommendations include:

Priority Evaluation Question #2: *How could MSHA's current Part 50 Audit process be made more effective at capturing accurate injury and illness*

- ***ERG recommends that MSHA continue to implement the procedures and training it has developed and is using as part of the PPOV audits in improving the Part 50 audits.***[14] ERG also recognizes that MSHA has already implemented the spirit of this recommendation on its own.
- ***ERG recommends that MSHA perform a set of random audits to capture accurate injury and illness information from mines.*** The number of audits to be performed will depend on (1) the desired accuracy MSHA wants to obtain, (2) the categories at which MSHA wants to obtain accurate counts, and (3) the resources that MSHA can devote to random audits. ERG provided guidance on the first two while the third is clearly a decision for MSHA to make. In short, ERG recommends that MSHA stratify by canvass codes (anthracite coal, bituminous coal, sand and gravel, stone, non-metal, and metal) and select a number of audits to perform based on the statistical guidance provided in the report and the resources available to perform audits.
- ***ERG provided guidance on the factors to consider in targeting Part 50 audits.*** The data indicate that coal mines have more violations than metal non-metal mines. Among the metal non-metal mines, metal mines tend to have the most violations followed by sand and gravel, stone, and then non-metal mines. In terms of mine type, facilities are the most likely to have violations followed by underground mines and then surface mines. In fact, the analysis indicated that both facilities and underground mines have significantly more violations than surface mines. There is some (weak) evidence that mines that have completed their Part 48 training are more likely to have violations. Mines with a larger number of employees and that work more hours also have more violations. Additionally, mines where the average number of hours per employee is high also have more violations.
- ***ERG is recommending that MSHA conduct both random audits and targeted Part 50 audits***. We are not, however, recommending *how* MSHA allocate its auditing resources between the two sets of audits. Nevertheless, MSHA should make an effort to perform both types of audits. Random audits allow MSHA to develop statistical estimates of injuries and illnesses in the industry while the targeted audits allow MSHA to target mines where violations are more likely to be found.
- ***We recommend that MSHA conduct the random audits using the PPOV approach and protocols to ensure that the audits are conducted with rigor.*** ERG further recognizes that

[14] Given that MSHA has evolved its approach to audits over the course of our project work by implementing the PPOV process, ERG's role in answering this question has changed over the project. In short, it is not possible for ERG to evaluate and assess a moving target.

MSHA must also balance the random and targeted audits with the need to conduct PPOV audits, as well as within its larger statutory requirement to inspect all mines two or four times annually.

- ***ERG also recommends that MSHA perform detailed analysis of the random audits to better target future audits.*** The types of violations found and the characteristics of mines should be analyzed to determine what types of mines should be targeted in the future.

Priority Evaluation Question #3: Other Strategies

The third question asks ERG to identify other strategies that could be implemented, in addition to Part 50 audits, to insure more accurate reporting of injuries. To accomplish this, ERG is making the following recommendations:

Priority Evaluation Question #3: *Are there other strategies that could be implemented, in addition to the Part 50 Audits, to insure more accurate and complete reporting of accidents, injuries and illnesses for the purposes of monitoring? What techniques have been used*

- ***ERG recommends that MSHA institute some form of a best practices dissemination so those who conduct the audits can learn from one another.***
- ***ERG recommends that MSHA review the research that ERG has provided for OSHA to assess whether any of the aspects that OSHA performs should be adopted***.

Finally, ERG reviewed a number of potential proxies for MSHA and found that none would be more reliable than its current data. Furthermore, we feel that our recommendation to perform random Part 50 audits would result in reliable estimates of the number of injuries and illnesses in the industry. ERG reviewed trauma registries and ambulance/medical records as potential proxies. However, neither of these are available on a wide-spread basis and data from different states are likely to be. Furthermore, ERG used workers' compensation data in our analyses to assess underreporting. ERG does not feel that workers' compensation data would be useful as a proxy since (1) it took significant effort to process these data to be consistent with MSHA Part 50 reporting requirements and (2) it was a significant effort to obtain the data and only two states allowed access to their data.

1.0 INTRODUCTION

Under a contract funded through the Department of Labor's Chief Evaluation Office[15] (CEO), Eastern Research Group (ERG) and its subcontractor the National Opinion Research Center at the University of Chicago (NORC), conducted a study to examine the level of accuracy and completeness of injury/illness[16] reporting in the mining industry and identify feasible improvement approaches that MSHA could implement.

The Statement of Work (SOW) underscored the impetus for the study. MSHA considers accurate data on injury and illness critical to the Agency's core mission of worker protection. Such data provide the basis for understanding trends over time, using limited resources effectively, and designing improved regulatory, training, and enforcement interventions. Mine operators who fail to file accident and injury reports are operating mines that are, by definition, less safe than they appear. Miners working in mines that are unsafe are at higher risk of severe injuries, occupational illnesses, and fatalities.

A body of research[17] suggests that many industries underreport occupational injuries and illnesses, although most studies are not specific to mining. In addition, some audit findings by MSHA have indicated, according to Assistant Secretary of Labor for MSHA Joseph Main, "an unsettling amount of underreporting at mines that already have troublesome compliance records."[18]

The SOW specified three priority evaluation questions as well as related sub-topics that the study should address as data and resources allowed. These are listed in the Table 1. These questions cut across the challenges MSHA faces in monitoring mine operator and contractor compliance with the submission of data for reportable accidents, injuries, and illnesses and in using these data to target limited resources and improve mine safety.

To address these questions, this report provides the results of the research we performed under this project. Section 2 begins by providing background for the study design (scope, definitions, and key terms) and also provides background related to Part 50 reporting. Section 3 then addresses the types of activities that MSHA performs (inspections, Part 50 audits, and Potential Pattern of Violations audits) to uncover underreporting and the numbers of violations found through those activities. Section 4 provides a detailed summary of the largest analysis we performed under this project: a case-level comparison of MSHA Part 50 data to workers' compensation data from Kentucky and California. The purpose of this analysis is to compare Part 50 data to an external source of information on worker injuries to determine the extent to which there is underreporting in the Part 50 data. Section 5 then turns to a process analysis we performed for the three interrelated processes: mine operator reporting, MSHA data flow, and MSHA audits. The process analysis is intended to identify areas where underreporting can occur and where MSHA's processes can be improved to lead to more accurate reporting. Section 6 reviews OSHA's approach to detecting underreporting and reviews two sources of data that we were asked to review as potential sources to uncover underreporting: medical records and trauma registries.

[15] "Evaluation of Accuracy and Completeness of Nonfatal Injury and Illness Reporting in the Mining Industry," DOLF109630909.
[16] Based on discussions during development of the study work plan, DOL decided that occupational illnesses should be excluded from the analysis (see Part I, Section 2.3 on study scope decisions).
[17] See literature review summary in Appendix A.
[18] U.S. DOL News Release. *MSHA audits uncover injury, illness underreporting at Kentucky mine*. October 13, 2011.

Table 1 - Priority Evaluation Questions and Related Sub-Questions

Priority Evaluation Question	Related Sub-Topics
1. To what extent is there underreporting of injuries and illnesses under Part 50? Is underreporting concentrated to specific types of mines and operators or does underreporting occur across the mining industry? Are particular types of injury more likely to be inaccurately reported?	• Characterizing mines and operations where underreporting occurs, type of injuries more likely to be underreported, and implications for targeting. • Determining consistency of size of underreporting over time, and for subgroups. • Assessing data consistency among state worker compensation agencies, hospitals, and others with reported Part 50 data. • Considering feasibility for MSHA to accurately estimate injury rates based on the level reported under Part 50.
2. How could MSHA's current Part 50 audit process be made more effective at capturing accurate injury and illness information?	• Assessing possible correlations between accuracy in reporting injury data and other measures of mine safety violations. • Suggestions for identifying mines at high risk of underreporting for audits, • Suggested activities to enhance the quality and comprehensiveness of the audits. • Determining if current methods for analyzing Part 50 data are sufficient to identify areas of inaccuracies. • Identifying incentives or disincentives involved in the reporting process.
3. Are there other strategies that could be implemented, in addition to the Part 50 audits, to insure more accurate and complete reporting of accidents, injuries and illnesses for the purposes of monitoring? What techniques have been used in other sectors to improve reporting of workplace injuries and illnesses?	• Determining proxies for injury and illness data which could be used to track injury and illness trends over time.

2.0 STUDY AND CONCEPTUAL BACKGROUND

2.1 Definitions and Key Terms

In order to operationalize the study scope, key terms in the three priority evaluation questions (PEQ) required definition. Table 2 presents these definitions that helped shape the study scope and methodology.

Table 2 – Definition of Key Terms for Priority Evaluation Questions

Priority Evaluation Question 1 To what extent is there ***underreporting*** of ***injuries*** and ***illnesses*** under ***Part 50***? Is underreporting concentrated to specific types of mines and operators or does underreporting occur across the mining industry? Are particular types of injury more likely to be inaccurately reported?	
Key Term	**Study Definition**
Part 50	Title 30 Code of Federal Regulations (CFR) § 50: Notification, Investigation, Reports and Records of Accidents, Injuries, Illnesses, Employment, and Coal Production in Mines. • Defines accident, injury, illness, employment, and production reporting requirements for mine operators and contractors. • Requires operators and contractors to file two reports*: Mine Accident, Injury, and Illness Report (Form 7000-1) and Quarterly Mine Employment and Coal Production report (Form 7000-2). • Requires operators to maintain copies of the reports at nearest mine offices. **Note*: These reports constitute the source documents of what is generally referred to as Part 50 data, which MSHA processes and maintains as part of the MSHA Standardized Information System (MSIS). (See Section 2.2 for more description.)

Injuries [reportable]	Any injury to a miner which occurs at a mine for which ***medical treatment**** is administered, or which results in death or loss of consciousness, inability to perform all job duties on any day after an injury, temporary assignment to other duties, or transfer to another job. (30 CFR § 50.) Additional clarification: Injury is a result of a single, instantaneous incident. (MSHA report PC-7014, "Yellow Jacket"[19].) **Note:* Part 50, categorizing treatment as first aid vs. medical treatment is based on: (1) the severity of injury or procedure following the injury and (2) the number of treatments received by the injured employee. Minor injuries such as first-degree burns, bruises, and cuts may receive first aid treatment, while injuries such as third –degree burn s or permanent loss of a body part receive medical treatment. Part 50 defines first aid (not reportable) as a "one-time treatment, and any follow-up visit for observational purposes, of a minor injury"…For example, first aid can include cleaning or soaking an abrasion or wound, but an injury requiring multiple soakings is classified as medical treatment. Classification of diagnostic or preventive procedures as first aid or medical treatment depends on the results or findings of the test or procedure.
Illnesses [reportable]	An illness or disease of a miner which may have resulted from work at a mine or which an award of compensation is made. (30 CFR § 50.) Black lung is an example of an occupational illness as defined by Part 50. Additional clarification: Illness is a result of prolonged or repetitive exposure. . (MSHA report PC-7014, "Yellow Jacket".) *Note:* Excluded from analysis. See scope decisions in Section 2.2.
Extent of Underreporting	For the purposes of this study*: Extent to which Part 50 data are not accurate or complete; i.e., the extent to which numbers (or rates) of injuries and illnesses found in the Part 50 data [reported to MSHA] are not comparable to numbers (or rates) of injuries and illnesses found in comparison data sources. As such, this definition reflects inconsistency in injury/illness counts between Part 50 and comparison data sources. **Note*: ERG and DOL agreed that ideally, the definition should reflect "the extent to which events that are reportable under 30 CFR Part 50 are actually reported". In other words, the "accuracy" of a value is close to a "true value". Because no "true" value for number of injuries and illnesses at mining operations is available, accuracy and completeness was defined as above.

[19] A technical review of 30 CFR Part 50 that provides practical guidance developed by MSHA and updated to include frequently asked questions and answers pertaining to implementation of the regulation.

Priority Evaluation Question 2 How could MSHA's ***current Part 50 audit process*** be made more effective at capturing accurate injury and illness information?	
Key Term	**Study Definition**
Part 50 Audit	For the purposes of this study: Procedure by which MSHA mine inspectors/auditors compare 7001-1 and 7000-2 data reported to MSHA over a certain time period on a mine's injuries, illnesses, accidents and employees hours worked with information they compile and analyze from other relevant data sources. Data sources include internal mine documents (e.g., payroll records, time sheets); external documents (e.g., medical records, emergency medical transportation records, workers' compensation filings), and employee interviews, if conducted. (Composite audit definition based on MSHA Audit Training slides and interviews with MSHA auditors and data processing staff.) To ensure compliance with the regulation, MSHA conducts Part 50 audits when a fatality occurs at a mine, randomly (as directed at the District level), and/or when the mine is considered to have a "potential pattern of violations (PPOV)"*. *Note: PPOV refers to a review procedure to determine if a mine meets a set of specified criteria that can trigger MSHA's notification of the mine and subsequent monitoring of the mine's corrective actions. PPOV mines meet all of the criteria for receiving a Pattern of Violations (POV) notice, except for the injury severity measure(SM): lost workday per 200,000 employee hours for the mine that is greater than the overall industry SM for all mines in the same mine type and classification over the most recent 12 months.
Part 50 Audit [and Related Data] Processes	For the purposes of this study: Processes related to Part 50 audits include both conducting the audits and MSHA's data flow/processing of the Part 50 reports.
Priority Evaluation Question 3 Are there other strategies that could be implemented, in addition to the Part 50 audits, to insure more **accurate and complete reporting** of accidents, injuries and illnesses for the purposes of monitoring? What techniques have been used in other sectors to improve reporting of workplace injuries and illnesses?	
Key Term	**Study Definition**
Accurate and Complete Reporting	See PEQ-1 Definition of Underreporting.

2.2 Study Scope Decisions

Based on discussions during the initial stages in developing the study Work Plan, DOL provided some direction on operationalizing the study evaluation questions and scope. These decisions impacted the methodology. In particular, several decisions on analytical priorities, approaches, and data sources provided a framework in which the ERG team developed and executed the final methodology. Specific study scope decisions included:

Focus of the analytical effort. MSHA requested that the focus of the evaluation be on estimating the extent of underreporting (i.e., priority evaluation question #1).

Type of events analyzed. Occupational illnesses would be excluded from the analysis. They tend to be defined in a more subjective manner than injuries and are often reported less consistently MSHA Part 50.

Approaches to monitoring worker injury/illness reporting by entities other than MSHA. Analysis of OSHA's experience would be a useful component.

In addition to the scope decisions that were made, ERG and DOL ruled out a number of other approaches and data collection. We present the list of these here to provide perspective on the methods that were implemented. Each method that was ruled out was ruled out based on the infeasibility of method itself or the infeasibility of performing the component within the study time frame and budget. The methods ruled out were:

- *A statistical survey of miners*. ERG and DOL determined that this could not be conducted within the study timeframe and resources. Furthermore, no reliable list of miners exists and thus a survey of miners is not feasible.
- *Random audits*. ERG and DOL discussed the possibility of MSHA performing random Part 50 audits to better characterize underreporting. However, this was ruled since it did not appear to be feasible within the study time frame.
- *Using medical records as a source of comparison to Part 50 data*. An initial assessment of medical records found that these records would not be available in formats that could be matched to the Part 50 data.
- *Transcripts for Hazardous Condition Complaints (HCC) Help Line*. Based on initial discussions with during the kickoff meeting, ERG requested the HCC transcript. However, after further discussions, MSHA indicated to ERG that it receives few calls related to underreporting and thus the HCC data would not be useful for analysis.
- *Interviews with Sentinel of Safety awardees*. ERG also proposed interviewing Sentinel of Safety awardees for the operator perspective on injury/illness reporting. ERG was not able to conduct these interviewees since MSHA had terminated its sponsorship of the program during the course of this project.

2.3 Part 50 Reporting

Injury and Illness Information

When an accident or injury occurs at the mine, the operator is required by MSHA to complete the Mine Accident, Injury, and Illness Report (Form 7000-1). A copy of the 7000-1 form is to be sent to the MSHA office in Denver, while a copy of the form is also maintained by the operator. The information captured by the 7000-1 Form includes:

- Mine and operator identification
- Incident-specific information (e.g., cause/description of the event, equipment involved)

- Injury-specific information (e.g., part of body injured or type of illness)
- Injured employee information (e.g., job experience)
- Return to work information[20]

Quarterly Employment and Production Data

Within 15 days of the end of each calendar quarter, operators are required to submit the production totals (e.g., tons of coal produced) and number of employee hours worked during that quarter to MSHA on the Quarterly Employment and Production Form (Form 7000-2).

Uses of Part 50 Data

Combining the information received from the 7000-2 and 7000-1 forms, MSHA calculates the injury severity rate for mines as presented below.

$$Severity\ rate = \frac{\text{Number of lost workdays } \times 200{,}000}{\text{Number of employee hours}}$$

The injury severity rate is one factor used by MSHA in determining whether a mine may exhibit a Pattern of Violations.[21] Part 50 data are used in calculating both the numerator and denominator of the severity rate; thus accurate data reported from operators is essential. Since 2010, MSHA has conducted a number of audits known as Potential Pattern of Violations (PPOV) or "but for" audits. The PPOV audits are conducted among operators that would be on the Pattern of Violations list, *but for* the severity rate. The PPOV audits are designed to determine whether injuries and lost days (numerator) or hours (denominator) are accurate for these operators.

MSHA also summarizes Part 50 data on an annual basis into the Mining Industry Accident, Injuries, Employment, and Production Statistics and Reports.[22] These reports summarize the data by mine type, work location, accident classification, part of body injured, nature of injury, and occupation.

2.4 MSHA's Approaches to Onsite Injury Records Review

MSHA has two primary methods for reviewing Part 50 information at the mine level for accuracy and completeness: (1) compliance checks conducted during routine mine inspections and (2) audits.

[20] This section of the 7000-1 Form is completed upon the employee's return to work. Although an initial version of the 7000-1 Form is sent to MSHA when the injury or illness is first reported, the injury illness report (or case) is not considered to be complete until the employee has been returned to full duty (or transferred or terminated) and a completed version of the 7000-1 Form is sent to MSHA.
[21] http://www.msha.gov/POV/POVsinglesource.asp.
[22] Mining Industry Accident, Injuries, Employment, and Production Statistics and Reports http://www.msha.gov/ACCINJ/accinj.htm

2.4.1 Part 50 Compliance Checks

During regular mine inspections, which occur two to four times per year (two for surface mines, four for underground operations), MSHA inspectors review Part 50 forms (7000-1s and 7000-2s) for internal consistency with MSHA records. According to MSHA audit interviewees, these records checks are primarily concerned with submittal and timeliness information. If an inspector finds that an operator has failed to submit the appropriate paperwork, or fails to do so within the required time period under Part 50, MSHA issues a citation to the operator.

2.4.2 MSHA Part 50 Audits

Under 30 CFR Part 50, operators are required to be audited by MSHA when fatalities occur, and they were previously required to be audited when seeking candidacy for a Sentinels of Safety Award.[23] MSHA also has the discretion to perform audits when it deems them necessary. This includes:

- *Regular Part 50 Audits.* MSHA districts have the discretion to conduct regular, random Part 50 audits on a portion of their operations each year. The audits might be the result of District Manager discretion or the result of miner complaints found to have merit.
- *Potential Pattern of Violations (PPOV)/"But for" Audits.* In 2010, MSHA began performing audits of operations on the potential pattern of violations list (PPOV audits). PPOV audits are triggered when a mine meets all of the criteria of receiving a Pattern of Violations (POV) notice, with the exception of the injury severity measure.[24]

[23] Sentinels of Safety Awards are no longer being sponsored by MSHA.
[24] Injury severity measure is derived from the number of lost workdays per 200,000 employee hours.

3.0 EXTENT OF UNDERREPORTING: MSHA INSPECTIONS AND AUDIT FINDINGS

This section provides summary data on the numbers of activities MSHA conducts to identify underreporting of injuries and illnesses and the numbers of violations found. The section also provides data on the total numbers of inspections and mines that MSHA is responsible for regulating to provide context for its activities to uncover underreporting.

