Mining, Metallurgy & Exploration
https://doi.org/10.1007/s42461-020-00347-x

Identifying Risk Factors from MSHA Accidents and Injury Data Using Logistic Regression

Richard Amoako[1] · Judith Buaba[1] · Andrea Brickey[1]

Received: 1 May 2020 / Accepted: 23 October 2020

Abstract
The global mining industry has recorded significant declines in accident and injury rates attributed to the advancement in technology, increased enforcement, and safety consciousness. A goal of the mining industry is to achieve zero injury and occupational illness on all mine sites, prompting increased research into ways to further reduce mine accidents. A machine learning technique known as multiclass logistic regression is applied on a 10-year injury dataset from the Mine Safety and Health Administration (MSHA) to determine a miner's susceptibility to a class of injury and to help identify significant risk factors associated with different classes of injury. The data is aggregated based on injury classification to provide statistically relevant results. The analysis identifies specific risk factors that influence a mine worker's susceptibility to a given class of injury, i.e., non-fatal with no days lost or restricted activity, non-fatal with days lost and/or days of restricted work activity, and fatal and total permanent or partial permanent disability. These factors include miner's age, mine type (coal vs. non-coal), experience on the current job (years), shift start time, employment type (operator vs. contractor), mining district, and type of accident. The results of the analysis indicate that a miner's experience on the job, i.e., the number of years worked in a current job, is a significant risk to injury occurrence, even for those with decades of total mining experience. We further show the differences and similarities between the surface and underground mine incidents.

Keywords Mine accidents · Machine learning · Logistic regression · Mine fatalities

1 Introduction

Mine workers are exposed to a variety of hazards within the mining environment. These hazards include rock falls, equipment malfunctions, fires, explosions, and harmful gases, thereby increasing susceptibility to mine accidents; however, over the past three decades, the mining industry has recorded significant declines in injury occurrence [1, 2]. This is attributed to the increased enforcement, improved safety consciousness, and implementation of new technology [3]. The establishment of the mine safety legislation, the Federal Coal Mine Health and Safety Act of 1969, has resulted in a steady decline of mine accidents by institutionalizing an enforcement agency to implement codes and regulations. In the late 1800s and early 1900s, the mining industry recorded thousands of fatalities annually [4]. This led to the insistence of increased safety regulations by society and the government. In recent decades, mine safety has become an integral part of the mining industry culture through increased safety training and hazard recognition. At present, the United States (US) Mine Safety and Health Administration (MSHA) enforces the health and safety rules outlined in the Federal Mine Safety and Health Act of 1977, as amended by the MINER Act of 2006. The National Institute for Occupational Safety and Health (NIOSH) works to develop innovative safety solutions through the provision of research grants with the focus of reducing safety and health-related accidents. Between research-focused NIOSH and regulation enforcement by MSHA, the USA has seen a steady decline in mine-related accidents. In 2019, 24 fatalities were reported for all mining sectors in the USA, one of the lowest numbers ever recorded [5]. Nieto and Duerksen [6] summarize the effects of mine safety legislation on the US mining industry from 1872 to 2007.

✉ Andrea Brickey
Andrea.Brickey@sdsmt.edu

Richard Amoako
Richard.Amoako@mines.sdsmt.edu

Judith Buaba
Judith.Buaba@mines.sdsmt.edu

[1] Mining Engineering and Management Department, South Dakota School of Mines and Technology, Rapid City, SD, USA

MSHA requires mine operators under their jurisdiction to disclose all reportable accidents, occupational injuries, and illnesses that occur on their mine site. Information on the work location, accident classification, body part injured, degree of injury, and occupation are reported by the principal officer responsible for health and safety in the mine or the supervisor of the mine area where the injury occurred. Other information on the age and mining experience of the miner, work shift time, and other statistics are also reported. The injuries reported to MSHA are categorized into ten degrees [7] as shown in Table 1. This study examines injured miners involved in a single accident while working on a mine site. The first six degrees of injury, outlined in Table 1, fall within the scope of this study. To achieve statistical relevance, we aggregate the relevant six degrees of injury into three classes, listed in Table 2. Fatalities and permanent disability (total or partial) only make up 0.3% and 1.2% of the final dataset used in modeling, respectively. Aggregating them into a common injury class provides sufficient data to enable validation of the model at a 95% confidence level. Aside from statistical relevance, the aggregation also reflects injury severity among the three injury classes, with fatality and permanent disability (FP) representing the highest form of severity, followed by injuries with days lost and/or days of restricted work activity (DLR) and injuries with no days lost or restricted activity (NDLR). Setting NDLR as a reference provides reasonable grounds for comparison among the three classes of injury. This injury classification differs from MSHA's injury classification, i.e., fatal, non-fatal with days lost and/or restricted activity, and non-fatal with no days lost. In this study, permanent disability is moved from non-fatal with days lost and/or restricted activity and aggregated with fatality to form a new injury class, fatality and total permanent or partial permanent disability (FP).

Many mining companies have set the goal of achieving zero harm. To this end, researchers have analyzed accidents and injury data with a goal of mitigating mine accident and injury occurrence using a variety of statistical, quantitative, and novel methods, among which includes machine learning (ML). In this era of "big data," ML techniques have been employed to derive valuable information implicit in data. ML methods have been applied in various fields such as engineering, finance, transportation, and medicine. As a branch of artificial intelligence (AI), ML allows computer systems to improve their performance at a task through experience (learning) for the purpose of predicting future outcomes [8]. It is a multidisciplinary field that includes probability and statistics, control theory, and computational complexity, among others. In general, ML is classified into two categories: supervised and unsupervised learning. In supervised learning, the relationship between the input and output data is known. Thus, there is foreknowledge (an estimation) of the output value. Here, the goal is to predict an outcome given a set of input data. Regression and classification methods fall under this category. Unsupervised learning, on the other hand, provides little or no information on the outcome. Its purpose is to discover patterns and trends that exist in a given input data. In mining, ML methods such as logistic regression have been applied to predict the presence or absence of gold mineralization in a deposit [9] to predicting roof fall risks in underground coal mines [10].