We look at three types of MSHA activities in this section:

- Regular inspections and other non-audit activities – These are regular MSHA inspections (E01) conducted by MSHA, which include a review of injury and illness reporting by mines, as well as all other non-audit activities conducted by MSHA. Thus, these activities can result in citations related to underreporting of injuries and illnesses.
- Part 50 Audits – These are actions denoted as "Part 50 audits" in the MSHA MSIS data. The focus of Part 50 audits is to assess the accuracy of a mine's injury and illness reporting.
- Potential Pattern of Violation (PPOV) audits – These are audits conducted by MSHA as part of its PPOV program. In brief, MSHA developed a set of criteria for identifying whether a mine exhibited a "potential pattern of violations." Some mines met the PPOV criteria *but for* a low severity rate.[25] Thus, the PPOV audits were targeted at these mines to determine whether their reporting of injuries and illnesses and production hours were accurate. These audits were also referred to as the "but for" audits.

In terms of violations, this section focuses on violations related to underreporting of injuries, illnesses, or accidents; these are classified as violations of 30 CFR 50.20. This does not include violations of timely reporting requirements (30 CFR 50.10), violations of employment or production reporting (30 CFR 50.30), or recordkeeping requirements (30 CFR 50.40). The reason for focusing on violations of 30 CFR 50.20 is that the comparison of workers' compensation to MSHA Part 50 data in Section 4.0 looks at underreporting of injury cases. Furthermore, an initial scope decision on this project was to focus on underreporting of injuries and not on the other aspects of Part 50. However, it should be noted that violations of the 30 CFR 50.10 and 50.30 are also considered underreporting by MSHA. Additionally, one focus of the PPOV audits was to find violations of 50.30 since those violations will affect the calculation of severity rates.

Table 3 summarizes the number of mines, numbers of MSHA actions, Part 50 audits and PPOV audits that MSHA has performed since 2000.[26] This table shows the scale of the work performed by MSHA. Specifically, between 2000 and late July 2012, MSHA conducted more than 700,000 activities for approximately 14,000 mines.[27] This reflects MSHA's statutory requirement to inspect all underground

[25] Severity rate is calculated as the number of lost workdays multiplied by 200,000 divided by total work hours.
[26] In this section we summarize data on violations from inspections, Part 50 Audits, and PPOV Audits. For inspections and Part 50 Audits, we use 2000-July 23, 2012 as the time period which corresponds to the MSIS data we used for this analysis. For PPOV Audits, MSHA provided data running from 2010 (inception of the PPOV audits) through March 2013. Given that the purpose of these tables is to provide a sense of where and through what mechanisms MSHA finds underreporting violations, we did not make an attempt to reconcile these time periods.
[27] As noted in the table, the number of mines in the table reflects the number in calendar year 2012.

mines four times a year and surface mines twice a year.[28] Over 2000 – July 2012 time period, MSHA performed 12,000 Part 50 audits. Finally, at the time these data were provided to ERG, MSHA had performed 82 PPOV audits between 2010 and March 2013 with most of those being performed at underground coal mines.

Table 3 – Numbers of Mines, Activities, Part 50 Audits, and PPOV Audits, by Mineral and Mine Type

Mine Type	Number of Mines, 2012 [a]	Number of MSHA Activities (excluding Part 50 Audits), 2000 -July 23, 2102 [b]	Number Part 50 Audits, 2000 -July 23, 2102	PPOV ("but for") Audits, 2010 - 2013 [c]
All Mines	14,058	708, 107	12,116	82
Coal	1,865	302,855	1,397	68
Metal/Non-Metal	12,193	405,252	10,719	14
Facility	869	67,213	853	5
Surface	12,359	431,802	10,158	4
Underground	830	208,897	1,105	73

[a] This is the number of mines in calendar year 2012
[b] Exclude Part 50 audits and PPOV audits.
[c] PPOV audits did not take place until 2010.

Table 4 summarizes the number of MSHA activities that find violations of 30 CFR 50.20 and the numbers of violations those activities find. These data indicate that 30 CFR 50.20 violations were found in 4,662 non-audit activities and 836 Part 50 audits between 2000 and July 2012. Combined, MSHA's activities, including Part 50 audits, led to identifying 9,076 (= 6,453 + 2,623) violations of Part 50.20. Part 50 audits that found violations of 30 CFR 50.20 found an average of 3.14 violations while other activities found violations of 30 CFR 50.20 found an average of 1.38 violations. Among the PPOV audits, 37 of the 82 audits found violations of 30 CFR 50.20. Those 37 PPOV audits found a total of 221 violations of 30 CFR 50.20 for an average of 5.97 violations per audit. It should be noted that the PPOV audits focused on assessing whether the mines' severity rates (injuries and illnesses per 200,000 hours worked at the mine) were accurate and included review of hours reported.

Table 4 – Part 50 Underreporting Violation Findings by Type of MSHA Activity

MSHA Activity	Number Performed	Number that Found a Violation of 30 CFR 50.20	Total Number of 30 CFR 50.20 Violations Found	Average Number of 30 CFR 50.20 Violations Found for Mines with Those Violations	Median Number of 30 CFR 50.20 Violations Found for Mines with Those Violations
MSHA Activities (excluding audits)	708,107	4,662	6,453	1.38	1
Part 50 Audits	12,116	836	2,623	3.14 [a]	1
PPOV Audits	82	37	221	5.97 [b]	3

[a] This average is skewed by the fact that three of the 836 found more than 50 violations; removing those three audits results in an average of 2.82.

[28] MSHA's regular inspections (i.e., its "twos and fours") are E01 inspections and are a subset of these activities. The data we used for this project indicated that MSHA performed 161,620 E01 inspections over the time period we looked at; however, we focused on all MSHA activities since many of these resulted in citations of Part 50.20 reporting requirements.

[b] As noted above, this section focuses on violations of 30 CFR 50.20 (non-reporting of injuries, illnesses, and accidents). PPOV audits were also concerned with violations of 30 CFR 50.30 (inaccurate reporting on hours) to identify situations where a mine's severity rate was reduced by an inaccurate reported number of hours.

Table 5 and Figure 1 summarize the distribution of the number of 30 CFR 50.20 violations from actions that resulted in finding violations. As can be seen, most activities (81 percent) that found a 30 CFR 50.20 violation, found one violation. The same is true for Part 50 audits, but a larger percentage of Part 50 audits found two or more violations of CFR 50.20 compared to inspections. Finally, for PPOV audits, the distribution is more evenly distributed across numbers of violations found with slightly less than one-third finding only one violation and about one-fifth finding between six and 10 violations.

Table 5 – Distribution of Numbers of 30 CFR 50.20 Violations from Actions that Found Those Violations by Type of MSHA Activity

Number of Part 50 Underreporting Violations Found	MSHA Activities, excluding Part 50 Audits		Part 50 Audits		PPOV Audits	
	Number	Percent [a]	Number	Percent [a]	Number	Percent [a]
1	3,790	81.3%	471	56.3%	11	29.7%
2	547	11.7%	140	16.8%	5	13.5%
3	149	3.2%	77	9.2%	7	18.9%
4	86	1.8%	38	4.6%	4	10.8%
5	30	0.6%	23	2.8%	1	2.7%
6 – 10	48	1.0%	48	5.7%	1	2.7%
11 – 50	11	0.2%	36	4.3%	8	21.6%
51 plus	1	< 0.1%	3	0.4%	0	0%
Total	4,662	100%	836	100%	37	100%

[a] This is the percentage of all inspections or audits that found a violation. Rounding errors may exist in totals.

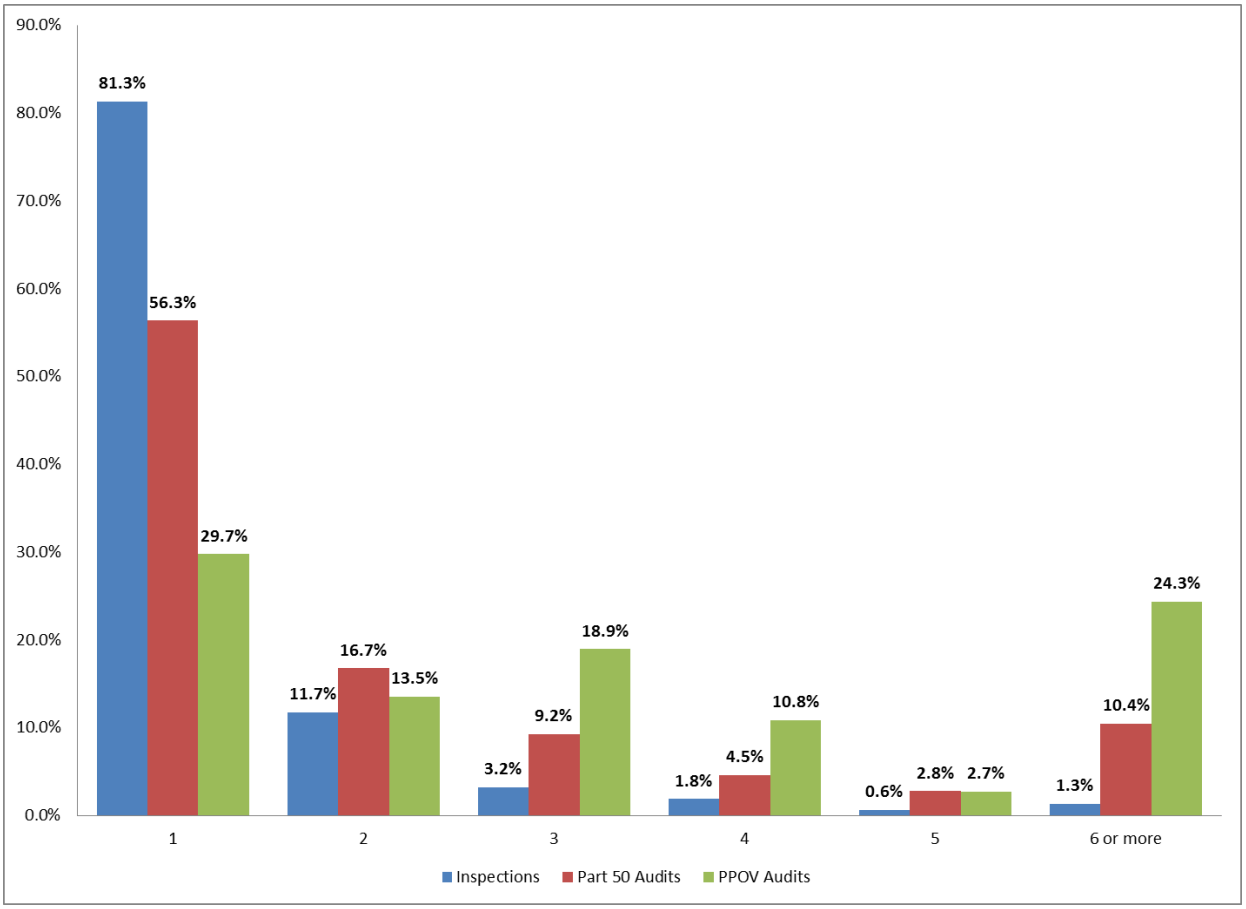


Figure 1 – Distribution of Violations of 30 CFR 50.20 for Actions that Found Violations by Type of MSHA Action

Table 6 breaks out the number of 30 CFR 50.20 violations by MSHA activity and also by coal and metal/non-metal, mine type, and characteristic of the unreported injury. The data indicate that, relative to their total number (see Table 3), coal mines are cited more frequently for 30 CFR 50.20 violations compared to metal/non-metal mines. The same is true for underground mines compared to surface mines and facilities. The data also indicate that most violations of 30 CFR 50.20 are for injuries or illnesses with no lost work days. The distribution of 50.20 violations for the PPOV audits for mineral type and mine type reflect that most of these audits were conducted at underground coal mines. For the PPOV audits, MSHA provided ERG with a special tabulation of the types of underreported violations. The PPOV audits conducted between 2010 and 2012 found 168 cases that involved lost work days among a total of 224 violations.[29] ERG also tabulated the types of violations by CFR paragraph for PPOV audits which is found in Table 7. As noted above, the PPOV audits focused on assessing whether the severity rate was accurate for mines that would have been on the pattern of violations list, but for the severity rate. Thus, Table 7 better reflects the full scope of the PPOV audits and their findings.

[29] As noted in the table, the special tabulation by MSHA for this table included 224 violations and not 221.

Table 6 – Numbers of 30 CFR Part 50.20 Violations by Mine Type and Type of MSHA Activity

Type of Mine	Number of 30 CFR 50.20 Violations Found During…		
	Inspections, Excluding Part 50 Audits	Part 50 Audits	PPOV Audits
All 30 CFR 50.20 Violations	6,453	2,623	221
Mineral Type			
Coal	2,913	1,083	198
Metal/Non-Metal	3,540	1,540	23
Mine Type			
Facility	866	463	7
Surface	3,197	1,187	0
Underground	2,390	973	214
Characteristic of Injury [a]			
Fatal	1	0	0
Lost Days	129	78	168
No Lost Days	6,252	2,540	45 [a]
Permanent	6	3	2
Not indicated	65	2	-
Other	-	-	9

[a] For PPOV audits, the data provided to ERG from a special tabulation added to 224 violations rather than 221.
[b] Also includes illnesses.

Table 7 – Numbers of Violations by CFR Paragraph for PPOV Audits 2010-2012/2013

Type of Violation	Number of Violations Found, 2010-2012/2013
30 CFR 50.11	179
30 CFR 50.20(a)	221 [a]
30 CFR 50.20(a)	76
30 CFR 50.40(a)	10

[a] These are the 221 that are summarized in the tables above.

Table 8 provides a summary of the number of 30 CFR 50.20 violations over time for non-audit activities, Part 50 audits, and PPOV audits. The series for non-audit activities shows a high of 727 in 2007 and a low of 303 in 2002; however no upward or downward trend is discernible from these data. For Part 50 audits, the high was in 2004 (333 violations) and the low in 2008 (55 violations), also with no discernible upward or downward trend. For PPOV audits, MSHA found 39 violations in 2010, 97 in 2011, and 85 in 2012/2013.

Table 8 – Underreporting Violations Over Time by Type of MSHA Activity

Year	Number of Part 50 Underreporting Violations Found During…		
	MSHA Activities, Excluding Part 50 Audits	Part 50 Audits	PPOV Audits
2000	422	257	-
2001	361	293	-
2002	303	133	-
2003	373	308	-
2004	550	333	-

Year	Number of Part 50 Underreporting Violations Found During…		
	MSHA Activities, Excluding Part 50 Audits	Part 50 Audits	PPOV Audits
2005	469	260	-
2006	628	78	-
2007	727	125	-
2008	547	55	-
2009	570	159	-
2010	561	305	39
2011	536	163	97
2012 [a]	406	154	85
Total	6,453	2,623	221

[a] For Inspections and Part 5 audits, the 2012 number reflects through 7/23/2012 and for the PPOV audits the number reflects the total number conducted in the 2012-2013 cycle.

In Table 9 we compare the number of 30 CFR 50.20(a) violations to the number of reported injuries and illnesses for 2006-2011. Violations of 50.20(a) are a subset of the violations in the table above and specifically reflect the non-reporting to MSHA of an injury, illness, or accident.[30] We compared these violations to the number of reported injuries and illnesses taken from MSHA's *Mine Injury and Worktime, Quarterly* reports for 2006 to 2012.[31] The final column of Table 9 reports the violations as a percentage of the number of reported injuries. This provides a rough estimate of underreporting from violations found during MSHA inspections and Part 50 audits sources with a few important caveats:

1. The violations may be for years prior to the year in which they were found. That is, a violation found in 2006 may reflect an underreported case in 2005.[32]
2. Not all violations are for injuries or illnesses underreporting, some may reflect unreported accidents.
3. Some of the unreported injuries and illnesses (violations) may have been added back into the number of reported injuries.[33]
4. Although all mines are subject to inspection each year, only a smaller subset is subject to Part 50 audits which concentrate on finding underreporting violations.

Nevertheless, these data provide a rough order to magnitude estimate of underreporting based on data available from MSHA activities. These data indicate underreporting of 5 to 9 percent. That is, through its activities alone, MSHA is finding that between 1 in 10 and 1 in 20 injuries go unreported. Assuming 7,500 annual injuries (a low number based on the data in Table 9), that means between 375 and 750 injuries may be unreported based on MSHA's findings from its inspections and Part 50 audits.

[30] However, most violations of 30 CFR 50.20 summarized in the preceding tables are violations of 50.20(a).
[31] http://www.msha.gov/ACCINJ/accinj.htm.
[32] This is mitigated to some degree if this cancels out over time; that is, although 2006 violations include some prior year unreported injuries and illnesses, unreported 2006 injuries and illnesses will be found as violations in 2007, etc.
[33] In Table 9 we account for this by calculating an upper and lower bound based on assumptions on whether or not the violations are added back into the number of reported injuries.

Table 9 – Numbers of Reported Injuries and Violations of 30 CFR 50.20(a)

Year	Number of Reported Injuries	Number of 30 CFR 50.20(a) Violations	30 CFR 50.20(a) Violations as a Percentage of the Number of Reported Injuries [a]
2006	10,540	600	5.4% - 5.7%
2007	10,039	751	7.0% - 7.5%
2008	9,528	507	5.1% - 5.3%
2009	7,842	584	6.9% - 7.4%
2010	7,561	670	8.1% - 8.9%
2011	7,690	552	6.7% - 7.2%

[a] The lower bound in the range is calculated by dividing the number of violations by the sum of the number of reported injuries and the number of violations while the upper bound is calculated by dividing the number of violations by just the number of reported injuries. The lower bound reflects an assumption that all violations are added back into the number of reported injuries and the upper bound assumes that none of the violations are added back.

4.0 EXTENT OF UNDERREPORTING: COMPARING MSHA PART 50 DATA TO STATE-LEVEL WORKERS' COMPENSATION DATA

This section describes our analysis to match state-level worker's compensation data to MSHA Part 50 data. The purpose of this matching is to use a non-MSHA source of data on injuries and illnesses to asses to the extent to which injuries and illnesses are not being reported to MSHA by operators. Based on discussions with MSHA, ERG decided to use workers' compensation data for this analysis. Workers' compensation laws are state specific, so the analysis must be performed on a state-by-state basis. ERG explored the possibility of performing this analysis for multiple important mining states; however, only two states allowed ERG access to its workers' compensation data for this analysis: Kentucky and California. To perform this analysis, ERG used its subcontractor, the National Opinion Research Center and the University of Chicago (NORC), to perform the matching of state workers' compensation data to MSHA Part 50 data. We describe the results for each state separately.

4.1 Kentucky

This section describes our analysis that matched Kentucky workers' compensation data to MSHA Part 50 data for mines in Kentucky. Workers' compensation records are generated in Kentucky based on a First Report of Injury (FROI). ERG requested and received FROI data from Kentucky. The FROI data we received from Kentucky reflect cases that resulted in more than one lost day. We begin by discussing the data processing steps that were taken to create comparable data sets for the analysis. Next, we describe the approach we took to matching the two data sets. Finally, we present the detailed results of the matching process.

4.1.1 Data Processing for Kentucky

In this section we describe the process used to develop two sets of comparable data for matching: one for the Kentucky workers' compensation data and one of the MSHA Part 50 data. Creating comparable data sets are necessary for effective matching.

Processing Kentucky State Workers' Compensation Data

Kentucky workers' compensation data for the years 2005-2010 (inclusive)[34] was received from the Kentucky Department of Workers' Claims in the form of Excel worksheets. The data file contained information on 204,111 cases over that time period. The data included all WC cases available (including injuries and illnesses) and had information on:

- Geographic location of injury (including city, state, zip code, and date)
- Demographics of the worker (including gender, date of birth, age, and occupation)
- Information on the injury (including body part and cause)
- Industry information (including industry name and description)

[34] Although 2010 was beyond the reference period of 2005-2009, data from 2010 was useful for identifying cases that occurred in 2009 but were not reported to KY WC until 2010. Consistent with this, cases that occurred before the reference period but were reported during the reference period were excluded from the analysis.

- Employer information (including name and location)

ERG and NORC took several steps to develop a file that could be matched to the MSHA data. These steps are described below and the resulting number of records following each step is summarized in Figure 2. Appendix B provide a summary table showing Kentucky WC cases retained or excluded at each step in the case selection process.

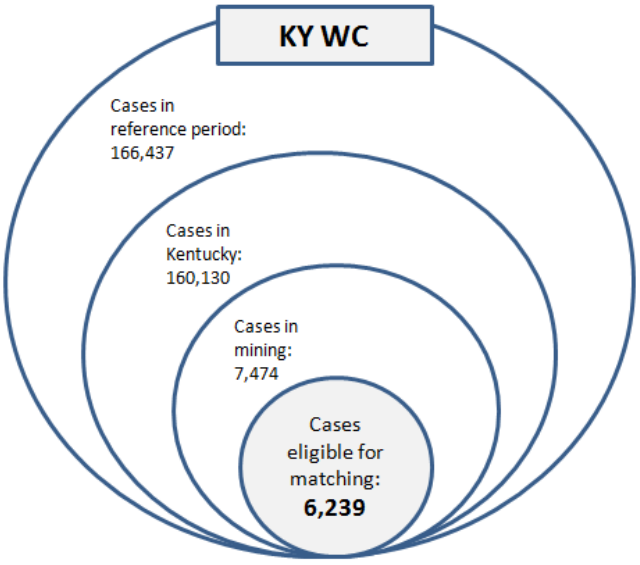


Figure 2 – Kentucky Workers' Compensation Data Processing Summary

- **Step 1 – Retain only cases within the reference period of 2005-2009.** The Kentucky WC files included an additional year beyond the MSHA dataset for a total of 204,111 cases across 2005-2010. Thus, cases from 2010 were filtered out. This resulted in 166,437 cases for the 2005-2009 time period.
- **Step 2 – Retain only cases that occurred in Kentucky.** Although the data was obtained from the state of Kentucky, it is possible for the cases to have occurred outside of the state. Workers' compensation covers the employees of businesses within the state; thus, someone who is injured while working for a Kentucky company outside of the state of Kentucky would appear in the file if a workers' compensation case was filed. The file included two relevant data elements: "Injury Zip" and "County of Injury." The majority of cases indicated a Kentucky location based on both data elements and thus were retained. Thus, 160,131 cases were identified as having occurred in Kentucky.
- **Step 3 – Retain only cases for mining-related injuries.**[35] The data provided by Kentucky covered a range of industries. Two relevant data elements were used to identify mining establishments: "Industry Name" (the superordinate category) and "Industry Description" (a more detailed classification). Specifically, the unique combined values of the data elements were identified and then each record was coded for whether it identified a mining business. Mining industries were identified by referring to the industry descriptions attached to all the sub-categories within NAICS 212. When Industry Name and Industry Description were in conflict (e.g., when Industry Name was 'animal specialty services, except veterinary' and Industry Description was 'metal mining'), whether to include the case was resolved by referring to the Occupation of the injured worker(s). Cases with an Occupation that accorded with a mining-related component (e.g., 'mining machine operators') were retained, while cases where the Occupation matched a non-mining component (e.g., 'animal caretakers except farm') were dropped.[36] This resulted retaining 7,474 cases as mining-related injuries.
- **Step 4 – Retain only cases that are reportable to MSHA as an *injury*.** MSHA defines injuries as resulting from "a work accident or exposure involving a single instantaneous

[35] A mining injury is defined as a person that was injured while working at a company classified by the Kentucky WC system as being in the mining industry.
[36] See Appendix C for the coded list of Kentucky WC industries.

incident in the work environment," while "repeated trauma or repetitive movement which produces tenosynovitis is considered an illness."[37] Also, injury to the lungs from inhaled substances is considered an illness by MSHA. In contrast, Kentucky WC considers repetitive-stress injuries and lung conditions to be injuries. Given these differences, these two conditions were excluded from the WC dataset. Thus, 6,239 cases were thus identified as eligible for matching.

Processing MSHA Part 50 Data from Kentucky Mines

Similar to the workers' compensation data, several database preparation steps were undertaken to establish a comparable set of cases in MSHA Part 50 from Kentucky. Cases were selected from the MSHA Accident/Illness (AI) files based on the steps described below. Figure 3 summarizes the results of these steps. Appendix B provides a summary table showing cases retained or excluded at each step in the case selection process.

- **Step 1 – Retain cases that match the time frame of the Kentucky worker's compensation data.** For purposes of this project, NORC used MSHA Part 50 data from 2005-2009. This corresponded to the data available from NIOSH at the time the analysis was initiated. This resulted in retaining 68,113 cases from 2005 to 2009.

- **Step 2 - Retain only injury cases.** MSHA data provides a field ("aii") that codes each incident based on the type of harm suffered by the employee. Values in this field that are 21 or under indicate the record is an injury; a value above 21 in this field indicates the record is an illness. As noted above, the focus of the evaluation was underreporting of injuries. Based on this, 66,488 cases were identified as involving an injury.

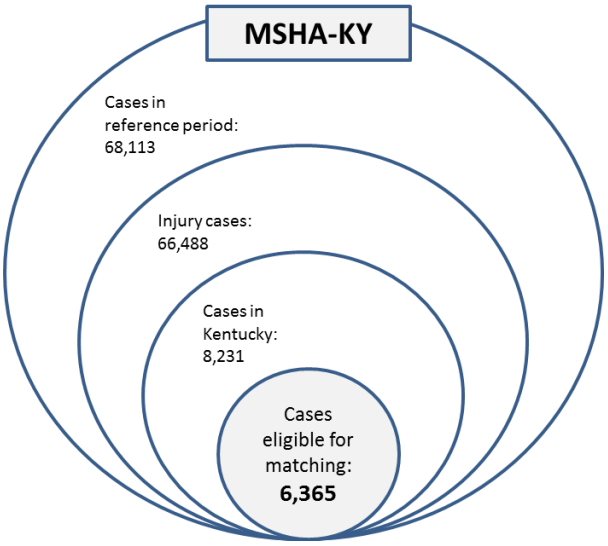


Figure 3 – Kentucky MSHA Part 50 Data Processing Summary

- **Step 3 – Retain only cases occurring in Kentucky.** MSHA's "state" field was used to identify the location of the mine where the injury occurred. This resulted in retaining 8,231 injuries as having occurred in Kentucky mines.

- **Step 4 – Retain only cases that caused the employee to lose more than one day of work time.** Kentucky workers' compensation data tracks only cases where the employee has lost one or more days of work. The variable 'deginj' classifies injuries by type and degree, including noting whether an injury was physical or not. When 'deginj' is set to zero, then no actual physical injury to the employee has occurred. Thus, we excluded cases where 'deginj'

[37] MSHA, 1986. "Report on Part 50," PC-7014. http://www.msha.gov/stats/part50/rptonpart50.pdf.

was set to zero. Based on this criterion, 6,365 Kentucky cases were thus identified as eligible for matching.

After case selection, company-related information on both the mine operator and the mining contractor (where relevant) was extracted from the MSHA Address/Employer (AE) files and merged into the AI case file so it would be available for use in future analyses once the matching process was complete.

4.1.2 Approach to Matching Kentucky Workers' Compensation Data to MSHA Part 50 Data

Determining Which Data Elements/Fields to Use in the Matching Process

ERG and NORC reviewed both data sets to determine the best approach to match the datasets. Specifically, it's necessary to determine which data elements/fields in both datasets can be used to match one data set to the other. Based on an initial examination, eight data elements/fields were identified for possible use in matching records across datasets. After further review and some coding adjustments, however, six of these were selected for matching attempts. Table 10 describes the eight data elements that were considered and the reasons for their use or exclusion on the matching process.

Table 10 – Data Elements Used in Matching WC and MSHA Data

Data Element	Used in Matching	Characterization
Employee age	✓	These data elements had either exactly the same coding in the two data sources or could be equalized with minimal recoding. Only the recoding of "body part injured" required any exercise of judgment. [a]
Employee sex	✓	
Body part(s) injured	✓	
Date of injury	✓	
County of injury	✓	
Company name	✓	For each reported case, MSHA data provide two data elements identifying the worker's employer: the name of the company that is the operator of the mine, and also the name of the contractor employing the worker (where applicable). Company name was not directly present in the WC data but was encoded indirectly in the WC data element LocationCode. So in response to a follow-on request, Kentucky WC administrators provided company-name information for about 4/5 of the cases in our WC file. For the MSHA data, preliminary investigations revealed that matching was more successful when company name was set to the name of the contractor company when one was present and to the name of the operator company otherwise. So that was the procedure followed. When company name information was received from the Kentucky WC administrators, it was in string format, as were the company name variables in the MSHA file. In order to maximize the power of the matching

Data Element	Used in Matching	Characterization
		process using company name, recoded versions of both data elements were produced in which the key one or two words were extracted from company name. [b]
Cause of injury	✘	This data element proved infeasible to recode into the same system because the codeframes used in the two data sources were too different. For cause of injury, the primary barrier to alignment is that the categories in the set of causes used by Kentucky WC are not mutually exclusive, and include both causes describing the nature of the injury (e.g., 'burn or scald contact with steam or hot fluids' or 'strain or injury by lifting') and causes describing the activity that led to the injury (e.g., 'fall slip or trip on stairs' or 'struck or injured by falling or flying object'). We were therefore unable to map the WC causes consistently onto the values of MSHA's 'natinj' variable, which codes strictly for the properties of the physical injury suffered by the worker. (In the hope that the values of 'Cause' as they were actually used might show more consistency than appeared likely given the codeframe, we selected out the subset of cases that matched uniquely across the two data sources for values of the five variables age, sex, county, injury date, and body part, and cross-tabbed this set's values for 'Cause' and 'natinj'. However, the results of this procedure confirmed our sense that there was no consistent mapping between values of these two variables.
Employee's job	✘	This data element proved infeasible to recode into the same system because the codeframes used in the two data sources were too different. Occupation is coded by the WC data element 'Occupation', with 99 categories represented in the final WC file, and the MSHA data elements 'jobtitle' and 'jobtitle2'. ('jobtitle2' is slightly less detailed than 'jobtitle', having 121 total values as opposed to 199 for 'jobtitle'.) The primary difference between the coding schemes is that MSHA's job-title variable was constructed to categorize occupations very specific to the mining industry, while WC's occupational categories cover occupations in every field. MSHA's categories are therefore highly technical and depend on detailed knowledge of the mining process. We attempted to construct a simplified occupational classification involving significant recoding of both WC and MSHA occupation variables, but were unable to interpret many of the MSHA job classifications with sufficient certainty to be confident that the recoded variables would contribute to matching success rather than detracting from it. We therefore decided not to include occupation in the list of data elements used for matching cases. It is worth noting that, in this case, there is no inherent incompatibility between the codeframes used by WC and MSHA, and future efforts to reconcile the two might succeed if carried out by persons with a very high level of expertise in the specifics of the mining industry.

[a] For example, while MSHA uses only the category 420 ('Back') to code all injuries to the back, Kentucky WC data differentiates among 'low back area (inc: lumbar and lumbo-sacral)', 'lumbar &/or sacral vertebrae', and 'upper back area (thoracic area)' as sites of injury. So for matching purposes, all three WC codes were collapsed into a single category for coding all injuries to the back. (The list of comparable categories across data sources is found in Appendix D.)

[b] For example, a company name like 'A B & J Coal Co., Inc.' would be designated 'AB&J Coal'; 'Matrix Energy LLC' would be listed as 'Matrix Energy'. This procedure was designed to simplify and standardize company name while taking a conservative approach that would avoid combining similar but distinct names into the same recoded category. For example, 'B & W Resources', 'B&W Resources Inc', and 'B&W Resources, Inc.' were all recoded into 'B&W Resources', since the differences among them involved only punctuation and incidental words. However, since 'Cheyenne Mining Co Inc' and 'Cheyenne Resources Inc' differed in substantive ways, they were maintained as separate company names in the recoded version.

Criteria Used to Determine a Match

The data elements/fields identified above tell us which characteristics of an injury case to compare, but do not provide a criterion for determining when two injury cases can be deemed a match. Specifically, it is necessary to determine the number of matched data elements/fields that constitute a match. If we require that all six data elements/field match for two cases be a match, then some potential matches will be missed. Additionally, defining a match based in the concurrence of 1-2 data elements/field would be too lenient and result in false matches.

To determine the appropriate number of data elements/fields to define a match, ERG and NORC performed an investigatory round of matching.[38] This round of matching yielded the following results:

- Using all six comparable data elements resulted in a 12.6 percent match rate. Of these, all cases but four match uniquely; that is, exactly one WC case matches exactly one MSHA case, with the remaining WC cases matching two cases in the MSHA file.
- Using any five of the comparable data elements:
  - 2,418 (38.8 percent) of the Kentucky WC cases match to the MSHA dataset.
  - Of these, 2,261 (93.5 percent) are unique matches.
- Using any four of the comparable data elements:
  - 5,233 (83.9 percent) of Kentucky WC cases match to the MSHA dataset.
  - Of these, 3,313 (63.3 percent) are unique matches.

With the relatively high rate of non-uniqueness found among the matches on 4 variables, we could not be confident in counting all these WC cases as matches. In particular, given that the WC file contains many more cases than the MSHA file, there would likely be many situations where more than one WC case shares values on at least 4 matching variables with just one MSHA case. Even if we accept that matching on 4 variables is sufficient in principle, allowing more than one WC case to match a single MSHA case would produce a significant over-estimate of the true match rate. This would not be true, however, if a single WC case shared values with more than one MSHA case. In this situation, although we might not be able to be sure which MSHA case was the correct match to the WC case, we could be reasonably confident that the WC case matched one or the other of them and that it would therefore be fair to count the WC case as having found a match in the MSHA data. Similarly, if a particular combination of values on 4 variables was shared by two WC cases and two or more MSHA cases, it

[38] This initial round of matching was done to assess the number of data elements to use as a match criterion. This initial round did not include as extensive post-matching double checking and did not account for the possibility of multiple matches per record. This is described in more detail in "Matching Limitations and Considerations" below.

would be reasonable to conclude that the MSHA data contained reports for those two cases, even if we could not be sure which particular MSHA case corresponded to each WC case.

Accordingly, *we counted cases as matching if they shared values with any four of the six data elements / variables available for matching, but only if the number of MSHA cases with those values was equal to or greater than the number of WC cases involved.*

Matching Limitations and Considerations

Regarding the level of confidence in the matching rates and other analytical results of this comparison between Kentucky WC system data and MSHA data, the following limitations/considerations should be kept in mind:

- ***Matching involved a small set of variables/data elements***. Both the Kentucky WC data and the MSHA dataset used in the study contain numerous variables describing the case records. In matching cases across the two files, however, we relied on the small set of key variables whose codeframes could be made to match exactly. We counted cases as matching when these key variables all had the same values.
- ***The existence of match duplicates***. For all sets of matching variables, there were initially at least a small number of instances where the match was not between exactly one case in the WC file and exactly one case in the MSHA file, but was one-to-many or even many-to-many. We made our best effort to resolve these duplicates with reference to other supporting information in the file, but resolution was not always possible. In these cases, the existence of a match is therefore not certain, but only probable.
- ***Uncertainty resulting from variable recoding.*** In order to harmonize the codeframes for some of the matching variables across the two files, some recoding was necessary. In the case of county or gender, the coding was straightforward; but for part of body injured and for company name, the recoding was more involved and could have resulted in matching errors. These matching errors could comprise both misses and false positives.
- ***Missing and inconsistent data.*** In some cases, data for key matching variables was missing and could have led to missed matches. For example, about 18 percent of the Kentucky WC cases were missing company name, and about 0.5 percent were missing body part.
- ***Data Errors.*** There is an unknown rate of error connected with each of the key variables used in matching that could affect the accuracy of the reported match rate. Errors in variables such as age, date of injury, county of injury, and gender would primarily arise because of simple miscoding by the WC claimant, WC staff, or MSHA reporter. For example, when a case in the WC data file is coded “Female” but Claimant First Name is “Lawrence” or “Mohamed”, it is likely that the correct gender for that case is “Male”. Further, ambiguities associated with company name may arise because there can be several corporate entities associated with a particular mine (a contractor company, an operator company, a parent company or subsidiary, a holding company, and so on) and the standards governing which company to report may differ between WC and MSHA or may not be followed by every reporter. An even more complex exercise in judgment is called for in the coding of which

body part was injured (e.g., it is not always clear whether a particular injury is to the "ankle" or the "foot").

4.1.3 Results for Matching Kentucky Workers' Compensation Data to MSHA Part 50 Data

Using the matching criterion and processed data described above, ERG and NORC matched the 6,239 Kentucky workers' compensation cases to the 6,365 cases in the MSA Part 50 data to determine the extent to which workers' compensation cases were also reported to MSHA under Part 50. Overall, ERG and NORC found that 4,209 cases (67.5 percent) of the workers' compensation cases had a match in the MSHA data. ERG and NORC also calculated that a 95 percent confidence interval around the estimate would be 66.3 percent to 68.7 percent.[39]

As noted in Section 4.1.2, there are a number of data-specific limitations to the matching process that concerns us in attributing all 2,030 unmatched cases to underreporting and using that as our only estimate of underreporting for these types of cases in Kentucky. We expect that with better data, some of the 2,030 may have found a match in the Part 50 data. Nevertheless, it seems reasonable to assume that the upper bound on underreporting in Kentucky is at most 32.5 percent,[40] but given the limitations in the matching, the upper bound is probably somewhere less than 32.5 percent.

In order to provide a range for underreporting, we assume the data limitations led to some "miss rate" (i.e., the percentage of matches that we missed due to the data limitations). A miss rate of zero assume that all non-matched workers' compensation cases truly did not have a corresponding cases in the MSHA Part 50 data and a miss rate of 100 percent assumes that everything we did not match should have found a match. The process we used to match the two set was rigorous and detailed and we expect that our miss rate should be less than 50 percent. Table 11 provides a sensitivity analysis on the underreporting estimate based on a few assumed miss rates.

Based on our best professional judgement, we expect that a 30 percent miss rate represents a reasonable upper bound. As mentioned above, ERG and NORC undertook a rigorous and detailed matching process. The process involved repeated attempts to match records based on all available information that could reasonably be used to match records from the two data sets. Additionally, the data elements we used for matching represented key data elements for injury cases (see Section 4.1.2); thus not matching on those key data elements would seem to indicate a non-match (i.e., the workers' compensation case was not found in the Part 50 data). Given these considerations, a 30 percent miss rate would seem to be a large number of misses.

[39] This range was calculated using the assumption that the categorization of cases as "matching" or "not matching" should behave approximately as a binomial event. This assumption should be robust with the available sample size and allows the estimation of a standard error (SE) term using the match percent as "p" and the non-match percent as "q," in the formula SE=SQRT(pq/n). Using the resulting SE to construct a 95 percent interval around (1-p) using the normal approximation to the binomial distribution results in the lower and upper bounds interpretable as indicating that the true match rate should be between the upper and lower bounds with 95 percent confidence.

[40] Alternatively, we could use the lower bound of the 95 percent confidence interval of the match rate (66.3 percent) to define the upper bound on underreporting. However, given the matching limitations, we expect a more prudent approach is to use the percent matched. The difference is small, however.

In summary, we expect that a reasonable range for underreporting in Kentucky for injury cases involving at least one lost day is between 23 percent and somewhere less than 32.5 percent.

Table 11 – Estimated Underreporting Percentage for Kentucky for Assumed Matching Miss Rates

Miss Rate	Number of Non-Matched Cases	Underreporting Percentage
0%	2,030	32.5%
20%	1,624	26.0%
30%	1,421	22.8%
40%	1,218	19.5%
50%	1,015	16.3%

Table 12 provides a set of cross-tabulations on the matching. We summarize these cross-tabulations below.

Lost Workday Cases

Lost workdays (LWD) counts for an injured worker were available for 29 percent of the cases in the Kentucky WC file (N=1,539). The median number of lost workdays for these cases is 5, with a range from 0 to 1,627 and a mean of 51. As shown in the Table 12 below, comparing the non-match rates for cases with a number of lost days below the median or above it reveals that cases with 0 – 5 days out of work have a non-match rate of 43 percent, while cases with more days lost have a non-match rate of 26.8 percent. Days lost from work is a reasonable proxy for severity, so this comparison provides some evidence that underreporting is more likely for injuries that are less serious.