Table 1 Degrees of injury

Degree	Injury	Description
1	Fatal	Injury occurrences resulting in the death of the miner
2	Permanent total or permanent partial disability	Injury occurrences resulting in loss or complete loss of use of any member or part of a member of the body
3	Non-fatal with days lost only	Injury occurrences not resulting in death or permanent disability but requires days away from work to recover from injury
4	Non-fatal with days lost and days of restricted work activity	Injury occurrences not resulting in death or permanent disability but requires days away from work to recover from injury and upon recovery might require the miner to be assigned to another job other than his primary job temporarily
5	Non-fatal with restricted work activity only	Injury occurrence not resulting in permanent disability or days lost but the miner is assigned to another job other than his primary job temporarily
6	Non-fatal with no days lost or restricted activity	Injury occurrences resulting only in loss of consciousness or medical treatment other than first aid
7	Occupational illness	Cases not caused by a single accident occurrence such as hearing loss, pneumoconiosis, silicosis, hepatitis and cancer.
8	Fatal and non-fatal cases due to natural causes to employees on company business	Cases not caused by a single accident occurrence such as hearing loss, pneumoconiosis, silicosis, hepatitis, and cancer
9	Fatal and non-fatal injuries involving non-employees on or off the mine property	Cases such as heart attacks and strokes
10	All other cases including first aid	Injury occurrence resulting in death or permanent disability reported to MSHA which do not result in charges or citations to the company

Table 2 Injury class

Degrees of injury	Definition	Abbreviation
1 and 2	Fatal and total permanent or partial permanent disability	FP
3, 4, and 5	Non-fatal with days lost and/or days of restricted work activity	DLR
6	Non-fatal with no days lost or restricted activity	NDLR

Logistic regression (LR), a supervised machine learning technique, is applied in this study. LR is well suited for describing and testing hypotheses about relationships between a categorical outcome, i.e., response or dependent variable, and one or more categorical or continuous predictors, i.e., independent variables [11]. A set of independent variables (predictors) is needed to predict the dependent variable (outcome). The nature of the outcome variable dictates the type of logistic regression to use—binomial (binary), multinomial (multiclass), or ordinal. Binomial or binary logistic regression is used when the outcome variable can have only two categories, e.g., fatal vs. non-fatal. If the outcome variable has more than two categories in no order of priority, e.g., fatality, injury, and no injury, multinomial or multiclass logistic regression is used. If the outcome variable is ordered, e.g., severe, moderate, mild, and minimal, then ordinal logistic regression is used [11]. This study applies multiclass logistic regression to a 10-year (2008 to 2017) injury dataset from MSHA to help identify potential risk factors associated with different classes of injury. The multiclass logistic regression model is expressed as [11] follows:

$$\text{logit}\ (Y) = ln\frac{p}{1-p} = \beta_0 + \beta_1 X_1 \ldots + \beta_n X_n \tag{1}$$

where

Y	the dependent or outcome variable, i.e., fatality and no days lost
p	the probability that Y occurs
$X_1\ldots$	the independent or predictor variables, i.e., age,
X_n	work location, and mine type
β_0	intercept of the regression line on the y-axis
$\beta_1\ldots\beta_n$	corresponding regression coefficient of X_i

The regression coefficient values $\beta_1\ldots\beta_n$ indicate the relationship between $X_1\ldots X_n$ and logit of Y. A coefficient value greater than 0 indicates an increase in logit of Y with an increase in X, and a coefficient value less than 0 indicates a decrease in logit of Y with an increase in X. A coefficient value of 0 indicates there is no linear relationship between logit of Y and X. The Wald test is used to measure the significance of the independent variables to the logistic regression model [11]. Suppose we want to predict the mood (outcome) of a miner as happy or sad (outcome) given some independent variables (predictors) such as food, bad weather, and sleep. If the Wald test shows that the parameters for certain independent variables, such as food and sleep, are zero, it proves that there is no association between these predictors and the outcome; hence, they can be removed from the model. If the test shows the parameters are not zero, then the opposite holds true and those variables are included in the model.

2 Literature Review

Research investigating the factors associated with mine accidents using logistic regression has provided some insight into the occurrence of mine accidents. Bennet and Passmore [12] use multinomial logit analysis to examine the relationship among the mine and injured miner characteristics, and the degrees of injury in underground bituminous coal mines in the USA from 1975 to 1982. Their study reveals that injury severity varies by the mining system used in the mine, the geographical region of the mine, circumstances surrounding the injury occurrence, the injured miner's age, the location in the mine where the injury occurred, and whether the injured miner used powered haulage in the year the injury occurred. They further reveal that for their period of study, injury severity is not associated with the injured miner's total mining experience, experience on the job, and experience in the mine. Friedman et al. [3] study the injuries, i.e., fatal and non-fatal, associated with long working hours, i.e., more than 9 h in a shift, among mine workers in the USA using binary logistic regression to analyze MSHA data from 1983 to 2015. They identify job change, lack of work routine, small mining operations, and being new at the mine as risk factors associated with injuries occurring during long working hours. Muzaffar et al. [13] investigate the factors associated with fatal accidents among contractors and operators by applying binary logistic regression on MSHA data. The results show that a higher proportion of fatal injuries are associated with contractors, less mining experience at the current mine, and working more than 8 h per day. A multinomial logistic regression model used by Maiti and Bhattacherjee [14] examines the differences in accident susceptibility among various groups of underground coal mine workers, given their personal and workplace characteristics. The authors conclude that, among the occupation groups, face workers are more susceptible to injury than other groups, e.g., haulage. The analysis, however, considers only one type of mine—underground coal—and consists of data from five mines operating under the same company. Risk factors associated with injuries in artisanal and small-scale mining operations include gender, less working experience, long working hours, poor supervision, and job stress, among others [15].