Table 12 – Number of Non-Matched Cases and Non-Match Rates from Comparing Kentucky Workers' Compensation Data to MSHA Part 50 Data for Kentucky (2005-2009) for Cases Involving at Least One Day Away from Work, Overall and by Various Categories

Category	Number of Kentucky Workers' Compensation Cases Used in Matching Process	Number of Kentucky Workers' Compensation Cases Not Matched to MSHA Part 50 Data	Percentage of Kentucky Workers' Compensation Cases that Did Not Match to MSHA Part 50 Data
All cases	6,239	2,030	32.5%
Number of Days Lost			
0-5 days lost	786	338	43.0%
5+ days lost	753	202	26.8%
Total [a]	1,539	540	35.1%
Cause of Worker Injury			
Burn or scald	144	53	36.8%
Caught in/under/between	421	96	22.8%
Cut, puncture, scrape	245	91	37.1%
Fall, slip, or trip	1,005	360	35.8%
Motor vehicle	510	192	37.6%
Strain	1,824	609	33.4%
Striking against/stepping on something	265	80	30.2%
Struck by something	1,592	435	27.3%
Other cause of injury	233	114	48.9%
Body Area of Worker Injury			
Head & neck	530	173	32.6%
Trunk & internal organs	1,759	629	36.8%

Category	Number of Kentucky Workers' Compensation Cases Used in Matching Process	Number of Kentucky Workers' Compensation Cases Not Matched to MSHA Part 50 Data	Percentage of Kentucky Workers' Compensation Cases that Did Not Match to MSHA Part 50 Data
Arms & shoulders	1,666	469	28.2%
Legs & hips	1,532	443	29.0%
Other/unknown	752	316	39.1%
Union Status			
Status unknown	1,999	1050	52.5%
Non-union	3,960	928	23.4%
Union	280	52	18.6%
Worker Age Quartiles			
17 - 28	1,596	512	32.1%
29 – 37	1,616	479	29.6%
38 – 47	1,516	499	32.9%
48 & over	1,511	540	35.7%
Managerial/Non-Management Status			
Non-Manager	5,993	1959	32.7%
Manager	246	71	28.9%
Year			
2005	1,340	480	35.80%
2006	1,397	457	32.70%
2007	1,133	362	32.00%
2008	1,190	375	31.50%
2009	1,179	356	30.20%

[a] Only 29 percent of the cases in the Kentucky worker's compensation data had a number of lost workdays available.

Cause of Worker Injury

The categories used by Kentucky WC to code for the cause of a worker's injury, although not compatible with the scheme used by MSHA, can be combined in a reasonable way to enable us to examine the relative match rates of the different categories. As shown in Table 12, the non-match rates range from 48.9 percent for the heterogeneous category "Other cause of injury",[41] to 22.8 percent for "Caught in/under/between." The cases with the three least-common causes of injury, "Burn or scald", "Cut, puncture, or scrape," and "Other cause of injury," along with "Motor vehicle," which has a slightly higher frequency, are also the least likely to find a match in the MSHA data. Cases with the remaining five causes of injury had higher match rates. Because WC does not code for relative severity of injury, it is not possible for us to identify whether these less-common causes of injury also involved less-serious injuries.

Body Area of Worker Injury

Table 12 shows match rates by injured body area, using values of Kentucky workers' compensation data's Body Part field collapsed into four major categories. The highest non-match rate (39.1 percent) was observed for injuries in the "Other/unknown" category. This category represents 13.0% of all reported injuries, and represents injuries to multiple body parts or systems along with cases for which body part was not recorded. Given that body part was used as one of the variables in the

[41] This category comprises such causes as Absorption Ingestion or Inhalation Not Otherwise Classified, Foreign Body in Eye, and Other Injury (Not Otherwise Classified).

matching process, it is not surprising that this category yielded lower match rates. The second highest non-match rate (36.8 percent) was observed for injuries of the trunk & internal organs, indicating that 629 cases of this type of injury went unrecorded in Kentucky during the reference period. This category is the one with the greatest frequency of occurrence, so it accounts for the largest number of unreported cases during this 5-year period. The next highest non-match rate was observed for head and neck injuries, which were not matched 32.6 percent of the time. This is also the lowest-frequency category of injury, representing only 8.5 percent of all cases. Although only 173 cases of head and neck injuries were apparently unreported during this period, it is still an important concern, given the critical areas involved. Injuries of the legs and hips and of the arms and shoulders had the lowest non-match, but because they represent approximately half of all injuries, these two categories accounted for 1,098 unrecorded injuries.

Union Status

We were able to obtain union status for approximately two-thirds of the cases in the Kentucky WC file. However, because the WC data does not allow us to identify the particular mine where an individual works, we could only attribute to a WC case the most common union status for each company as found in the U.S. Energy Information Administration's annual coal production reports.[42] Along with differences in union status from company to company, different mines run by the same company can also have different union status. Thus, there is some uncertainty associated with the assignment of union status to a given injured worker's data. As might be expected, the match rate (see Table 12) was higher in the union mines than in non-union mines, though this observation must be interpreted cautiously because of the relatively small number of union mines in the sample and because the Kentucky WC data did not permit us to resolve an injured worker's employment to the level of the individual mine, but only to the level of the mining company. Although degree of unionization does differ by company, different mines run by the same company can have different union status, so there is some uncertainty associated with the assignment of union status to a given injured worker's data.

Age of Injured Worker

Table 12 shows match rates for different age groups of injured workers. Workers were divided into quartiles to assess difference among age groups (aged 17-18, aged 29-37, aged 38-47, and aged 48 and older. The non-match percentage is roughly equal across the four groups.

Gender of Injured Worker

Although not presented in Table 12, we also looked at the non-match rate by gender. Workers' compensation cases involving females were much less likely to find matches in the MSHA file: only 31.3% of female WC cases matched. However, the total numbers involved are so small (only 83 WC cases, or 1.3% or the total) that this difference cannot be relied upon.

Worker Management Status

Although Kentucky WC data does not specifically code workers for whether they are part of management or not, the workers' compensation field "Occupation" does give some indication of a

[42] http://www.eia.gov/cneaf/coal/page/databases/coalpublic05-coalpublic09

worker's status which we used to code each case as with "Manager" or "Non-manager."[43] By this accounting, the number of managers with injuries recorded in the WC data was very small, 71 cases or 3.9 percent of the total cases in the file. The non-match rate for non-managers was slightly higher than that for managers (32.7 percent compared to 28.9 percent; see Table 12). Nevertheless, these results should be interpreted with caution because of the small number of manager cases involved.

Year-to-Year Overall

Table 12 also provides a summary of the non-match rates over time. These rates are also plotted in Figure 4 along with their corresponding 95 percent confidence intervals. These results show little improvement in non-match rates over time, though the rates do decline steadily from 35.8 percent in 2005 to 30.2 percent in 2009. These data indicate that underreporting may be declining over time in Kentucky. Figure 4 provides a graphical display of these rates over time showing the 95 percent confidence interval for each year.[44]

[43] We recoded the "Occupation" values 'Management related occupations', 'Managers & administrators NEC', 'Production Inspectors checkers & examiners', 'Purchasing managers', 'Supervisors food preparation & service occupations', 'Supervisors & proprietors sales occupations', 'Supervisors construction occupations', 'Supervisors extractive occupations', 'Supervisors general office', 'Supervisors handlers equipment cleaners & laborers NEC', 'Supervisors mechanics & repairers', and 'Supervisors production occupations' as 'Manager', while all other occupation categories were coded as 'Non-manager'.

[44] Calculation of the 95 percent confidence interval is described in footnote 39.

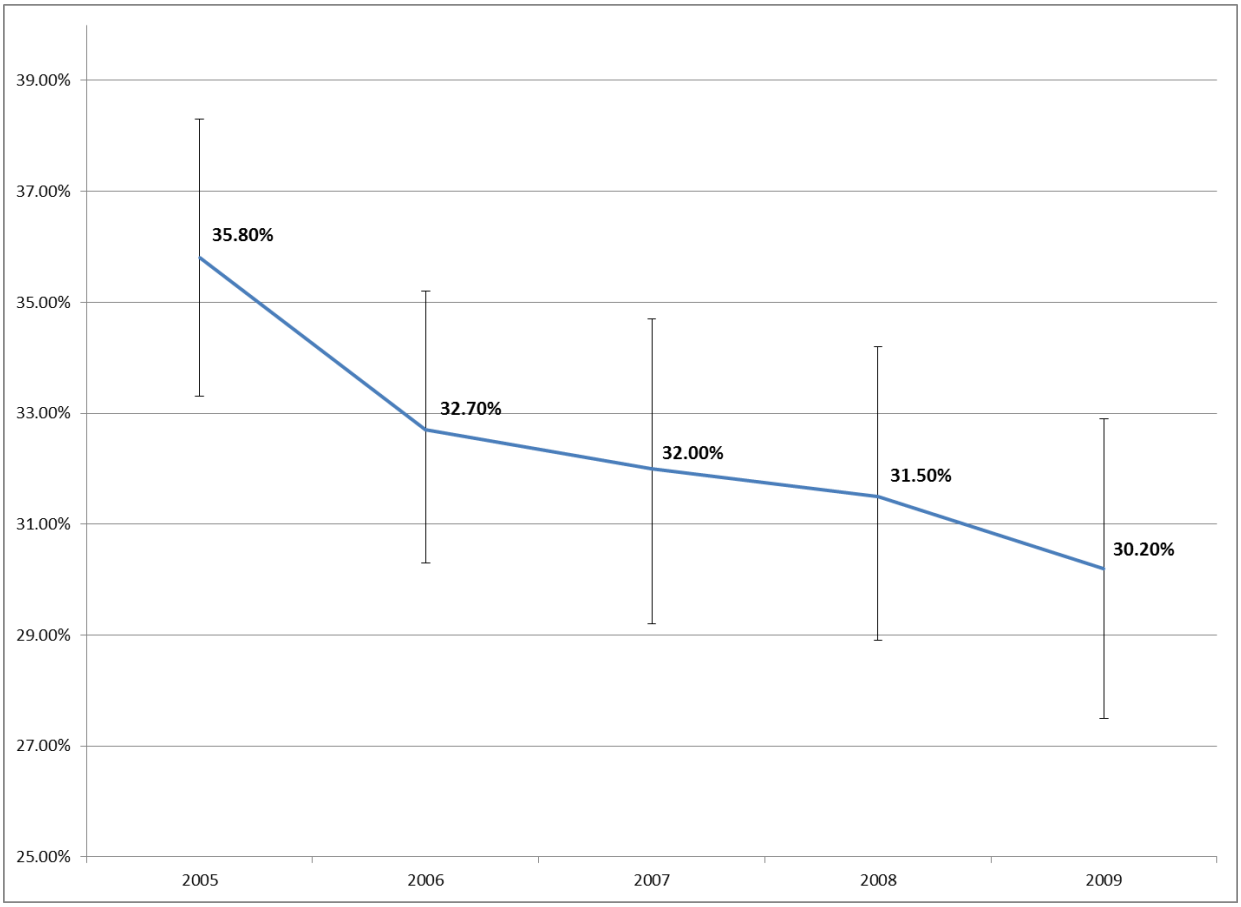


Figure 4 – 2005-2009 Non-Match Rates between Kentucky Workers' Compensation Data and MSHA Part 50 Data for Cases Involving at Least One Lost Day

4.2 California

4.2.1 Data Processing for California

California workers' compensation data for the years 2005-2009 (inclusive) was received from the California Division of WC's Workers' Compensation Information System (WCIS) in the form of Excel worksheets with 4 tabs organized respectively by NAICS, SIC, "Class Code," and "Industry Description." NAICS was not collected by for the California WC data throughout the reference period and various cases contained several of the four codes specified. The case selection steps described below regarded NAICS, when present, as primary, followed by SIC, followed by Class Code, and finally by Industry Description. At each step of the process, selected cases were de-duplicated based on case-specific identification numbers to eliminate duplicate cases having both the code at the previous step and the code at the current step, retaining the NAICS code where present.

California Workers' Compensation Program Data: Selection of Case Records

Several database preparation steps were undertaken to establish a file of California workers' compensation cases for matching with MSHA data. These steps are described below and the results are summarized in Figure 5. Appendix E provides a summary table showing cases retained or excluded at each step in the case selection process.

- **Step 1 – Retain only case that occurred within the 2005-2009 reference period.** The California WC file included a total of 9,097 cases and 17 cases were filtered from the file as outside the reference period. Thus, 9,080 cases were identified as occurring within the reference period.

- **Step 2 – Retain only cases that were "mining injuries".**[45] This step entailed a number of sub-steps:

  - *Sub-step 1 – Match SIC to NAICS:* 651 cases were filtered out by matching a NAICS-based case listing against a SIC-based case listing and then de-duplicating based on the Jurisdiction Claim Number (JCN), which is a unique number assigned to each claim in California's Workers' Compensation Information System (WCIS).

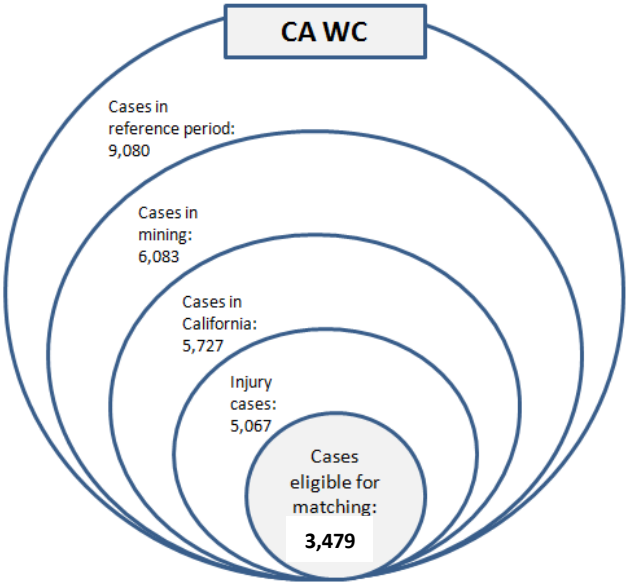


Figure 5 – California Workers' Compensation Data Processing Summary

  - *Sub-step 2 – Match the result of sub-step 1 against CA business classification codes:* 214 cases were filtered out by matching the sub-step 1 result against a Class Code-based case listing. (CA's Standard Classification System, which contains approximately 500 industry classifications, describes groups of employers whose businesses are relatively similar. The primary purpose of the classifications is to divide payroll data into groups.)

  - *Sub-step 3 – Match the result of sub-step 2 against industry descriptions:* 2,132 cases were filtered out by matching the sub-step 2 result against an industry-based case listing case. Two relevant data elements were used to identify mining establishments: "Industry Name" (the superordinate category) and "Industry Description" (a more detailed classification). Specifically, the unique combined values of the data elements were identified and then each record was coded for whether it identified a mining business.

  Mining industries were identified by referring to the industry descriptions attached to all the sub-categories within NAICS 212. When Industry Name and Industry Description were in

[45] A mining injury is defined as a person who was injured while working at a company classified by the California WC system as being in the mining industry.

conflict (e.g., when Industry Name was 'animal specialty services, except veterinary' and Industry Description was 'metal mining'), whether to include the case was resolved by referring to the Occupation of the injured worker(s). Cases with an Occupation that accorded with a mining-related component (e.g., 'mining machine operators') were retained, while cases where the Occupation matched a non-mining component (e.g., 'animal caretakers except farm') were dropped. (See Appendix F for the coded list of California WC industries). As a result, 6,083 cases were identified as occurring at mining businesses.

- **Step 3 – Retain only cases that occurred in California.** As with the Kentucky workers' compensation data, the actual injury does not need to occur within the state to be included in the workers' compensation data. Cases in the California WC file included the ZIP code location where the injury occurred. Using the ZIP code, 224 cases were filtered out as not occurring in California. Thus, 5,727 cases were identified as having occurred in California.
- **Step 4 – Retain only cases that are reportable under Part 50.** MSHA defines injuries as resulting from "a work accident or exposure involving a single instantaneous incident in the work environment," while "repeated trauma or repetitious movement which produces tenosynovitis is considered an illness" (Report on 30 CFR Part 50, p. ii). Also, injury to the lungs from inhaled substances is considered an illness by MSHA. In contrast, California WC considers repetitive-stress injuries and lung conditions to be injuries. Given these differences, these two conditions were excluded from the WC dataset; 660 cases were filtered out based on injury type. No cases were filtered from the California WC dataset based on days away from work because 447 of the California WC cases with counts of zero for days away from work matched to cases in the MSHA California dataset. Thus, 5,067 cases were identified for matching based on injury type.
- **Step 5 – Resolve conflicting industry coding information and remove cases where conflicts could not be resolved.** A non-trivial number of cases were found to have conflicting information in combinations of NAICS, SIC, Class Code, Industry Description. These cases were reviewed and, where possible, conflicting information was resolved. Based on this, 1,575 cases were filtered out due to conflicting information that could not be resolved. Thus, 3,479 cases were thus identified as eligible for matching.

MSHA California Data: Selection of Case Records

Similar to the workers' compensation data, several database preparation steps were undertaken to establish a file of MSHA Part 50 California cases that aligned with the requirements of the California WC system. Cases were selected from the MSHA Accident/Illness (AI) files based on the following criteria, as described below and illustrated in Figure 6. Appendix E provides a summary table showing MSHA Part 50 California cases retained or excluded at each step in the case selection process.

- **Step 1 – Retain only cases that occurred within the reference period of 2005-2009.** This resulted in selecting 68,113 cases in 2005-2009.

- **Step 2 – Retain only cases that involved an injury.** MSHA Part 50 data provides a field ("aii") that codes each incident based on the type of harm suffered by the employee. Values in this field that are 21 or under indicate the record is an injury; a value above 21 in this field indicates the record is an illness. As noted above, the focus of the evaluation was underreporting of injuries. Thus, 66,488 cases were identified as involving an injury (i.e., had a value in the "aii" field of 21 or under).
- **Step 3 – Retain only cases that occurred in California.** MSHA's "state" field was used to identify the location of the mine where the injury occurred. Thus, 2,026 injuries were identified as having occurred in California mines.

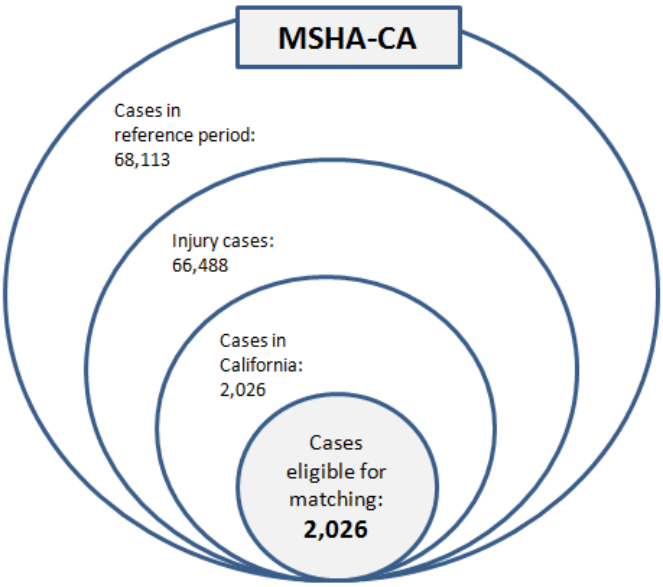


Figure 6 – California MSHA Part 50 Data Processing Summary

- **Step 4 – Retain only cases where the employee lost more than one day of work time.** The California WC system only includes injuries that are serious enough to cause more than one day of lost work time (i.e., the day the injury was sustained). Thus, injury cases resulting in fewer days away from work were filtered from the MSHA dataset. MSHA's "dayslost" field records the count of actual days lost from work as the result of a particular injury, and the "deginj" field classifies injuries by type and degree, including indication of whether an injury resulted in any days lost from work. These data elements were linked to identify cases involving more than one lost workday. No cases were filtered from the MSHA dataset based on this criterion, however, because 447 of the MSHA California cases with counts of zero for days away from work matched to cases in the California WC dataset.[46] Thus, 2,026 cases were thus identified as eligible for matching.

4.2.2 Approach to Matching California Workers' Compensation Data to MSHA Part 50 Data

Determining Which Data Elements/Fields to Use in the Matching Process

As with the Kentucky data, ERG and NORC performed an initial review of both data sets to determine the appropriate data elements to be used in the matching process. Based on this initial review, eight data elements/fields were identified for possible use in matching records across the two datasets.

[46] California workers' compensation recording guidelines indicate that cases in the record should only be those with more lost time than the day of injury itself, so cases with 0 lost days (days away from work) should not likely match. However, given the large number of such cases in the data, it was decided to retain those cases in the MSHA file and assess the number that did indeed find a match.

After further review and some coding adjustments, however, only seven of these were selected for matching attempts. The eight data elements considered are described in Table 13.

Table 13 – Data Elements Used in Matching WC and MSHA Data

Data Element	Used in Matching	Characterization
Employee age	✓	These data elements had either exactly the same coding in the two data sources or could be equalized with minimal recoding. Only the recoding of "body part injured" required any exercise of judgment. [a]
Employee sex	✓	
Body part(s) injured	✓	
Date of injury	✓	
County of injury	✓	
Company name	✓	For each reported case, MSHA data provide two data elements identifying the worker's employer: the name of the company that is the operator of the mine, and also the name of the contractor employing the worker (where applicable). California WC data contains the name of the injured worker's employer, and this variable was used in the matching process. For the MSHA data, preliminary investigations revealed that matching was more successful when company name was set to the name of the contractor company when one was present and to the name of the operator company otherwise. So that was the procedure followed. Company name information was in string format in both the WC and MSHA data files. In order to maximize the power of the matching process using company name, recoded versions of both data elements were produced in which the key one or two words were extracted from company name. [b]
Nature of injury	✓	Nature of injury is coded in the MSHA data by the field 'natinj' and in the WC data by NATURE_OF_INJURY_CODE'. These two fields are conceptually similar but have small differences in exactly what distinctions are made among the various types of injuries. For example, while the WC data contains one general code for burns, the MSHA data distinguishes among '120 Burn or scald (heat), '130 Burn, chemical, '301 Non-Contact electric arc burn', '302 Laser burn', '304 Sunburn', and '360 Electrical burn'. There were also instances where the WC codeframe made distinctions that the MSHA data did not. We constructed a joint codeframe for nature of injury and combined separate values on each side where necessary to equalize the two data sets.

Data Element	Used in Matching	Characterization
Cause of injury	✖	We also examined variables coding for cause of injury for possible use in matching. However, the fields involved ('sourcinj' on the MSHA side and 'CAUSE_OF_INJURY_CODE' in the WC data) proved to be impossible to recode into the same system because the codeframes used in the two data sources were too different. The MSHA field contains many more values than the WC file, distinguishing among, for example, 'Pulverized mineral', 'Sand, gravel, shell', and 'Loose dirt & mud' as sources of injury, and focuses solely on the material source of the injury, while the WC values are fewer but combine source and nature of injury, using such codes as '61 Strain or injury by: Wielding or Throwing'.

[a] For example, MSHA distinguished between injuries to the ears as such and injuries affecting hearing, while the California WCIS does not. So, the two values of 121 and 122 for MSHA 'partbody' were collapsed into one for matching with the WCIS variable PART_OF_BODY_INJURY_CODE value 13 "Head: Ears". (The list of comparable categories across data sources is found in Appendix G.) Note also: Employee age can be compared using an age data element or a date-of-birth data element. The MSHA data available for the study included only age and did not explicitly list the date of birth (although this value is collected by MSHA on the Accident form 7000-1). However, the California data includes a date-of-birth data element, and this was used to calculate age for matching with the MSHA data.

[b] For example, a company name like 'J D Edwards Company, Inc.' would be designated 'JD Edwards'; 'Burnett Group LLC' would be listed as 'Burnett Group'. This procedure was designed to simplify and standardize company name while taking a conservative approach that would avoid combining similar but distinct names into the same recoded category. For example, 'Shea Kenny Joint Venture', 'Shea Kenny JV', and 'Shea Kenny JV Center' were all recoded into 'Shea Kenny', since the differences among them involved only punctuation and incidental words. However, since 'Hansen Aggregates' and 'Hansen Bros. Enterprises' differed in substantive ways, they were maintained as separate company names in the recoded version.

Approach to Matching the Datasets

As with the Kentucky analysis, it was necessary to determine the appropriate criterion for saying that two records matched to one another. Specifically, ERG and NORC needed to determine how many of the data elements/field needed to match to say that two records were a match. To do this, ERG and NORC performed some initial investigatory matching. Accordingly, based on investigatory matching, ERG and NORC counted cases as matching if they shared values with any five (or more) of the seven data elements / variables available for matching, but only if the number of MSHA cases with those values was equal to or greater than the number of WC cases involved.

Matching Limitations and Considerations

Regarding the level of confidence in the matching rates and other analytical results of this comparison between California WC system data and MSHA data, the following limitations / considerations should be kept in mind:

- ***Relative differences in case counts between the two datasets.*** Even before attempting the matching of cases between datasets, the counts of cases available for matching lead to the initial conclusion that there must be some level of underreporting in the MSHA Part 50 data. The California workers' compensation data contained 3,492 cases while the MSHA Part 50 data had only 2,026 cases. As a result, it was apparent before matching was conducted that there were an insufficient number of cases in the MSHA dataset for possibly finding a substantial match rate.

- ***Data errors.*** These data have some of the same limitations/considerations as noted above for the Kentucky WC data. But also, specifically to California WC data, there is an unknown rate of error associated with each of the key matching variables of age, injury date, injured body part, county, company name, and gender that could affect the accuracy of the reported matching rate. Furthermore, data from the California WCIS system were not totally clean (e.g., administrators indicated they are aware of possible errors in coding of industry). Specifically, earlier data from California WC did not use NAICS, some WC data relies on SIC codes, still other data are based on California "Class Codes." Some cases include more than one of these codes, and the coding is not always consistent. We attempted to minimize the impact of this potential source of error, but it remains a possibility in the dataset.
- Other limitations/considerations that are the same as those noted above for the Kentucky WC data:
  - Matching involved a small set of variables/data elements.
  - The existence of match duplicates.
  - Uncertainty resulting from variable recoding.
  - Missing and inconsistent data.

4.2.3 Results for Matching California Workers' Compensation Data to MSHA Part 50 Data

Using the matching definition described above,[47] ERG and NORC found that 54.1 percent (1,881 cases) of California workers' compensation cases found a match in the MSHA data. NORC calculated a 95 percent confidence interval for the match rate of 52.4 percent to 55.8 percent.[48] This would indicate that underreporting in California could be as high as 45.9 percent (1,598 cases) for injury cases involving at least one day away from work with a 95 percent confidence interval of 44.2 percent to 47.6 percent.

Once again, it may not be reasonable to assume that all non-matched cases are underreported cases. That is, due to the data limitations, we may have missed some matches. Following the reasoning from the Kentucky analysis, we calculated a range for underreporting percentages using a set of assumed "miss rates." These are presented in Table 14. As before, using a 30 percent miss rate as a worst-case scenario (see logic for this in Section 4.1.3 above), then we can state that underreporting is somewhere between 32.2 percent and somewhere less than 45.9 percent in California for injury cases where the worker lost more than one day of work.

[47] Cases are considered to match if they share any five (or more) of the seven variables and if the number of MSHA cases with those matching values is equal to or greater than the number of WC cases involved in the match.
[48] See footnote 39 for details on this calculation.

Table 14 – Estimated Underreporting Percentage for California for Assumed Matching Miss Rates

Miss Rate	Number of Non-Matched Cases	Underreporting Percentage
0%	1,598	45.9%
20%	1,278	36.7%
30%	1,119	32.2%
40%	959	27.6%
50%	799	23.0%

Table 15 provides a set of cross-tabulations on the matching. We summarize these cross-tabulations below.

Cause of Worker Injury

The categories used by California WC to code for the cause of a worker's injury, although not compatible with the scheme used by MSHA, can be combined in a reasonable way to enable us to examine the relative match rates of the different categories. The non-match rates range from 57 percent for incidents involving motor vehicles,[49] to 33.7 percent for "Caught in/under/between. The cases least likely to find a match in the MSHA data were: "Burn or scald," the previously mentioned motor vehicle category, and incidents involving a fall, trip, or slip. The remaining five causes of injury had lower non-match rates. Because WC does not code for relative severity of injury, it is not possible for us to identify whether these less-common causes of injury also involved less-serious injuries.

Body Area of Injured Worker

Table 15 shows non-match rates by injured body area, using values of Body Part collapsed into four major categories. The highest non-match rate (60.2 percent) was observed for "Multiple and unclassified" injuries. This is also the lowest frequency category, and this match rate indicates that 157 cases of multiple and unclassified injury went unreported to MSHA in California during the reference period. The second highest non-match rate (52.1 percent) was observed for injuries of the Head, Neck, and Spinal Cord, indicating that 291 cases of this type of injury went unreported to MSHA in California during the reference period. The next two highest non-match rates correspond to the categories of upper and lower extremities. Taken together, these categories represent over half of the injuries recorded by the workers' compensation data in California during the reference period, and the non-match rate for each of these categories is between 40 and 43 percent, suggesting that over 800 cases of these injuries were not reported to MSHA. The final category, "Trunk and Internal Organs" had a non-match rate of 48.5 percent, indicating that an additional 344 cases went unreported.

Age

Table 15 shows non-match rates for age quartiles. As with the Kentucky analysis, these rates vary little across age.

[49] Motor vehicle-related injuries may go unreported to MSHA if they occur on a mine site (making it reportable under Part 50), but the injured worker is employed by an entity delivering (or visiting) the mine site. Thus, the mine operator may not know that a reportable injury even occurred.

Table 15 – Number of Non-Matched Cases and Non-Match Rates from Comparing California Workers' Compensation Data to MSHA Part 50 Data for California (2005-2009) for Cases Where the Worker Lost More Than One Day or Work, Overall and by Various Categories

Category	Number of California Workers' Compensation Cases Used in Matching Process	Number of California Workers' Compensation Cases Not Matched to MSHA Part 50 Data	Percentage of California Workers' Compensation Cases that Did Not Match to MSHA Part 50 Data
All cases	3,479	1,598	45.9%
Cause of Worker Injury			
Burn or scald	121	62	51.2%
Caught in/under/between	208	70	33.7%
Cut/puncture/scrape	300	104	34.7%
Fall/slip/trip	612	294	48.0%
Motor vehicle	114	65	57.0%
Strain	1043	448	43.0%
Striking against/stepping on something	179	77	43.0%
Being struck by or abraded by something	487	208	42.7%
Other cause of injury	357	125	35.0%
Total	3,421 [a]	1,861	54.4%
Body Area of Injured Worker			
Head & neck (incl. Spinal cord)	558	291	52.1%
Trunk & internal organs	709	344	48.5%
Upper Extremities	1,186	478	40.3%
Lower Extremities	774	332	42.9%
Multiple and unclassified	260	157	60.2%
Total	3,487 [a]	1,602	45.9%
Worker Age Quartiles			
17 - 28	879	393	44.7%
29 - 37	930	434	46.7%
38 - 47	829	382	46.1%
48 & over	840	388	46.2%
Year			
2005	879	416	47.3%
2006	783	355	45.3%
2007	803	377	46.9%
2008	535	223	41.7%
2009	492	233	47.4%

[a] Total is less than 3,479 for these sets of categories since not all cases could be coded for matching.

Year to Year

These results show little change in non-match rates over time (see Table 15) with a high of 47.4 percent in 2009 and a low of 41.7 percent in 2008 and most rates in the 45-47 percent range. Figure 7 plots these rates over time and provides the 95 percent confidence interval around each year's rate.[50]

[50] See footnote 39 for calculation of the confidence interval.

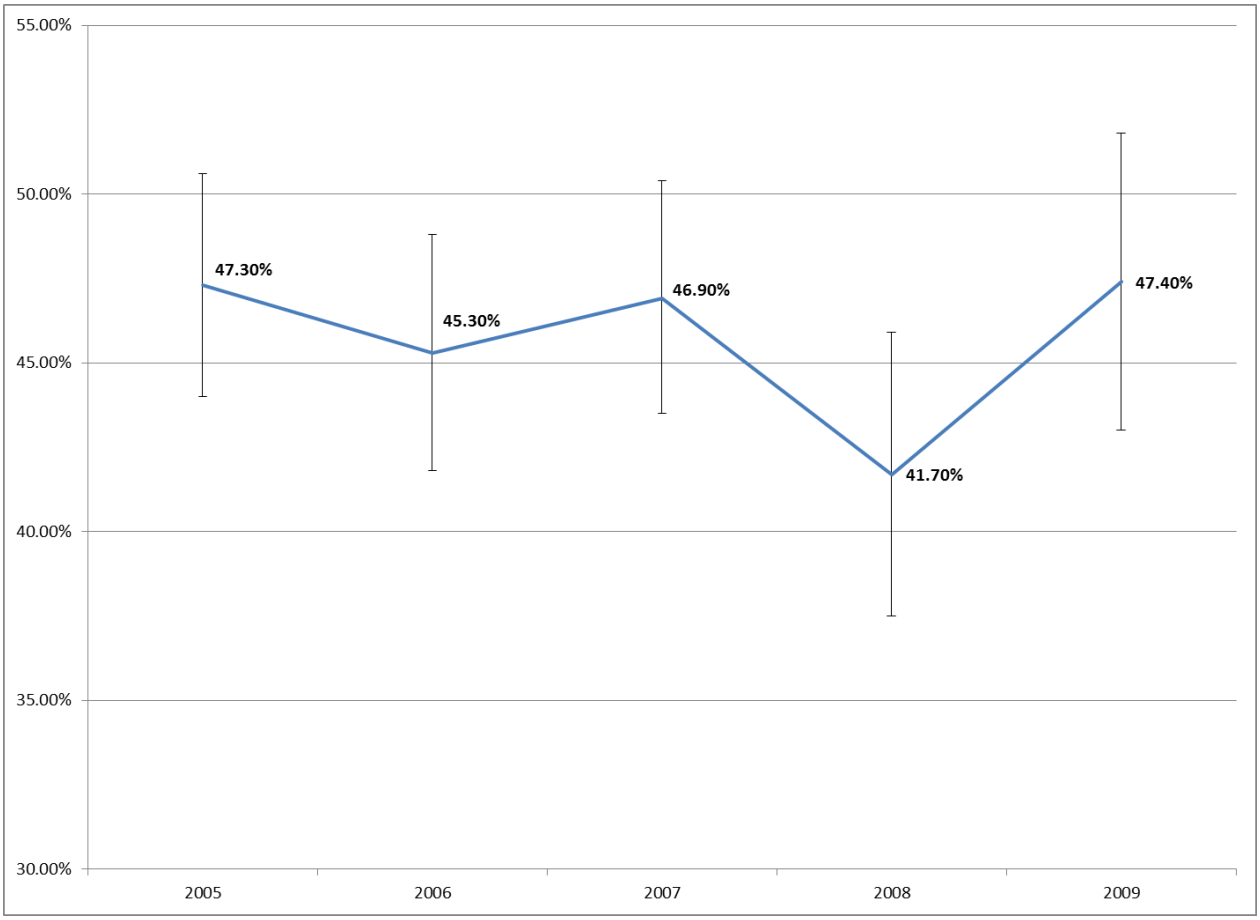


Figure 7 - 2005-2009 Non-Match Rates between California Workers' Compensation Data and MSHA Part 50 Data for Cases Where the Worker Lost More Than One Day of Work

5.0 PART 50 REPORTING AND RECORDKEEPING PROCESS ANALYSIS

This section describes the data and information collected for the process analysis we performed for this project. The process analysis drew from two primary sources of information: ERG's review of internal and public documentation on the Part 50 reporting and audit process and interviews with MSHA personnel. To provide context for this analysis, we also performed a small number interviews with miners. The process analysis we performed review three interrelated process:

- **Operators' reporting processes.** This is the process that occurs once an injury occurs and the steps that need to be taken by operators to ensure accurate reporting to MSHA.
- **Data flow processes at MSHA.** The process used by MSHA in processing data it receives from operators. In order for data to be reported accurately, it is necessary for accurate and timely processing to occur.
- **Audits.** The process used by MSHA in performing Part 50 audits. This is the process used by MSHA to audit operators' reported Part 50 data. MSHA must be able to verify the information that operators provide to ensure its accuracy.

5.1 Injury and Illness Reporting by Mine Operators

5.1.1 Process

The Part 50 reporting and recordkeeping process begins at the time that an injury occurs and ends when the injury information is either: (1) not communicated to MSHA or (2) when MSHA receives injury information from the operator. Figure 8 shows the primary steps in the reporting and recordkeeping process that ERG distilled from a review of background documentation and clarified through interviews MSHA staff. The shaded area of the map indicates the operator responsibilities for reporting under Part 50. The map has four end-points for an injury:[51]

- **Not reported, no violation.** This represents an endpoint where the miner does not report an injury that is subject to Part 50 reporting requirements. No violation occurs since miners are not subject to Part 50 requirements, only operators are required to report.
- **Not reportable under Part 50.** These are injuries that are either not work-related or not reportable under Part 50.
- **Violations.** These are situations where the operator has not followed Part 50 requirements for reporting and recordkeeping. Any injury that ends in the violation end-point represents underreporting.
- **Included in Part 50.** These are injuries where the operator has followed the requirements under Part 50 and the injury is included in MSHA's data reporting on injuries and illnesses.

[51] The discussion here focuses on injuries based on the scope for this project.

As can be seen in Figure 8, there are several opportunities for underreporting of injuries in the process. The following describes the issues that arise at each of the step in the process:

- *Worker reporting of injuries*. When a miner is injured, he or she may have the option of not telling his or her supervisor of the injury.[52] If the miner does not inform his or her supervisor of the injury (and it is the type of injury that is covered by Part 50), then the injury is not captured in the Part 50 data reporting and no violation has occurred (operators, not miners, are required to report under Part 50). The concern here, however, is why the worker did not report the injury. If the worker did not report due to fear of retaliation by his/her employer, then this is an issue. Additionally, operators may use incentive programs that link bonuses to low injury counts creating an incentive for miners to not report a reportable injury. That is, an operator's actions or messages may be deterring miners from reporting their injuries to their supervisors, leading to underreporting of injuries.
- *Determining work-relatedness and reportability under Part 50.* Once the injury has been reported to the mine operator, then the operator must determine if the injury is work-related and if it is reportable under Part 50. For work-relatedness, MSHA's Part 50 regulations are clear that injuries occurring on the mine site are work-related.[53] Furthermore, MSHA has provided clear guidance on the reportability of injuries under Part 50.[54] If the operator determines that the injury is both work-related and reportable under Part 50, then the process moves to the recording and reporting phase of the process. However, an incorrect determination about the work-relatedness or the reportability of the injury is a violation.
- *Recording and reporting to MSHA.* Once an injury is determined to be reportable under Part 50, the operator must record that injury on a 7000-1 form and send that form to MSHA.[55] Failure to do so constitutes a violation.
- *Tracking and recordkeeping.* For cases that involves days away from work, the operator is required to track return to work status of the injured miner to provide a form 7001-1(d) to MSHA when the miner returns, detailing the time the worker was away from work related to the injury. Failure to provide this information to MSHA is also a violation of Part 50 requirements.

5.1.2 Sources of Underreporting

Figure 8 depicts a number of ways (red arrows) in which a violation, and hence underreporting, can occur in the Part 50 reporting process. The primary source of underreporting in the process is that operators do not perform their responsibilities under Part 50 when a reportable injury or illness occurs. However, underreporting can also stem from other, related, sources, including:

[52] In many cases this may not be the case since the miner is injured in a place or way that it will become known to his/her supervisor.
[53] There may be some injuries, however, where the mine operator may dispute the work-relatedness of the injury. For example, an operator may dispute the claim by miner that a back strain is work-related and instead claim the strain stems from the miner's personal life.
[54] http://www.msha.gov/stats/part50/rptonpart50.pdf.
[55] Immediate notification is required under 30 CFR 50.10 in case of: (a) a death of an individual at the mine; (b) an injury of an individual at the mine which has a reasonable potential to cause death; (c) an entrapment of an individual at the mine which has a reasonable potential to cause death; or (d) any other accident.