A 2004 study by Chau et al. [16] examined the relationship between individual characteristics and occupational injuries for various jobs in the construction industry. Personal and work-related data from 880 male workers who have had at

Table 3 Initially examined (MSHA data) dependent and independent variables

Variables	Bins
Dependent variable	
Injury class	Fatal and total permanent or partial permanent disability (FP)
	Non-fatal with days lost and/or days of restricted work activity (DLR)
	Non-fatal with no days lost or restricted activity (NDLR)
Independent variables	
Mine type	Coal; non-coal
Work location	Surface; underground
Employment status	Apprentice or trainee; permanent
Employment type	Operator; contractor
Gender	Male; female
District	Multiple (6)
Accident type	Multiple (21)
Hours at work prior to injury occurrence	<9
	≥ 9
Age (years)	18–30
	31–40
	41–50
	51–70
Total mining experience (years)	0–1
	1–3
	3–6
	6–10
	10–20
	20–30
	>30
Experience at the current mine site (years)	0–1
	1–3
	3–6
	6–10
	10–20
	20–30
	>30
Experience on the current job (years)	0–1
	1–3
	3–6
	6–10
	10–20
	20–30
	>30
Shift start time	7 am
	3 pm
	11 pm

least one occupational injury during a 2-year period are analyzed using logistic regression. Results from this study show that young workers, workers with sleep disorders, and smokers are susceptible to occupational injuries. Zhang and Hassan [17] and Robin [18] analyze injury severity in work zone accidents using multinomial logistic regression in the construction industry. Both studies identify weather conditions (snow and rain) and gender and age group of the driver as risk factors. Akboga Kale and Baradan [19] identify factors contributing to the severity of construction injuries using a binary logistic regression model. Their findings show that work experience and accident type, among others, are significant factors affecting injury severity. In this study, we apply multiclass logistic regression on mine injury classifications, i.e., fatal and permanent disability, non-fatal with days lost and/or restricted activity, and non-fatal with no days lost and/or restricted activity, from the MSHA database to provide in-depth analysis on potential risk factors associated with different classes of injury in the mining industry. This study covers all mines required to report incidents to MSHA between the years 2008 to 2017.

3 Data and Methodology

Multiclass logistic regression is applied to accidents and injury data reported to MSHA from 2008 to 2017 to (i) determine a miner's susceptibility to a class of injury given characteristics such as age, mining experience, and work location and (ii) identify the various risk factors associated with each class of injury. We develop an analytics process by which safety managers can identify areas that require prioritized training or attention in helping advance the goal of zero injuries and fatalities in the mining industry.

The Federal Mine Safety and Health Act of 1977, Public Law 91-173, requires operators of mines in the USA to report occupational injuries and illnesses to MSHA [20]. The data is summarized by work location, mine type, accident classification, body part injured, nature of injury, and occupation of the miner. In this study, Part 50 accident, injury, and illness data files from 2008 to 2017 are downloaded from MSHA's website. The data format collected by MSHA is converted into the Statistical Package for the Social Sciences (SPSS) [21] file formats for data processing.

The study initially examines the following independent variables: (i) miner's age; (ii) gender; (iii) mine type (coal vs. non-coal); (iv) total mining experience (years); (v) experience at the current mine (years); (vi) experience on the current job (years); (vii) shift start time; (viii) hours at work prior to injury occurrence; (ix) employment type (operator vs. contractor); (x) employment status (permanent vs. trainee); (xi) work location (surface vs. underground); (xii) district; and (xiii) accident type. These variables have been categorized into subgroups. A description of the categories is presented in Table 3. We aggregate and categorize the degree of injury (dependent variable) into three classes (Table 3): fatal and total permanent or partial permanent disability (FP), non-fatal with days lost and/or days of restricted work activity (DLR), and non-fatal with no days lost or restricted activity (NDLR).

3.1 Data Cleaning and Validation

This study looks at mine workers, aged 18 to 70 years, who suffered occupational injuries while working on a mine site.

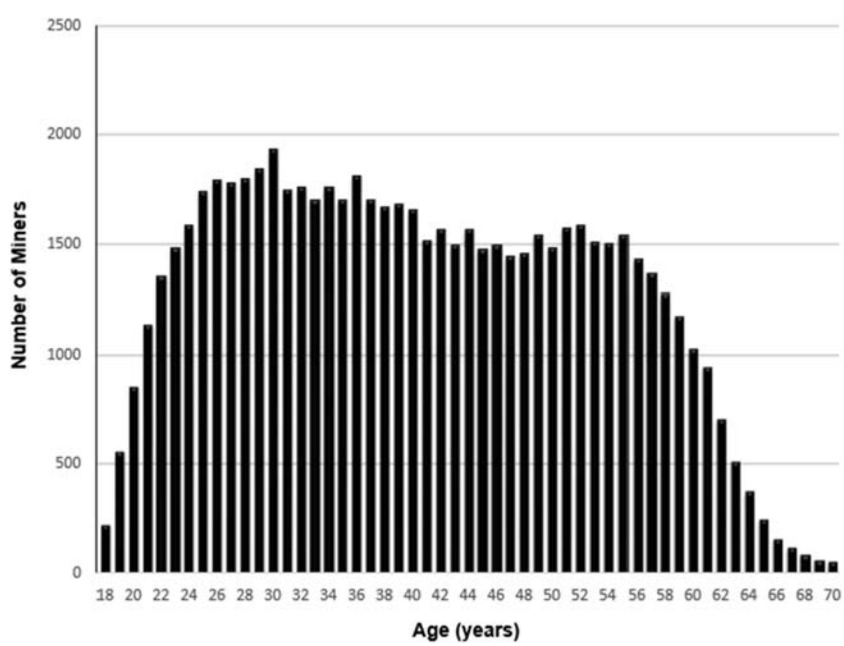

Fig. 1 Age distribution of injured miners (processed accident data)

MSHA reported a total of 96,834 incidents during the period 2008 to 2017. We exclude 6707 records (6.93% of total incidents) due to incidents involving minors, non-employees, and occupational illness not caused by a single accident. These occupational illnesses include cancer, hepatitis, lung diseases, hearing loss, and heart attacks. We further exclude 16,170 incidents with missing records such as age, total mining experience, and experience on the current job. This represents 16.70% of the total data. We perform preliminary analysis, hereafter, on the remaining 73,957 records (processed data), representing 76.39% of the original data.

3.2 Statistical Analysis and Modeling

The authors perform descriptive statistical analyses on the dataset using IBM SPSS software [21]. The independent variables are divided into categories. Age is divided into four categories, i.e., 18–30, 31–40, 41–50, and 51–70 years, to approximately reflect the distribution of age in the dataset. Total mining experience, experience on the current job, and experience at the current mine have similar data distribution; therefore, they are divided into similar categories. The shift start time shows a large number of miners starting work at 7 am, 3 pm, and 11 pm, yielding three categories. Figures 1, 2, and 3 show the distributions of the following variables: age, work location, and total mining experience (categorized). Figure 1 shows the age distribution of miners who were injured, ranging from 18 to 70 years. Figure 2 shows that over the past decade, more incidents occurred at surface work locations than underground. Tables 4 and 5 summarize statistics and distributions on the continuous and categorical independent variables, respectively. A preliminary step to justify the consideration of the independent variables involves conducting a two-tailed chi-squared contingency test. This test is performed on contingency tables to determine whether a significant relationship exists between the row and column variables of the table. A contingency table is a frequency table of two categorical variables where all the levels (categories) of one variable are listed as rows and the levels of the other variable as columns. The cells of the table are populated with the joint frequencies of the corresponding row and column levels of the two variables. The test verifies if the levels of the row variable are differentially distributed over the levels of the column variable. Obtaining a significant result means that a relationship exists between the two variables; i.e., one variable varies according to the other variable [22]. In this study, we conduct the test between the dependent variable and each independent variable. The results confirm that a significant association exists between the dependent variable and each independent variable. That is, the degree of injury (dependent variable) varies according to variations in the predictors (independent variables). This warrants further analysis to characterize how variations in the predictors influence injury occurrence.

Preliminary analysis reveals the major accidents that have occurred over the 10-year period under study. Figure 4 presents the ten most common accidents. For the purpose of achieving statistical relevance, our study focuses on the first six common accidents, being the ones with the most occurrences over the decade (each with at least 3000 cases). These include handling material, slip or fall, hand tools, machinery, powered haulage, and fall of roof, back, or brow. To this end, we exclude cases involving all other accidents, leaving a total of 67,580 cases for modeling. This represents 91.38% of the processed data.

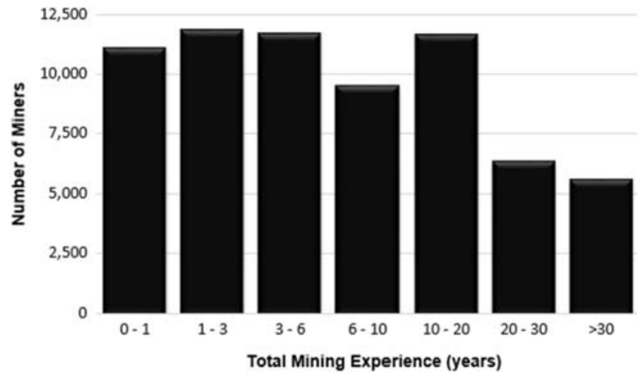

Fig. 3 Total mining experience of injured miners (processed accident data)

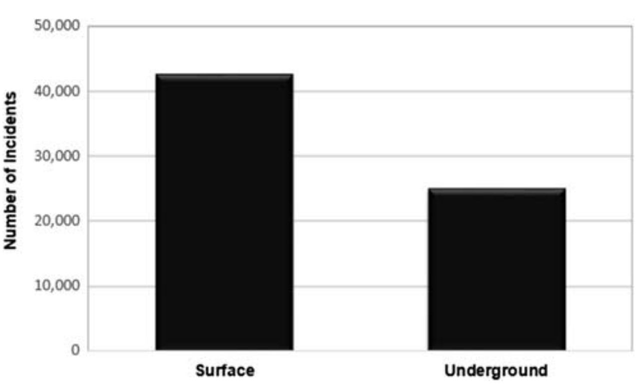

Fig. 2 Work location of injured miners (processed accident data)

Table 4 Summary statistics for continuous variables (processed accident data)

Variable	Mean	Standard deviation
Age (years)	40.38	12.31
Hours at work prior to injury occurrence	4.72	2.85
Total mining experience (years)	10.21	10.75
Experience at the current mine (years)	5.95	8.26
Experience at the current job (years)	6.62	8.41

Table 5 Data distribution for categorical variables (processed accident data)

Variable	Subgroup	Number of incidents	Percent (%)
Mine type	Coal	31,215	46.2
	Non-coal	36,365	53.8
Work location	Surface	42,581	63.0
	Underground	24,999	37.0
Employment type	Operator	59,570	88.1
	Contractor	8010	11.9
Employment status	Permanent	67,410	99.7
	Trainee	170	0.3
District	Multiple (6)	67,580	100.0
Accident type	Multiple (21)	67,580	100.0

We subject the independent variables to Pearson correlation analysis. The Pearson correlation measures the strength of the linear relationship between two continuous variables. It results in values ranging from + 1 to − 1; these are known as correlation coefficients. Obtaining a value of zero implies that there is no linear association between the two variables. A value greater than zero implies that the association between the two variables is positive, i.e., an increase in one variable is associated with an increase in the other variable. A value less than zero signifies a negative association between the two variables, i.e., an increase in one variable is associated with a decrease in the other variable. Values closer to + 1 indicate a strong positive relationship while values closer to − 1 indicate a strong negative relationship [23]. Results of the analysis show a strong positive relationship between pairs of the following variables: total mining experience, experience at the current mine, and experience on the current job. For each pair, we obtain a correlation coefficient in the range 0.59 to 0.71. In all cases, the correlation proves significant at a 1% level of significance. The condition where correlation exists between independent variables is termed multicollinearity. This condition could render the output of the logistic regression inaccurate. In order to avoid the potential effect of high multicollinearity, only one of these variables is considered for modeling, i.e., experience on the current job.