- Fear of reprisal
- Incentives or disincentives programs
- Gray areas in reporting

Fear of reprisal

One possible reason for a miner to not report an injury is if he or she fears some form of reprisal from the mine operator. This could occur if the operator somehow penalizes the worker for reporting an injury such as putting the worker on undesirable workshifts or jobs.

Incentive and disincentive programs

Some operators offer incentive programs that discourage injury and illness reporting. For example, an operator may offer a bonus to its miners for having no injuries over a certain time frame. Thus, if a miner reports an injury, then he and other miners at his mine may lose bonus money. This type of program could discourage reporting of injuries.

Gray areas in reporting related to first aid and medical treatment

A final area that may contribute to underreporting has to do with "gray" areas in what types of injuries are reportable. Under Part 50, only injuries that require medical treatment need to be reported and injuries only require first aid do not. The text of 30 CFR 50 provides details on what must be reported and MSHA's "Yellow Jacket" report[56] provides guidance on distinguishing between the medical treatment and first aid. Thus, MSHA staff indicated to ERG that there should be little or no uncertainty as to what is and what is not medical treatment (and hence, reportable under Part 50). However, other MSHA staff commented that the nuances between first aid and medical treatment may create a perceived "gray area" for some mine operators that leads to underreporting. Nevertheless, it is the responsibility of the mine operators to know and understand the law and the text of 30 CFR 50 and to accurately report injuries.

[56] http://www.msha.gov/stats/part50/rptonpart50.pdf.

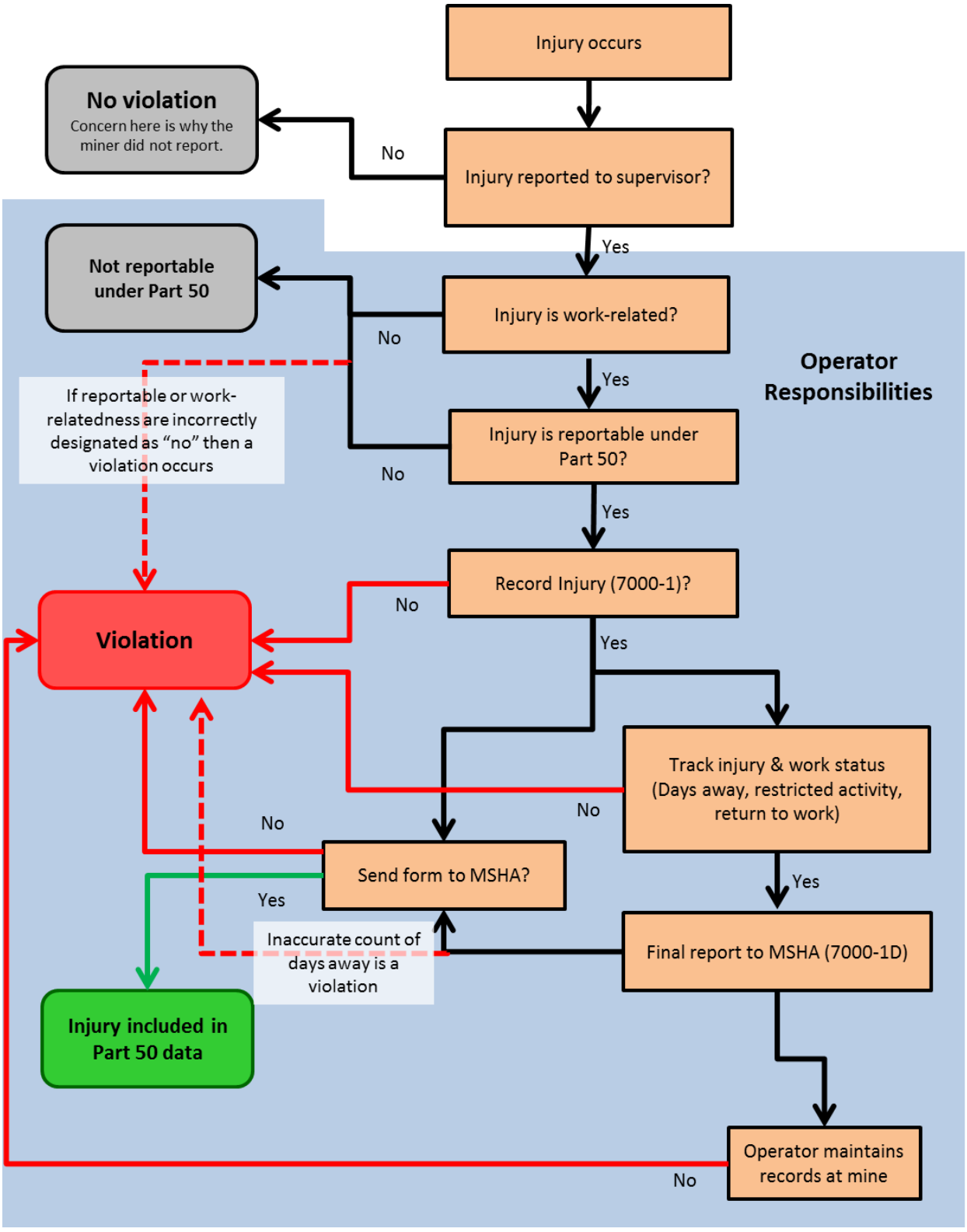


Figure 8 –High-Level Process Flow Map for Injury and Illness Reporting at the Operator Level

5.2 Part 50 Data Flow Analysis

Another potential influence on the accuracy and completeness of Part 50 data is the way that the Part 50 information flows from the operator level to MSHA and how that data is received, stored, processed, and utilized by the Agency. ERG interviewed MSHA staff responsible for collecting, processing, and summarizing Part 50 data in order to better understand the data itself as well as how the Part 50 data flows from the mine level to MSHA and is used by the Agency.

5.2.1 Process Description

Figure 9 shows how the various sources of Part 50 data flow from operators to MSHA. The process map shows how the Part 50 information stemming from the regular Part 50 reporting and compliance checks is similar, and ultimately follows the same processing path as the Part 50 data collected through audits.

Data for use in the Part 50 reporting are generated under the following scenarios:

- Regular Part 50 reporting requirements require operators to submit 7000-1 forms for accidents and reportable injuries as well as quarterly employment and production information via 7000-2 forms. This is described in Section 5.1 above.
- Revised or updated 7000-1 and 7000-2 forms may be submitted as a result of an audit, where the operator might need to resubmit previously submitted forms or send entirely new 7000-1 forms to reflect MSHA's audit findings (e.g., submit a 7000-1 form for an injury that the audit revealed had not been reported by the operator, re-submit a 7000-2 for to reflect an inaccuracy in employment hours)

5.2.2 Processing Part 50 Data

The primary steps for processing Part 50 data (as indicated in Figure 9) include:

- Copies of 7000-1 and 7000-2 forms come into the Office of Injury and Employment Information (OIEI) and are scanned and saved in the Hummingbird database.
- Each incoming document receives a unique number. Based on the date and a code denoting whether the form is addressing employment, injury, or return to work information.
- OIEI Coders review the 7000-1 and 7000-2 forms, and if there are no evident problematic issues, the forms are entered into MSHA's Sunguard database.
- The lead in OIEI uses an excel spreadsheet to divide the incoming documents among coders, and the OIEI coders code their forms into the MSIS database.
- Coders may also be assigned to "miscellaneous mail", which includes any amended 7000-1 and 7000-2 forms that are sent to MSHA. Following up with miscellaneous mail may include discussions with Operators as to how their injuries/illnesses should be reported and/or coded. If a form has been amended, the file will be shown as having an adjustment. For example, the file will state a reason for a form having a correction.
- The MSHA Sunguard database (updated nightly) feeds the data to Data Warehouse.
- The webmaster puts information from Data Warehouse onto the data retrieval system (DRS).

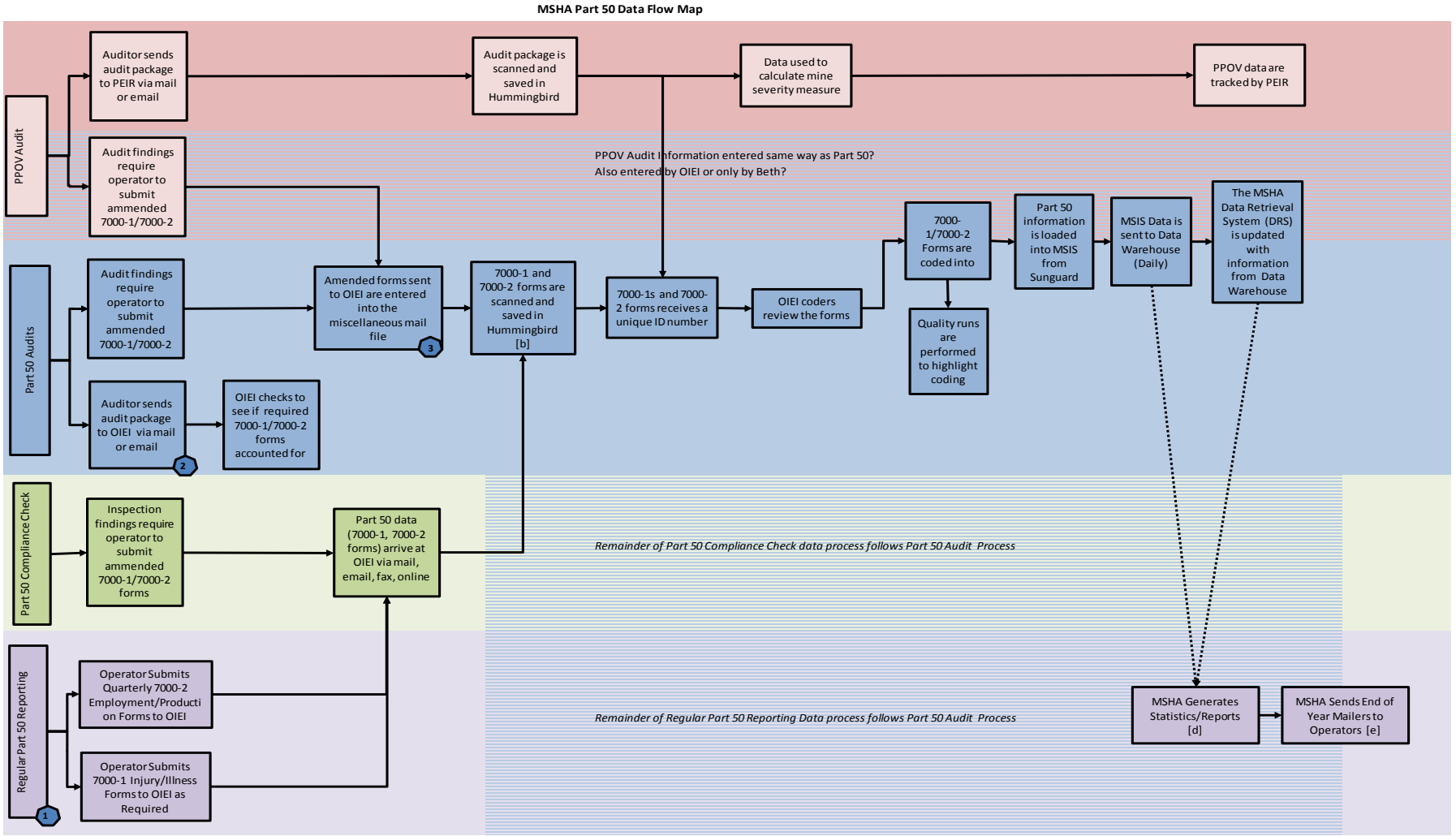
In addition to the Part 50 information gathered on 7000-1 and 7000-2 forms, completed audit packages submitted to MSHA by auditors might also present a source of Part 50 information that the Agency could use to improve injury reporting information. Audit packages contain the auditor's record of the number of injuries and illnesses that were underreported to MSHA. Although these underreported injuries should be reflected in the revised 7000-1 forms submitted by the operator, the information included in the audit packages can directly be captured by MSHA and use it to formulate an internal measure of injury underreporting.

5.2.3 Data Limitations

MSHA staff interviewees noted the following limitations to Part 50 Data:

- *Potential data coding ambiguities in some cases.* There may be some injuries which operators have trouble determining the coding for. In these cases, operators can request guidance from MSHA's Denver data process staff or their District Manager. MSHA has protocols in place for determining how injuries are coded and these are followed by the MSHA Denver staff and by DMs. Nevertheless, it is possible for the same type of injury at two different mines to be coded differently based on differing guidance provided by DMs. ERG expects, however, this is rare and that most injuries are coded consistently.
- *Closing out the data before all information on return to work is received.* MSHA closes its data for a calendar year on June 30 of the following year. For most injuries this does not pose an issue. However, for injuries where the miner has not returned to work by June 30 in the following year, MSHA truncates the days away from work measure using June 30 as a cut off. For example, if a miner was injured on December 2, 2011 and had not returned to work by June 30, 2012, then MSHA would record the worker as being away from work for 150 days.[57] Thus, for injuries where MSHA does not receive a 7000-1(d) form before June 30 of the following year, the data on return to work will not be accurate. For injuries where the worker has truly not returned by June 30 of the following year, the data underestimate return to work. For cases where the operator has neglected to send in a 7000-1(d) form (a violation Part 50 requirements) for a worker who returned before June 30 of the following year, the data will overestimate days away. Despite this, however, this only applies in cases of long-term injuries (i.e., the worker is out at least 6 months) or where the operator fails to return the 7000-1(d) form.

[57] There are 30 weeks between December 2, 2011 and June 30, 2012. Assuming the worker works 5 days per week, then the total number of lost work days would be 30 × 5 = 150.



[a] All 7000-1s and 7000-2s arrive at OIEI with the exception of PPOV audit packages, which are received by the Data Branch Manager for PEIR
[b] Only the 7000-1 and 7000-2 forms from the audit package are coded and entered. The remainder of the audit package is only maintained for PPOV audit packages and Part 50 Audit packages are discarded except for their 7000-1/7000-2 forms. District Offices may keep the full audit package on file.
[c] The Part 50 coding manual specifies how information should be coded.
[d] Statistics and Reports include: Injury and illness incident rates; Injury Experience Reports; Injury and Worktime Reports (annual, quarterly); Mine Accident, Injury, Illness, Employment and Coal Production Statistics; Standard reports showing mines with missing employment and production information
[e] End of Year Mailers are sent after the 4th quarter closes. The mailers show the information that MSHA has on file for the operators/contractors.

Figure 9 – MSHA Data Processing Flow Diagram

5.3 Audit Processes

5.3.1 Overview of Audit Processes

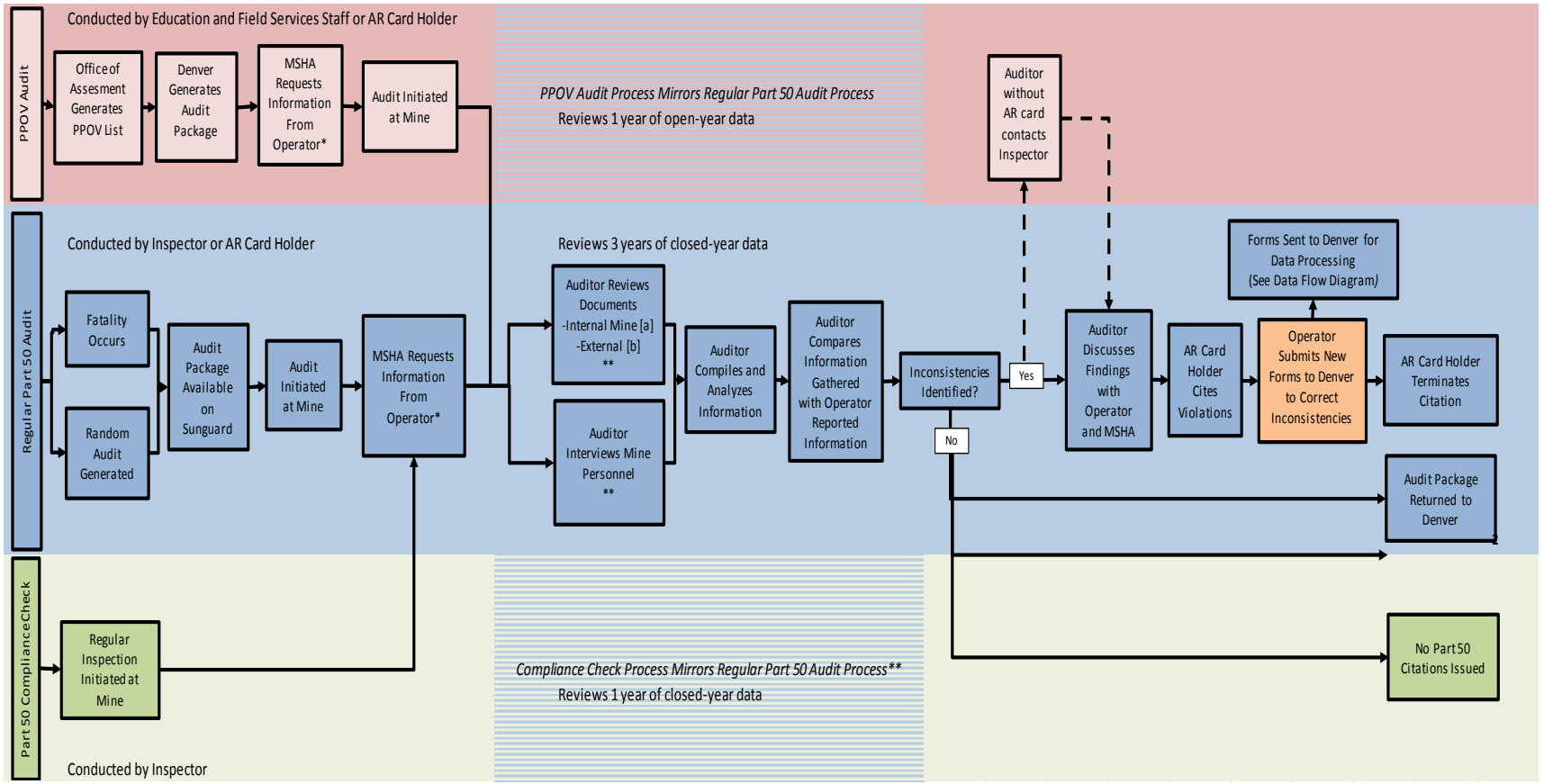
MSHA audit processes vary, depending on the type of audit being conducted as well as the circumstances that triggered the audit. Figure 10 depicts the various MSHA audit processes as well as the Part 50 compliance checks described in Section 2.4.1.

As Figure 10 shows, there are shared steps as well as differences among the different MSHA audit and compliance check processes. Process differences generally revolve around the: 1) type of MSHA personnel conducting the audit or records check; 2) procedure for checking and comparing Part 50 information, 3) type of data sources used as comparison sources with Part 50 records, and 4) time frame of Part 50 records reviewed.

Table 16 highlights these differences between the MSHA audits and records checks and is followed by further detail on each of these components.

5.3.2 MSHA Personnel Conducting Records Checks and Audits

The type of MSHA personnel conducting records checks and audits can vary by monitoring activity. As indicated in Table 16, records checks and audits might be conducted by MSHA inspectors or Education and Field Safety (EFS) staff. EFS began helping conduct audits as MSHA began additional rounds of audits under PPOV. EFS staff currently receive in-depth training to better conduct audits. MSHA inspectors hold current AR cards allowing them to cite violations. EFS staff, who are generally former inspectors, hold “Right of Entry” cards and do not hold AR cards, and therefore might not be able to cite violations during an audit. If EFS personnel do not hold AR cards, they must present any possible violations that need to be cited under Part 50 as a result of the audit to MSHA personnel that hold AR cards in order for citations to be issued to the operator.