We further perform a multicollinearity test on experience on the current job and the remaining independent variables. Multicollinearity is measured by examining the variance inflation factor (VIF). This is the increase in the variance of an estimated regression coefficient if the independent variables are correlated. According to Minitab [24], if all the VIFs are equal to one, no

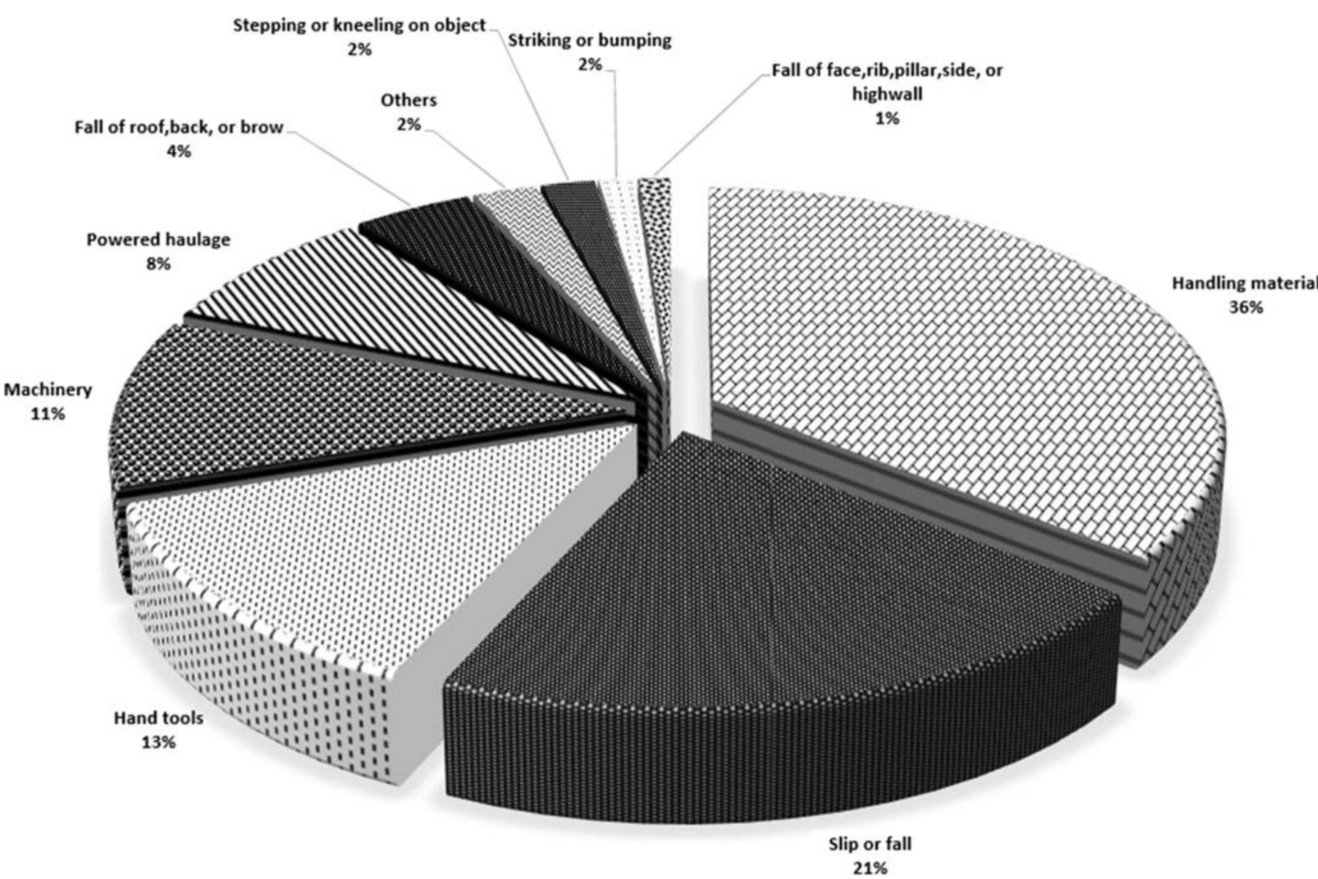

Fig. 4 Most frequent accidents (2008–2017)

multicollinearity exists; however, if some of the VIFs are > 1, the independent variables are correlated. Multicollinearity is of significant concern when the VIF of any of the independent variables is above (>) five. When this happens, the estimated regression coefficient may not be accurate. For this study, the VIFs for all the independent variables considered are less than 5 (between 1.01 and 2.05). This implies that there is no significant multicollinearity in the dataset.

4 Results

We develop the multiclass logistic regression model using IBM SPSS software [21]. The dataset used for modeling comprises a total of 67,580 incidents, representing valid records on the six major accidents. Gender and employment status variables are excluded from the model since they are highly skewed. Gender data consists of 97.4% males, and employment status comprises 99.7% permanent employees. The independent variables that are finally considered for modeling include (Table 6) miner's age, mine type (coal vs. non-coal), experience on the current job (years), shift start time, hours at work prior to injury occurrence, employment type (operator vs. contractor), work location (surface vs. underground), district, and type of accident. Degree of injury is the dependent variable in the final model, and it comprises the following categories (Table 6): fatal and total permanent or partial permanent disability (FP), non-fatal with days lost and/or days of restricted work activity (DLR), and non-fatal with no days lost or restricted activity (NDLR).

We validate results from the model and proceed with analysis to determine which independent variables have had a significant impact on injury occurrence over the decade under study. These variables are identified as risk factors and include miner's age, mine type (coal vs. non-coal), experience on the current job (years), shift start time, employment type (operator vs. contractor), mining district, and accident type. Using the odds

Table 6 Dependent and independent variables in the logistic regression model

Variables	Bins
Dependent variable	
Injury class	Fatal and total permanent or partial permanent disability (FP)
	Non-fatal with days lost and/or days of restricted work activity (DLR)
	Non-fatal with no days lost or restricted activity (NDLR)
Independent variables	
Mine type	Coal; non-coal
Work location	Surface; underground
Employment type	Operator; contractor
District	Multiple (6)
Accident type	Multiple (6)
Hours at work prior to injury occurrence	< 9
	≥ 9
Age (years)	18–30
	31–40
	41–50
	51–70
Experience on the current job (years)	0–1
	1–3
	3–6
	6–10
	10–20
	20–30
	> 30
Shift start time	7 am
	3 pm
	11 pm

Table 7 Significance of predictors

Significant ($p < 0.025$)	Not significant ($p \geq 0.025$)
District	Time until injury occurrence
Experience on the job (years)	Work location
Shift start time	
Age (years)	
Employment type	
Mine type	
Accident type	

ratio output of the model, we determine injury susceptibility among different groups of miners. We also conduct a comparative study between the surface and underground incidents; this is informed by the significant differences that exist between the surface and underground operations. For the surface incidents, we identify the following risk factors: miner's age, mine type (coal vs. non-coal), experience on the current job (years), shift start time, employment type (operator vs. contractor), mining district, and accident type. For underground incidents, we identify the following risk factors: miner's age, shift start time, mining district, and accident type.

4.1 Model Validation and Risk Factors

We validate the fitted model at a 5% level of significance using the likelihood-ratio chi-squared test. The purpose of the test is to determine if the model is a good fit for the data. Significance of the test statistic implies that the model fits the data well and that one or more of the predictors (independent variables) have a significant influence on the model. From the validation results, a value of $p = 0.00$ is obtained signifying that the model is a good fit for the data. We identify the specific predictors that have a significant impact on the model by conducting a series of the likelihood-ratio chi-squared test, focusing on each predictor at a time. At a 5% level of significance, all predictors prove significant apart from time until injury occurrence and work location. This means that these two variables have not had a significant impact on the three classes of injury over the study period. Table 7 summarizes the significant and insignificant predictors. The significant predictors are termed risk factors. We subsequently assess whether there is a significant difference in injury susceptibility between the different categories of each risk factor.

The output of the model also presents interactions among the three injury classes as follows: non-fatal with days lost and/or days of restricted work activity (DLR) vs. non-fatal with no days lost or restricted activity (NDLR) and fatal and total permanent or partial permanent disability (FP) vs. non-fatal with no days lost or restricted activity (NDLR). Thus, we set NDLR as a reference response variable to enable assessment of the relative impact the various predictors have on the injury classes. For DLR vs. NDLR, the results in Table 8 show how the established risk factors influence one's susceptibility to DLR, relative to NDLR (the reference injury class). The results in Table 9 show how the risk factors determine one's susceptibility to FP, relative to NDLR. A detailed discussion of these results follows in Section 4.2.

4.2 Injury Susceptibility

A means of assessing injury susceptibility among different groups of miners lies in understanding the odds and odds ratio. The odds is the ratio of the probability that an event will occur to the probability that the event will not occur. Suppose the numerical values of 0 and 1 are the outcomes of a binary event, where 1 represents event occurrence (success) and 0 represents event non-occurrence (failure). If p is the proportion of observations with an outcome of 1 (probability of success), then $1-p$ is the probability of an outcome of 0 (probability of failure). The ratio $\frac{p}{1-p}$ is called the odds [25]. It provides a sense of how many times an event is likely to succeed than it is to fail.

Suppose this binary event can be observed among two groups of miners, A and B, with the former being

Table 8 DLR vs. NDLR

Predictor	Category	Wald Statistic	Significance (p value)	Odds Ratio
District	**Northeast**	**62.64**	**0.00**	**1.32**
	Southeast	**5.68**	**0.02**	**1.09**
	North Central	4.57	0.03	0.92
	South central	**20.57**	**0.00**	**1.19**
	Rocky Mountain	**24.75**	**0.00**	**0.84**
	Western (Reference)			
Experience on the current job (years)	**0 - 1**	**42.95**	**0.00**	**1.45**
	1 - 3	**31.87**	**0.00**	**1.37**
	3 - 6	**30.18**	**0.00**	**1.37**
	6 - 10	**25.63**	**0.00**	**1.34**
	10 - 20	**12.48**	**0.00**	**1.22**
	20 - 30	1.68	0.20	1.08
	>30 (Reference)			
Shift	7 am	1.21	0.27	0.97
	3 pm	**8.27**	**0.00**	**1.10**
	11 pm (Reference)			
Age (years)	**18 - 30**	**84.19**	**0.00**	**0.77**
	31 - 40	**14.22**	**0.00**	**0.90**
	41 - 50	2.58	0.11	0.96
	51 – 70 (Reference)			
Accidents	**Handling material**	**6.84**	**0.01**	**1.12**
	Hand tools	**264.64**	**0.00**	**0.48**
	Powered haulage	**253.39**	**0.00**	**2.26**
	Machinery	**9.12**	**0.00**	**0.87**
	Slip/fall of a person	**530.58**	**0.00**	**2.83**
	Fall of roof, back, brow (Reference)			
Employment type	**Operator**	**53.58**	**0.00**	**1.21**
	Contractor (Reference)			
Mine type	**Non-coal**	**15.32**	**0.00**	**1.09**
	Coal (Reference)			

Categories with significant odds ratios are in bold. Arrow shows decreasing susceptibility