* Information requested of the operator includes: 1) 7000-1 forms, 2) 7000-2 forms, 3) 30 CFR 50.11 reports of accidents and injuries, 4) employee timesheets, 5) payroll records, 6) sick leave requests or records, 7) medical records,8) medical claims forms, 9) doctor's slips, 10) emergency medical transportation records, and 11) drug screening documents. For PPOV audits, an information request letter is sent to the operator before the auditor arrives at the mine site.
** Compliance checks focus on comparing operator Part 50 information with MSHA records and checking timeliness of Part 50 information submitted.
***If the auditor (e.g. some EFS staff) does not have an AR card, an inspector would need to cite any violations found.
[a] includes information sources # 1-6 listed above.
[b] includes information sources # 7-11 listed above.

Figure 10: MSHA Part 50 Audit and Record Checks: Revised High-Level Process Flow Diagram

Table 16: Comparison of Audit Types

Audit Type	Timeframe of Records Reviewed	MSHA Personnel Responsible for Conducting Audit	Comparison Data Sources Potentially Used	Procedural Differences
PPOV	12 months[c]	MSHA Auditors, EFS Staff	• 7000-1 forms • 7000-2 forms • Payroll records • Timesheets • Training records • Ambulance Records • Medical records • Medical claims forms • Drug screening documents • Employee Interviews • Part 50 information on file in Denver	PPOV audits are among the more detailed audits conducted, however, their primary focus is not Part 50 reporting. Part 50 records are assessed and compared with other data sources during PPOV audits.
Regular/ Random Part 50 Audit*	3-5 years [b]	MSHA Inspectors	These audits would use most of the comparison sources used for PPOV audits; however, these audits may not use all of them to the same degree.	Procedure for conducting Part 50 audits and tools used to conduct them differ among auditors, however these audits tend to use more comparison data source than the compliance check and regular inspections
Sentinels Of Safety [a]	1 year	MSHA Auditors	• 7000-1 forms • 7000-2 forms • Payroll records • Timesheets • Training records • Employee Interviews • Part 50 information on file in Denver	May just address a specific department where the hours are reported, and this audit only uses one year of data. Does not necessarily use medical records as comparison data source.
Part 50 Compliance Check	1 year	MSHA Inspectors	• 7000-1 forms • 7000-2 forms • Timesheets • Part 50 information on file in Denver	Primarily looking for operator compliance with Part 50 submittal and timeliness information [c]

*Includes fatalities, random/regular audits initiated with district manager discretion
[a] MSHA is no longer a co-sponsor of Sentinels Safety Award and will no longer be completing this type of audit.
[b] Only 1 interviewee said 3-5 years, while the other two said 3 years.
[c]12 months of data may overlap fiscal years
[d] There was disagreement among interviewees as to whether compliance checks varied in scope to regular part 50 audits.

5.3.3 Audit Comparison Data Sources

As indicated in Table 16, there are a variety of information sources that MSHA personnel might use to verify injury and illness reporting, including:

- 7000-1 forms
- 7000-2 forms
- Payroll records
- Timesheets
- Training records
- Ambulance service records
- Medical records[58]
- Medical claims forms
- Drug screening documents
- Interviews with operators and employees

During interviews with MSHA staff, ERG asked about the types of information most frequently used in the audit process.

Ambulance Records

One potential comparison data source of interest to MSHA in this study is the use of ambulance records during Part 50 audits. Ambulance records are primary used for the following purposes:

- To determine if an injury occurred at a mine that involved medical treatment beyond first aid but might not have been reported. That is, after reviewing ambulance service records, the auditor can ask the operator to estimate the number of ambulance runs during a particular period and then can compare the response with the ambulance records.
- Since ambulance records generally include notes identifying the name of the person transported, a general characterization of the injury, and the hospital or clinic where the worker was treated, the auditor can check these records against those at the hospital or clinic, and then against the mine operator's records.

PPOV audits currently use ambulance records in assessing underreporting at mines.

Although ambulance records can provide a valuable source of information in determining underreporting of injuries, interviewees indicated that they are often utilized as a second layer of information that confirms worker injury-related information from other sources. Other key findings from interviews with MSHA personnel on the use of ambulance records in audits include:

- Reviews of ambulance records are not consistently conducted, or required, during Part 50 audits.

[58] MSHA inspectors and Auditors receive Health Insurance Portability and Accountability Act (HIPAA) training to ensure proper handling of these data.

- Interviewees did not consider obtaining and reviewing ambulance records to place an unnecessary time burden on auditors, with the process for obtaining the records estimated at anywhere from two to 24 hours.
- Ambulance records are one of the only sources of injury information that the operator cannot affect. For example, the operator may pay cash for medical bills so that there are no WC records associated with the injury. Alternatively, the operator might not record or mis-record injury information on the 7000-1 form that goes to MSHA.

5.3.4 Audit Resource Requirements

Table 17 summarizes the resource use for Part 50 audits by calendar year (2000-2012), mineral type, and mine type. MSHA spends slightly less than a third of an hour per mine employee conducting Part 50 audits. However, this ratio has increased over time. Additionally, more time per mine employee is needed for surface mines compared to underground or facilities.

Table 17 - Summary of Resource Use for Part 50 Audits, by Calendar Year, Mineral Type, and Mine Type

Calendar Year or Category	Number of Audits	Average Number of MSHA Staff Hours	Average Number of Employees at Mine	MSHA Staff Hours Per Mine Employee
Total	12,309	8.88	31.2	0.28
Calendar Years				
2000	922	10.00	42.0	0.24
2001	1118	8.23	34.4	0.24
2002	990	8.56	35.9	0.24
2003	1151	8.72	30.7	0.28
2004	1332	8.95	25.6	0.35
2005	1305	8.20	29.2	0.28
2006	1129	6.23	23.0	0.27
2007	966	6.83	27.3	0.25
2008	498	9.67	45.5	0.21
2009	718	9.37	30.9	0.30
2010	898	10.68	28.8	0.37
2011	659	11.77	27.1	0.43
2012	623	11.67	37.7	0.31
Mineral Type [a]				
Coal	1,405	23.38	82.0	0.29
Metal/Non-Metal	10,904	7.03	24.7	0.28
Mine Type				
Underground	1,117	21.87	100.5	0.22
Surface	10,335	7.20	21.1	0.34
Facility	857	12.41	63.7	0.19

[a] The large difference in the number of Part 50 audits in coal compared to those in metal/non-metal stems from the fact that there are significantly more metal/non-metal mines compared to coal mines in the United States. Table 3 provides summary information on the numbers of mines in 2012. For 2012, there were 12,193 metal/non-metal mines but only 1,865 coal mines.

Table 18 provides resource use data for PPOV audits from fiscal years 2011 to 2013. In 2011, the PPOV audits required 0.6 hours per mine employee and this increased to 2.5 and 2.0 hours per mine employee in fiscal years 2012 and 2013, respectively. Furthermore, the PPOV audits require significantly more time and are typically being performed at larger mines.

Table 18 – Summary of Resource Use for PPOV Audits by Fiscal Year

Fiscal Year	Number of Audits	MSHA Auditors Used (Range)	Average Number of MSHA Staff Hours	Average Number of Employees at Mine	MSHA Staff Hours Per Mine Employee
2011	39	1 - 4	74.4	124.4	0.6
2012	27	3 - 5	288.6	116.2	2.5
2013	16	2 - 5	294.0	148.4	2.0

6.0 ALTERNATIVE STRATEGIES TO AUDITS

This section discusses ERG's research into alternative strategies. We begin by discussing OSHA's approach to verifying the accuracy of injury and illness reporting. As discussed in Section 7, we did not find any strategies used by OSHA directly applicable to MSHA. Nevertheless, we have summarized our research findings here in this section. This section also discusses use of medical records and trauma registries for verifying injury and illness reporting.

6.1 OSHA's Approaches to Onsite Injury and Illness (I/I) Records Review[59] and Applicability to MSHA

ERG explored OSHA's experience related to recordkeeping compliance for possible approaches and/or mechanisms that might be relevant to MSHA. ERG reviewed OSHA's internal reports and public documents related to its efforts and conducted interviews with OSHA staff. OSHA has demonstrated an ongoing commitment to monitoring the accuracy of employer injury/illness recordkeeping and enforcing the requirements defined in 29 CFR Part 1904, *Recording and Reporting Occupational Injuries and Illnesses*. External drivers also shaped specific OSHA actions over time, including:

- Congressional interest based on research reports claiming extensive employer injury/illness underreporting.
- OMB requirements for OSHA to demonstrate a reasonable level of accuracy of employer injury/illness recordkeeping for approval to collect these data from a sample of employers and use to target inspections.

Like MSHA, finding a feasible approach for maintaining a resource-intensive process of records review represents an ongoing challenge for OSHA. This analysis traces the evolution of OSHA's approaches to monitor and enforce employer injury/illness recordkeeping[60] with a focus on two programs:

- The Annual Recordkeeping Audit Program (RK Audit Program)
- The National Emphasis Program on Recordkeeping (RK NEP)

The analysis excludes the more limited "records checks" reviews of the OSHA Log of Injuries and Illnesses conducted by compliance officers at the outset of an inspection.

[59] ***Onsite injury and illness records review*** is used in this report as a general term to refer to both OSHA recordkeeping audits and recordkeeping inspections.
[60] ***Injury/illness recordkeeping*** includes requirements for employer recording of incidents that meet workplace relatedness criteria on specified forms and selective reporting to OSHA.

Highlights of OSHA I/I Recordkeeping Requirements

Coverage. Applies to approximately 1.4 million private sector establishments under OSHA jurisdiction, except for designated low hazard SIC codes and establishments with fewer than 10 employees that are exempt from most requirements. A list of exempt SIC codes is available at http://www.osha.gov/recordkeeping/ppt1/RK1exempttable.html.

Forms. Employers not exempt from the Rule must use three forms to record work-related injuries and illnesses: Log of Work-Related Injuries and Illnesses (OSHA Form 300), Summary of Work-Related Injuries and Illnesses (OSHA Form 300A), and Injury/Illness Incident Report (OSHA Form 301).

Recording. Work-related cases that must be recorded involve: death, days away from work, restricted work, transfer to another job, medical treatment beyond first aid, loss of consciousness, significant injury/illness as diagnosed by a physician or other licensed health care professional. OSHA considers work-related cases as those caused, contributed to, or significantly aggravated by events or exposures in the work environment; that is, occurring in the workplace or in locations where the employee is located as a condition of employment.

Reporting to OSHA. Any work-related death or incident resulting in 3 or more in-patient hospitalizations within 8 hours of the incident. Also, establishment calendar year Injury/Illness Summary, if selected as part of the annual OSHA Data Initiative (ODI) collection from about 80,000 establishments in high-hazard industries.[61]

Employee involvement. Employers must: establish a procedure and notify employees how to report work-related injuries/illnesses, post Log Summary annually for three months of the year following its completion, allow employee access to own injury/illness records and protect their privacy. Under Section 11(c) of the OSH Act, employers are prohibited from discriminating against employees who do report. Employees, former employees, and employee representatives also have a right to access the Form 300 Log and any Form(s) 301 for their own injury or illness cases.

Despite differences between OSHA and MSHA, both agencies face some similar challenges in their efforts to detect employer injury/illness underreporting and enforce their respective regulations. Common issues for onsite injury/illness records review include how best to:

- Establish and maintain onsite monitoring of employer injury/illness recording and reporting.
- Identify establishments that are more likely to have underreporting.
- Manage resource constraints related to this effort.
- Protect employee rights to report injuries/illnesses and receive benefits to which they are entitled.

Still, lessons learned from OSHA's experience with the RK Audit Program and RK NEP may be applicable to improving MSHA's audit processes and the accuracy and completeness of injury/illness reporting. Examples of program elements contributing to OSHA's successes with injury/illness recordkeeping enforcement and improvement include:

[61] In addition, the Bureau of Labor Statistics (BLS) annually mails the *Survey of Occupational Injuries and Illnesses* to randomly selected establishments nationwide. As a result, the employee injury/illness records required by OSHA provide the source data for BLS's occupational injury and illness statistics.

- ***Commitment*** to accurate recordkeeping through audits and injury/illness recordkeeping enforcement at all levels of the Agency.
  - *Established and maintained onsite records review:* Based on its conviction that onsite records review is the best approach to monitoring and enforcing injury/illness recordkeeping requirements, OSHA maintained an RK Audit Program for 11 years (1998-2009) and then implemented the recently concluded RK NEP (2010-2012). For both programs, accuracy determinations focused on the types of underrrecording and misrecording errors that affect an employer's injury/illness rate along with the reported hours worked.
  - *Revised recordkeeping rule to clarify and simplify:* OSHA's recordkeeping rule (29 CFR Part 1904) defines requirements for employers covered, use of forms and records retention, injury/illness recording criteria, reporting to OSHA, and employee involvement and protections related to injury/illness records. About 10 years ago, OSHA released major revisions to the rule that clarified regulatory requirements, created simpler recordkeeping forms, allowed employers to maintain electronic injury/illness records, expanded provisions for employee involvement, and provided greater employee privacy protection related to injury/illness records.
  - *Provided explicit worker protection for reporting:* OSHA's revised recordkeeping rule emphasizes that the OSH Act of 1970 prohibits employers "from discriminating against an employee for reporting a work-related fatality, injury or illness" (29 CFR Part 1904.36). In addition, a March 2012 memo from OSHA's Deputy Assistant Secretary alerts Regional Administrators and Whistleblower program managers to "several employer practices that can discourage employee reports of injuries and violate section 11 (c) or other whistleblower statues." The memo identifies types of situations OSHA knows have occurred and provides guidance on investigating similar potential cases.
- ***Process standardization***.
  - *Trained auditors and provided feedback on adherence to established procedures.* For both onsite records review programs, the OSHA National Office conducted training for at least one inspector from each Area Office. In addition, the National Office provided a point-of-contact to answer questions related to implementing the records review methodology and to review collected information from audits/inspections for adherence to the protocol, then providing feedback as needed. OSHA found that adherence to the protocol improved over time along with inspectors' general skills and awareness of potential underreporting during non-recordkeeping-related site visits.
  - *Developed IT tools to support auditors, facilitate standardization and efficiency, and document data for analysis.* OSHA provided field staff with software for use onsite when conducting the recordkeeping reviews and collecting information (e.g., for audits, the tool tracks the steps in the recordkeeping review as established in the protocol and provides an integrated statistical sampling routine to determine the number of employees at the establishment for which to review injury/illness records). For both programs, the software supported conducting and documenting a recordkeeping review as well as aggregating and analyzing the results across establishments.

- ***Continuous improvement.***
  - *Analyzed findings with feedback loop for next iteration:* Following each annual cycle of the RK Audit Program, independent-contractor analysis of the audits provided a nationwide estimate of overall employer recordkeeping accuracy based on the sampling universe, a comparison of the estimate with prior years, and a case-level assessment of types of recordable cases and recording errors found by auditors. In addition, each year the reports included aggregated results from a comparison of employer summary injury/illness recordkeeping required to be maintained onsite with the same information submitted to OSHA under the ODI; also, a protocol implementation review resulted in ongoing refinements to the audit procedures, methodologies, and software. Similarly, an analysis was conducted at the conclusion of the RK NEP.
  - *Defined criteria to select establishments for records review based on underreporting hypotheses and then results:* The RK NEP was implemented with an initial and then revised criteria for selecting inspection establishments, with both approaches based on the assumption that underrecording is most likely to exist in lower-rate establishments operating in high-rate industries. While initial implementation targeted establishments with notably low rates despite operating in high-rate industries, a revision adjusted targeting to focus on establishments with a reported rate just below the cut-off used for OSHA's annual Site Specific Targeting (SST) Program. (In contrast, targeting for the RK Audit Program involved drawing a sample stratified by industry, establishment size, and OSHA Region of establishments subject to the OSHA Data Initiative collection.)
  - *Integrated enforcement with alternative approaches (compliance assistance and informational outreach):* OSHA conducts recordkeeping presentations at the regional and local level, generally focused on larger industry groups or unions. Also, OSHA's website includes an entire section on injury and illness recordkeeping, making extensive guidance information and tools readily accessible to the user. Further, OSHA's On-site Consultation Program provides free and confidential guidance to small and medium-sized businesses nationwide; although not its focus, the program does provide recordkeeping assistance and training as needed.

6.2 Use of Medical Records for Monitoring Injury and Illness Reporting

Interviews with MSHA staff that perform Part 50 audits indicate that the use of medical records may be helpful in determining the extent of underreporting; however, there are several limitations. Under The Health Insurance Portability and Accountability Act of 1996 (HIPAA), MSHA auditors have the ability to request and obtain medical records for mine employees. Medical records can provide auditors with additional information such as the date of an accident, the nature or severity of an injury, or clues about the time that the miner was away from regular work duties due to the type of injury or number of follow-up doctor visits noted. There are, however, limitations to the use of medical records in helping determine levels of reporting at the mine level, including:

- Medical records are not consistently reviewed by MSHA personnel performing Part 50 audits (See Section 2.4.2, Audit Comparison Data Sources).

- Medical records may be time consuming to obtain relative to the time under which the Part 50 audit is to occur. Auditors may request that medical records, along with other necessary audit information be present upon their arrival at the mine, but there may be reluctance of the part of mine staff to supply this information, and therefore, the auditor may communicate directly with the hospital or medical facility to obtain the records.

Medical records may not fully capture mining injuries or their severity. Medical records may not attribute work-relatedness for all mining injuries, if the miner seeks medical attention after work or is not taken by company ambulance to be treated. Some miners interviewed also noted that the company-specified doctor that they are required to see downplays the severity of their injuries so that the miner can return to work earlier than for a more severe injury or so that the company can reduce the amount of its workers' compensation claim. Therefore, the severity of the miner's injury might not be accurately captured on the medical record.

6.3 Use of Trauma Registries for Monitoring Injury and Illness Reporting

A summary of ERG's research and evaluation of trauma registry data for the six coal and metal/non-metal mining states with the highest numbers of reported injuries is presented in Table 19. Of the 11 trauma registries reviewed for top coal and metal/non-metal mining states:

- Data are publically available for 4 states, and possibly obtainable for the remaining 7 states.
- Occupational data are included in four of the registries and might be available in an additional four registries.
- Texas and Utah are two states that have publically available data and also include occupation in their data.

Despite the availability of data from registries, there are some limitations of these data:[62]

- Although many of the registries provide industry and/or occupation information, those data fields cannot be relied on to be coded consistently across cases.
- Although some of the registries indicate if the trauma was work-related, information on work-relatedness is generally not reliable. That is, there may be multiple cases that are work-related, but are not designated as such in the registry. Furthermore, there is no guarantee that the cases designated as "work-related" in a trauma registry would also meet the Part 50 designation of work-related.

Based on these limitations, ERG would not recommend the use of trauma registries for surveillance purposes.

[62] ERG uncovered these limitations while attending the National Occupational Injury Research Symposium (NOIRS) in October of 2011; ERG attended sessions that dealt with trauma registries and spoke with researchers that have experience with registries. The conference proceedings can be found at http://www.cdc.gov/niosh/noirs/2011/pdfs/TheNOIRS2011AgendaFINALSept72011.pdf.

Table 19: Overview of Trauma Registries for Coal and Metal/Non-Metal Mining States with the Most Reported Mining Injuries

ACS Verified Hospital	Coal or Metal/Non-Metal	Data Include Occupation? Y/N	Data Information	Data Available? Y/N/Maybe
ALABAMA	Coal	Maybe	Data collection began in 2000, and it appears that head and spinal cord injuries are required to be reported. 2010 report lists head/spine injuries as occurring in farm/mine	Maybe.
ARIZONA	MNM	No, but lists place of injury occurrence and notes mines	2008 report shows tracking of injuries occuring on mine site; injury diagnosis; injury severity, length of stay	Yes
CALIFORNIA	MNM	Yes	CA's data dictionary closely mirrors NTDS (nxt tab) and includes occupation (mining code included)	Maybe
COLORADO*	Coal?	Yes	Data lumps construction/extraction occupations in data.	Maybe
ILLINOIS	Coal	No	Data can be queried on web page. Data dictionary includes: Age; Gender; Race/Ethnicity; Cause of Injury; Prehospital Transport; Payor; Discharge Status; Time Period;EMS Region.	Yes
KENTUCKY	Coal	Maybe	Trauma registry data include hospital name, patient gender, sex, birthday, race, county of injury and	Maybe/No
MISSOURI	MNM		Only see head/spinal cord registry.	
NEVADA	MNM	Maybe	Included (but not limited to) are data on the event causing the injury, severity of the injury, place of the injury, length of hospital stay, diagnosis(es) of the patient, discharge destination of the patient and payer source.	Maybe
PENNSYLVANIA (No ACS verified facilities)	Coal/ MNM	No/Maybe	Data dictionary does not require/list occupation	Maybe
TEXAS	MNM	Yes	Data dictionary includes "place of injury" and "mine/quarry" is an option.	Yes
UTAH*	MNM	Yes	Includes occupation: "natural resources/mining".	Yes
VIRGINIA	Coal	No, but place of occurrence is listed	Included is E code 849.2 "Place of occurrence, Mine And Quarry"	Yes/Maybe
WEST VIRGINIA	Coal		Having trouble finding data details	-

* Indicates states that were not a top three state for reported injuries but had potential data available.

7.0 CONCLUSIONS AND RECOMMENDATIONS

This section provides ERG's conclusions and recommendations based on the research we performed under this project. We organized our discussion around the three priority evaluation questions and their associated sub-topics.

7.1 Priority Evaluation Question #1: Extent of Underreporting

Priority Evaluation Question #1 is primarily concerned with the extent of underreporting and aspects of and underreporting (e.g., characteristics of mines, trends, usefulness of Part 50 for accurately reporting injury rate). The sub-topics also explores whether other data sources are consistent with Part 50.

Priority Evaluation Question #1: *To what extent is there underreporting of injuries and illnesses under Part 50? Is underreporting concentrated to specific types of mines and operators or does underreporting occur across the mining industry? Are particular types of injury more likely to be inaccurately reported?*

Related Sub-Topics

- *Characterizing mines and operations where underreporting occurs, type of injuries more likely to be underreported, and implications for targeting.*
- *Determining consistency of size of underreporting over time, and for subgroups.*
- *Assessing data consistency among state worker compensation agencies, hospitals, and others with reported Part 50 data.*
- *Considering feasibility for MSHA to accurately estimate injury rates based on the level reported under Part 50.*

Extent of underreporting. ERG reviewed data from MSHA's own inspection and Part 50 audit activities and performed an extensive matching exercise between Kentucky and California workers' compensation data and MSHA Part 50 data. In reviewing MSHA's inspection and Part 50 audit activities, we found that those activities alone indicate underreporting on the 5 – 9 percent range. The workers' compensation to Part 50 matching, however, shows more extensive underreporting. For Kentucky, we concluded that underreporting could be in the 23 – 33 percent range and for California in the 32 – 46 percent range.

Types of mines. Unfortunately, the workers' compensation to Part 50 matching process cannot provide us with insights into the types of mines that underreport since the match looks for missing worker's compensation cases among the Part 50 and the workers' compensation data does not contain information on mines types. Instead, we must rely on data from MSHA inspections and Part 50 audits. Data from Table 3 and Table 6 above combined with a separate calculation for the rate of violations per 1,000 workers were used to create the data summarized in Table 20. These data indicate that inspections appear to find close to the same rate of violations across mine and mineral types. Part 50 audits, however, appear to find different rates among mine and mineral types. Based on the results from Part 50 audits, coal mines appear to have larger violation rates per 1,000 workers compared to metal non-metal mines. Furthermore, facilities have higher violation rates per 1,000 workers compared to either underground or surface mines and underground mines have a higher rate compared to surface mines.

Types of injuries. In assessing the types of injuries that have higher underreporting rates, it is important to note that it was necessary to limit the workers' compensation to Part 50 matching to cases that involved days away from work. There are two ways to view injury types: causes and body parts affected:

- Causes. For the Kentucky analysis, we found that the causes that were the least common in the data were least likely to find matches in the Part 50 data; this included burns/scalds, cuts/punctures/scrapes, and an "other" category.[63] For the California analysis we found that burns/scalds, motor vehicle-related incidents,[64] and falls/slips/trips were the most commonly underreported.
- Body part affected. In Kentucky we found that trunks and internal organs, head and neck, and other/unknown injuries were the most commonly underreported. In California we found that head and neck, multiple and unclassified, and trunk and internal organ injuries were the least likely to find a match.

Table 20 – Rates of Violations by Mineral Type and Mine Type for Inspections and Part 50 Audits Conducted between 2000 and July 23, 2012

Mineral Type or Mine Type	Inspections			Part 50 Audits		
	Number of Inspections	Number of Part 50.20 Violations	Rate of Violations per 1,000 Workers [a]	Number of Audits	Number of Part 50.20 Violations	Rate of Violations per 1,000 Workers [a]
Coal	302,855	2,913	0.68	1,399	1,083	49.36
Metal Non-Metal	405,252	3,540	0.66	10,719	1,540	7.18
Facility	67,213	866	0.74	853	463	42.45
Surface	431,802	3,197	0.70	10,158	1,187	7.97
Underground	208,897	2,390	0.59	1,105	973	26.03

[a] These columns are calculated by dividing the number of violations by the number of employees at each inspected or audited mine and multiplying by 1,000.

Over time. The data we have from both MSHA's inspections and Part 50 audits and from the workers' compensation matching analysis provides data over time. Table 8 shows no discernible trend in the number of violations found during MSHA inspections and Part 50 audits over time. For the WC analysis, however, we see a slight downward trend in Kentucky from 2005 to 2009 (Figure 4), but no such trend in California (Figure 7). Given that we only reviewed two states in the WC analysis and only one showed a trend, we cannot conclude there has been any trend over time in underreporting.

Data consistency. The sub-topics asked ERG to assess whether data sources such as workers' compensation data and hospital data are consistent with MSHA Part 50. Workers' compensation data are the most consistent with Part 50. However, even workers' compensation data has issues with consistency with Part 50. Most importantly, workers' compensation data is coded slightly differently in terms of causes of injuries and body part affected. In our matching, we had to adjust for this by re-coding the workers' compensation data to be consistent with the Part 50 data. Additionally, we had to develop matching protocols (see Sections 4.1.2 and 4.2.2) to match workers' compensation data to Part 50. The need to develop these protocols indicates that work is needed to ensure consistency before matching.

[63] The "other" category included such causes as Absorption Ingestion or Inhalation Not Otherwise Classified, Foreign Body in Eye, and Other Injury (Not Otherwise Classified).
[64] Motor vehicle incidents, however, may occur off the mine site and may not be subject to reporting under Part 50.

7.2 Priority Evaluation Question #2: Improving MSHA Part 50 Audit Processes

Priority Evaluation Question #2: *How could MSHA's current Part 50 Audit process be made more effective at capturing accurate injury and illness information?*

Related Sub-Topics

- *Assessing possible correlations between accuracy in reporting injury data and other measures of mine safety violations.*
- *Suggestions for identifying mines at high risk of underreporting for audits,*
- *Suggested activities to enhance the quality and comprehensiveness of the audits.*
- *Determining if current methods for analyzing Part 50 data are sufficient to identify areas of inaccuracies.*
- *Identifying incentives or disincentives involved in the reporting process.*

The second priority evaluation question focuses on ways in which Part 50 audits can be improved to be more effective at capturing accurate information. Our work under this project involved a process analysis of the Part 50 reporting process, including the way in which MSHA conducted Part 50 audits (see Section 5.3). During our project work, MSHA has implemented an extensive PPOV audit process. As noted in Section 5.3, the PPOV audits are extensive and time consuming, as well as productive at finding violations.[65] As part of this process, MSHA has also implemented more extensive training. Thus, ***ERG recommends that MSHA continue to implement the procedures and training it has developed and is using as part of the PPOV audits in improving the Part 50 audits.***[66] ERG also recognizes that MSHA has already implemented the spirit of this recommendation on its own.

Nevertheless, priority evaluation question #2 does not just cover Part 50 audit procedures and processes. The question is broader and covers how best to select mines for audit. The question and its sub-topics point to two objectives. First, the question asks how Part 50 audits can be used in "capturing accurate injury and illness information." This includes the use of the data from the audits to provide an accurate national estimate of injuries and illnesses in mining. The best way to accomplish that would be to perform a set of random audits. Second, the sub-topics ask ERG to identify factors that can be used to better target Part 50 audits. Targeting, however, is not best accomplished through random audits, but through taking into account which mines are more likely to have violations. In what follows, we make a recommendation on the numbers of random audits that are needed to provide accurate injury and illness information from the audits. We also make a recommendation on the factors to consider in targeting audits based on a statistical analysis of violations. That is, MSHA should conduct both a set of random audits (Section 7.2.1) and should conduct a set of targeted audits (Section 7.2.2).

7.2.1 Random Audits

ERG recommends that MSHA perform a set of random audits to capture accurate injury and illness information from mines. The number of audits to be performed will depend on (1) the desired accuracy MSHA wants to obtain, (2) the categories at which MSHA wants to obtain accurate counts, and

[65] Table 4 indicates that PPOV audits find almost six 30 CFR 50.20 violations per audit that finds violations while the comparable number for Part 50 audits is only 3 (per audit that finds a violation).

[66] Given that MSHA has evolved its approach to audits over the course of our project work by implementing the PPOV process, ERG's role in answering this question has changed over the project. In short, it is not possible for ERG to evaluate and assess a moving target.

(3) the resources that MSHA can devote to random audits. ERG can provide guidance on the first two while the third is clearly a decision for MSHA to make. Nevertheless, the guidance we provide on the first two should assist MSHA in how to allocate resources to random audits. To calculate the number of random audits to conduct, ERG relied on statistical sampling theory. We used a standard sampling scenario that involves selecting a sample size to meet a specified confidence level (e.g., 95 percent) and precision (i.e., length of the confidence interval measured in number of injuries). We used 2011 Part 50 data as a basis for population parameters (sample means and standard deviations). Details on the sample size calculations are provided in Appendix F.

ERG calculated sample size ranges for six canvass codes used by MSHA: anthracite coal, bituminous coal, sand and gravel, stone, non-metal, and metal. That is, *we are recommending that MSHA select random samples from each group separately*. To calculate the sample sizes, ERG used a basic sample size calculation for estimating a mean value from a sample for a continuous variable.[67] The purpose of the calculation was to determine the number of audits necessary to detect a number of underreported injuries as being a statistically significant level of underreporting. For example, in bituminous coal, how many audits are needed to determine if there are 500 underreported injuries in that canvass code? Naturally, MSHA does not need to find 500 underreported injuries among the audits (which are a sample). What we are looking for is whether the mean number of injuries per audit in the sample would indicate whether the total number of injuries among all bituminous coal mines is 500 injuries greater than the reported amount. In this case, an estimate of the total number of injuries among the population is the sample mean (injuries per audit) multiplied by the *total* number (audited and non-audited) of mines in the bituminous coal canvass code.

Table 21 to Table 26 provide the estimated sample sizes needed to detect different levels of underreporting among the mines in the six canvass codes. As discussed in Appendix F, it was not possible to use a uniform set of underreporting levels since the sample sizes are dependent on population means and standard deviations and the six canvass codes have varying values for these parameters. In general, the first few sample sizes (number of audits) reflecting relatively small levels of underreporting for each canvass are probably infeasible amounts for MSHA to undertake. We present these values, however, as an indication of how quickly sample size increases with precision. That is, to detect small levels of underreporting as being statistically significant, MSHA would need to expend significant resources.

Guidance on using these sample size estimates. Suppose MSHA decides to conduct 30 random audits among sand and gravel mines (see Table 23) and that these audits result in finding 9 injuries for an average of 0.3 injuries per mine. For this example, we will use the 2011 Part 50 data to project to the population; from Table F-1 of Appendix F there are 6,525 sand and gravel mines. This means the estimated population total for injuries is 1,958 (= 6,525 ×0.146) with a 95 percent confidence interval of 1,458 to 2,458. That is, MSHA can be 95 percent certain that that true number of injuries in coal mining is somewhere between 1,450 and 2,450.[68] MSHA can also conclude there was significant underreporting in this sector since the number of injuries in the population (654; see Appendix F, Table F-1) is well below the lower bound of the confidence interval. That is, if there was no (or minimal) underreporting in this sector, then finding 9 injuries among 30 audits would highly unlikely. Since 9 injuries were found, MSHA can conclude there must be significant underreporting. One issue is that the confidence interval

[67] https://www.dssresearch.com/knowledgecenter/toolkitcalculators/samplesizecalculators.aspx.
[68] For simplicity, ERG rounded these values.

above is 1,000 injuries in length (i.e., from 1,450 to 2,450). Naturally, a larger number of random audits leads to a smaller confidence interval.

Table 21 – Number of Audits Needed to Detect Statistically Significant Differences at 95 Percent Confidence, Anthracite Coal Canvass Code

Number of Underreported Injuries to Detect as Statistically Significant	Number of Audits Needed
10	137
20	60
30	32
40	20
50	14
60	10
70	7
80	6

Table 22 – Number of Audits Needed to Detect Statistically Significant Differences at 95 Percent Confidence, Bituminous Coal Canvass Code

Number of Underreported Injuries to Detect as Statistically Significant	Number of Audits Needed
500	229
600	166
700	127
800	100
900	81
1000	67
1100	57
1200	48

Table 23 – Number of Audits Needed to Detect Statistically Significant Differences at 95 Percent Confidence, Sand and Gravel Canvass Code

Number of Underreported Injuries to Detect as Statistically Significant	Number of Audits Needed
100	635
200	176
300	81
400	46
500	30
600	21
700	16
800	12

Table 24 – Number of Audits Needed to Detect Statistically Significant Differences at 95 Percent Confidence, Stone Canvass Code

Number of Underreported Injuries to Detect as Statistically Significant	Number of Audits Needed
100	1310
200	432
300	207
400	121
500	80
600	57
700	42
800	33

Table 25 – Number of Audits Needed to Detect Statistically Significant Differences at 95 Percent Confidence, Non-Metal Canvass Code

Number of Underreported Injuries to Detect as Statistically Significant	Number of Audits Needed
50	268
100	104
150	53
200	33
250	23
300	16
350	12
400	10

Table 26 – Number of Audits Needed to Detect Statistically Significant Differences at 95 Percent Confidence, Metal Canvass Code

Number of Underreported Injuries to Detect as Statistically Significant	Number of Audits Needed
100	201
200	103
300	60
400	39
500	28
600	21
700	17
800	10

7.2.2 Targeting Guidance

ERG is providing some guidance on the factors to take into account when targeting Part 50 audits. To determine what factors that MSHA should consider when targeting its actions, ERG used MSHA data on violations combined with data on mines where the violations occurred. These data were provided to ERG as part of this project. ERG used data from calendar years 2009 – 2011 on numbers of

Part 50.20 violations found during Part 50 audits[69] combined with descriptive information on the mines over those same years. The available information on the mines included information on:

- Type of mineral (coal, metal, nonmetal, sand and gravel, stone)[70]
- Type of mine (underground, surface, facility)
- Whether the mine was a portable operation
- If the mine had completed its Part 48 training requirement
- Whether or not the mine had a safety committee
- The number of days the mine operated per week
- The number of shifts per day
- The number of employees at the mine
- The number of hour worked at the mine

ERG estimated nonlinear regression models to determine how these different characteristics relate to the number of violations found by MSHA during Part 50 audits. Appendix G provides the estimated regression models. The results indicate the following regarding which types of mines are more likely to be found in violation of Part 50.20:

- The data indicate that coal mines have more violations than metal non-metal mines.
- Among the metal non-metal mines, metal mines tend to have the most violations followed by sand and gravel, stone, and then non-metal mines.
- In terms of mine type, facilities are the most likely to have violations followed by underground mines and then surface mines. In fact, the regression models indicate that both facilities and underground mines have significantly more violations than surface mines.
- There is weak evidence that mines that have completed their Part 48 training are more likely to have violations.
- Mines with a larger number of employees and that work more hours also have more violations. Additionally, mines where the average number of hours per employee is high also have more violations.

MSHA should use these considerations in selecting which mines to select for non-random Part 50 audits. ERG is not recommending, however, using the equations we estimated and reported in Appendix G to prioritize among mines for targeting. Rather, MSHA should use its best professional judgment in selecting mines for audit, but should consider the factors we have discussed above when determining which mines to audit.

7.2.3 Summary: Random and Targeted Audits

As noted in the introduction, ***ERG is recommending that MSHA conduct both random audits and targeted Part 50 audits***. We are not, however, recommending *how* MSHA allocate its auditing resources between the two sets of audits. Nevertheless, MSHA should make an effort to perform both types of audits. Random audits allow MSHA to develop an estimate of injuries and illnesses in the

[69] These are the same data that were tabulated in Section 3 above. However, ERG limited its analysis to only calendar years 2009 to 2011 in the analyses we conducted here.
[70] MSHA refers to these as the canvass for the mine.

industry with known accuracy (the confidence interval) and confidence (95 percent) while the targeted audits allow MSHA to target mines where violations are more likely to be found. MSHA should clearly denote which audits are part of the random set and which are part of the targeted set in its data. ***Finally, we recommend that MSHA conduct the random audits using the PPOV approach and protocols to ensure that the audits are conducted with rigor.*** ERG further recognizes that MSHA must also balance the random and targeted audits with the need to conduct PPOV audits, as well as within its larger statutory requirement to inspect all mines two or four times annually.

ERG also recommends that MSHA perform detailed analysis of the random audits to better target future audits. The types of violations found and the characteristics of mines should be analyzed to determine what types of mines should be targeted in the future. MSHA does not need to perform the type of regression analyses that we performed in Section 7.2.2. MSHA should review tabulations and cross-tabulations of the data to determine trends in violations. This should help MSHA target better as it moves forward since the set of mines selected is random.

7.3 Priority Evaluation Question #3: Other Strategies

The third question asks ERG to identify other strategies that could be implemented, in addition to Part 50 audits, to insure more accurate reporting of injuries. This involves assessing what has been done in other sectors to improve reporting of injuries and illnesses. The related sub-topic asks whether there are proxies that could be used to track injury and illness data over time.

Priority Evaluation Question #3: *Are there other strategies that could be implemented, in addition to the Part 50 Audits, to insure more accurate and complete reporting of accidents, injuries and illnesses for the purposes of monitoring? What techniques have been used in other sectors to improve reporting of workplace injuries and illnesses?*

Related Sub-Topics

- *Determining proxies for injury and illness data which could be used to track injury and illness trends over time.*

Other strategies. ERG has recommended above that MSHA conduct both random and targeted audits. MSHA has also been conducting more intensive PPOV audits recently. ERG expects that an opportunity exists to use the strategies employed in the PPOV audits to improve the Part 50 audits. ***ERG recommends that MSHA institute some form of a best practices dissemination so those who conduct the audits can learn from one another.*** One potential form of this would be an annual or semi-annual meeting (webinar or in-person) where auditors come together to discuss best practices in identifying reporting violations. These discussions would identify the most effective data sources to use, what things to look for in finding unreported injuries and illnesses, disseminate information on what strategies operators may be employing to knowingly conceal injuries and illnesses. Furthermore, having some inspectors sit in on these sessions would assist inspectors who do not conduct audits in finding underreporting during regular inspections.

Other sectors. ***ERG recommends that MSHA review the research that ERG has provided for OSHA to assess whether any of the aspects that OSHA performs should be adopted***. Our own assessment of OSHA's approaches did not reveal any specific practices that OSHA uses that would substantially improve on MSHA's approach to verifying the accuracy of injury and illness reporting. For one, at the time we performed our research, OSHA was itself undergoing a review of its procedures. Nevertheless, ERG has provided detailed information on OSHA which MSHA should consider.

Proxies. ERG reviewed a number of potential proxies for MSHA and found that none would be more reliable than its current data. Furthermore, we feel that our recommendation to perform random Part 50 audits would result in reliable estimates of the number of injuries and illnesses in the industry. ERG reviewed trauma registries and ambulance/medical records as potential proxies. However, neither of these are available on a wide-spread basis and data from different states are likely to be inconsistent (see Sections 6.2 and 6.3). Furthermore, ERG used workers' compensation data in our analyses (see Section 4.0) to assess underreporting. ERG does not feel that workers' compensation data would be useful as a proxy since (1) it took significant effort to process these data to be consistent with MSHA Part 50 reporting requirements and (2) it was a significant effort to obtain the data and only two states allowed access to their data.