Table 9 FP vs. NDLR

Predictor	Category	Wald statistic	Significance (*p* value)	Odds ratio
District	Northeast	0.01	0.92	0.99
	Southeast	1.61	0.20	0.84
	North central	4.00	0.05	0.75
	South central	0.01	0.93	1.01
	Rocky Mountain	1.04	0.31	0.88
	Western (reference)			
Experience on the current job (years)	0–1	0.01	0.93	0.98
	1–3	0.10	0.75	1.06
	3–6	0.09	0.77	0.95
	6–10	0.39	0.53	0.89
	10–20	0.10	0.75	0.94
	20–30	1.04	0.31	0.81
	> 30 (reference)			
Shift	7 am	0.03	0.86	0.98
	3 pm	0.27	0.60	1.07
	11 pm (reference)			
Age (years)	*18–30*	*42.83*	*0.00*	*0.51*
	31–40	*28.37*	*0.00*	*0.59*
	41–50	*5.50*	*0.02*	*0.80*
	51–70 (reference)			
Accident type	Handling material	1.59	0.21	0.82
	Hand tools	*59.99*	*0.00*	*0.21*
	Powered haulage	*58.07*	*0.00*	*3.50*
	Machinery	*10.42*	*0.00*	*1.68*
	Slip/fall of a person	*27.96*	*0.00*	*0.33*
	Fall of roof, back, or brow (reference)			
Employment type	*Operator*	*10.04*	*0.00*	*0.75*
	Contractor (reference)			
Mine type	Non-coal	0.48	0.49	0.94
	Coal (reference)			

Categories with significant odds ratios are in italics

a reference (control group) and the latter being a target (treatment group). If O_A is the odds of the event in the reference group, and O_B is the odds of the event in the target group, then the ratio of O_B to O_A is called the odds ratio (see Eq. 2). The odds ratio provides a sense of how many times the event is likely to occur in the target group, relative to the reference group [25].

$$\text{Odds ratio} = \frac{O_B}{O_A} \tag{2}$$

Applying this concept to the results in Table 8, odds will be the ratio of the probability that a miner suffers DLR to the probability that they do not. The age variable in the table shows four groups of miners, with the last group (miners aged 51–70) being the reference. For each age group, the odds is determined as the ratio of the probability that a miner in that group suffers DLR to the probability that they do not. Thereafter, we can determine the odds ratio using the reference group and any other group in the age variable. With the first group (miners aged 18–30) as a target, we initially determine the odds as the ratio of the probability that a miner aged 18–30 suffers DLR to the probability of not suffering DLR. Having done the same for miners in the reference group, we proceed to determine the odds ratio for the target group. We determine this as the ratio of the odds that a miner aged 18–30 suffers DLR to the odds that a miner aged 51–70 suffers DLR. The odds ratio provides a measure of how more or less susceptible the first group is to DLR, relative to the reference group. We subject the remaining groups to a similar process to obtain the corresponding odds ratios. A group with an odds ratio greater than one denotes higher susceptibility to DLR injuries, relative to the reference group. An odds ratio less than one denotes lesser susceptibility, while an odds ratio of one means there is no significant difference in susceptibility between the reference and target groups. Reference groups always have an odds ratio of one, since they serve as their own target groups.

From the age variable in Table 8, we note that miners aged 51–70 (reference category) are the most susceptible to DLR; all the other categories have odds ratios that are less than one. In determining injury susceptibility via the odds ratios, one thing to consider is the significance (p values) of the odds ratios. The susceptibility of a target group is not significantly different than that of the reference group if the corresponding p value for the target group is greater than 0.025. For instance, we observe that though the odds ratio for miners aged 41–50 in Table 8 is lower than one, their susceptibility is not significantly different than that of the reference category, since $p = 0.11 > 0.025$. Thus, miners aged 41–50 have about the same susceptibility to DLR as do miners in the reference category. We obtain the p values for the odds ratios from a two-tailed Wald test at a 5% level of significance.

All categories of the accident type variable show a significant difference in susceptibility when compared to the reference category (fall of roof, back, or brow). A worker who suffers a hand tools incident has the least susceptibility to DLR. Among the remaining accidents, slip or fall presents the highest susceptibility to DLR. The northeastern district has the highest susceptibility to DLR among all mining districts, with the Rocky Mountain district showing the least susceptibility. An operator starting work in the swing shift (2nd shift) in a non-coal mine is highly susceptible to DLR.

From the results in Table 9 (FP vs. NDLR), an increase in age is associated with increased susceptibility to FP. With a change in shift, there is no significant change in susceptibility to FP. Though shifts 1 and 2 have different odds ratios than shift 3 (the reference category), they do not present a significant change in susceptibility to FP (since $p > 0.025$). Likewise, there is no significant difference in susceptibility to FP among the various districts and experience on the current job. Among the six accident categories, powered haulage incidents make a worker most susceptible to FP. The category with the least susceptibility is hand tools. Operators are less likely to sustain FP injuries than contractors (0.75 times as likely as contractors). This result has some similarity with the study by Muzaffar et al. [13], which reports that contractors are more susceptible to fatalities than operators. Though workers in non-coal mines have lesser odds ratios than those in coal mines, the difference is not significant ($p = 0.49 > 0.025$). This implies that the susceptibility to FP for the coal and non-coal miners, over the study period, has been about the same. This again has some similarity with Muzaffar et al.'s [13] work, where results show there is no significant difference in susceptibility to fatality between the coal and non-coal miners ($p = 0.68$). Further analysis reveals that fatality and permanent disability records gathered over the study period by MSHA, i.e., original data, have been about the same for the coal and non-coal mines (Fig. 5a). From Fig. 5a, the number of cases for coal mines is relatively higher during the first half of the period. This is reversed in the second half of the period, with non-coal mines recording a relatively higher number of cases. Figure 5b, which summarizes fatality and permanent disability records for the final dataset used in modeling, corroborates the observation we make in Fig. 5a. Both figures show similar trends for the coal and non-coal mines over the 10-year period. While logistic regression can identify this trend, the analysis does not take into consideration the population size, e.g., employee hours worked. This limitation, which has the potential to yield biased results for the mine type variable, is further discussed in Section 5.

4.3 Risk Profile

At a 5% level of significance, the odds ratios provide a means to assess how the independent variables interact to influence a miner's susceptibility to a class of injury under consideration. Specifically, the categories of a given risk factor (independent variable with significant influence on the model) are compared to one another using the odds ratios to assess relative susceptibility to the class of injury being considered. Figures 6, 7, 8, and 9 summarize these interactions and show how categories of a given risk factor vary in susceptibility to DLR and FP. The odds ratios in these figures have been adjusted to account for the 5% level of significance; i.e., odds ratios that are not significant at the 5% level are recomputed to one. Figure 6 shows a decreasing trend of susceptibility to DLR with an increase in experience on the job, i.e., experience working in a specific job title, not experience as a miner. An increase in a miner's experience on the job does not cause a significant change in their susceptibility to FP. An increase in age generally comes with increased susceptibility to both DLR and FP (Fig. 7). From Figs. 6 and 7, the susceptibility to DLR decreases with increasing experience on the job but increases with an increase in age. A general convention would be for DLR susceptibility to

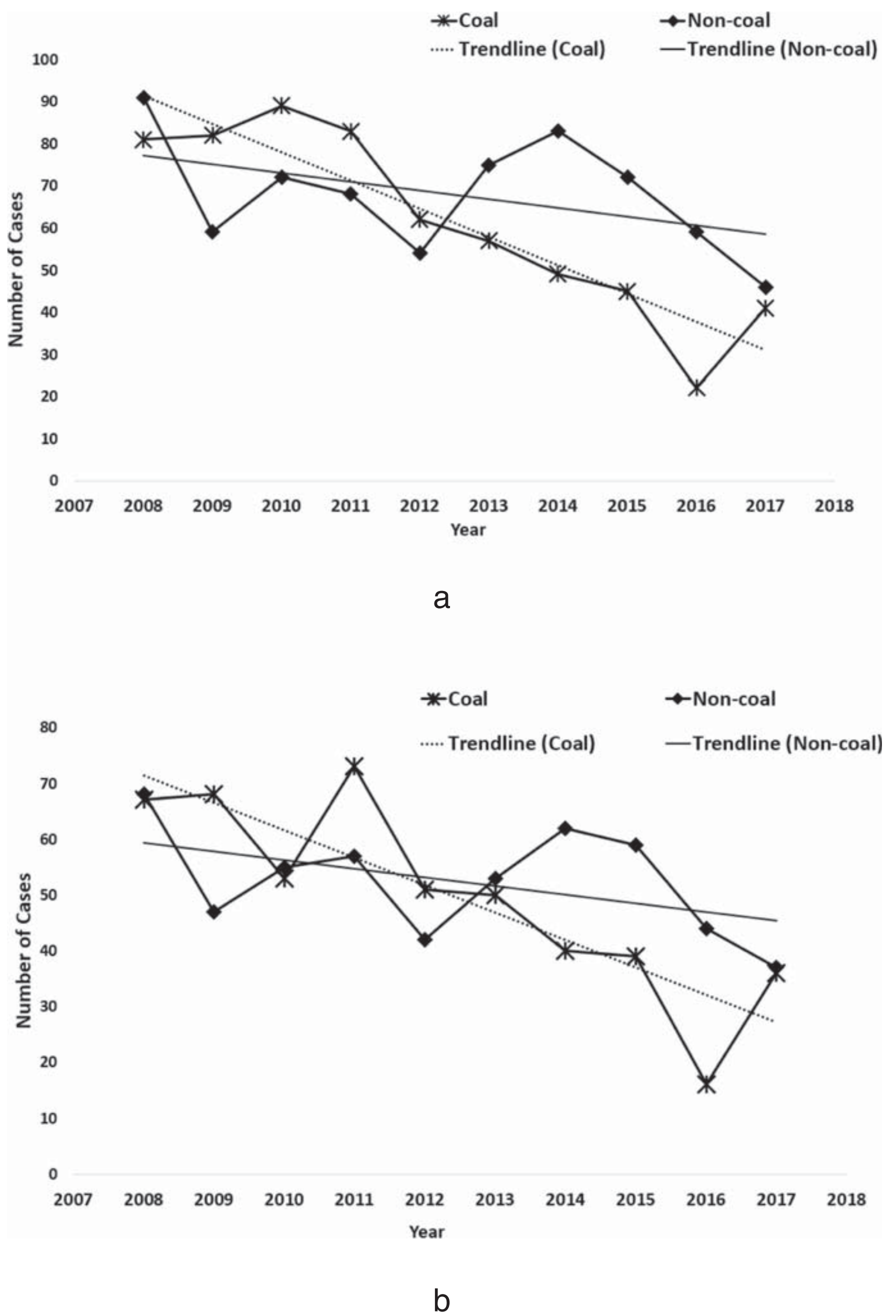

Fig. 5 **a** Original MSHA data summary for fatality and permanent disability (FP) for the coal and non-coal mines (unprocessed data). **b** Modified MSHA data summary for fatality and permanent disability (FP) for the coal and non-coal mines (final data used in modeling)

decrease with an increase in age, just as we observe for experience on the job, since we expect experience on the job and age to be generally proportional. This is, however, not always the case, considering that the experience we refer to here is the experience on a miner's current job, prior to the incident. Thus, a miner might be relatively older but with less experience on a particular job. Figure 8 shows that the accident type one encounters comes with significant variation in susceptibility to either DLR or FP. In both cases of injuries, hand tools presents the lowest susceptibility among the six major accidents under study. Powered haulage has the highest susceptibility to FP, while slip or fall has the highest susceptibility to DLR. While there is no significant change in susceptibility to FP from one district to the other, susceptibility to DLR varies from one district to the other (Fig. 9). Miners in the swing shift (with a peak start time of 3 pm)

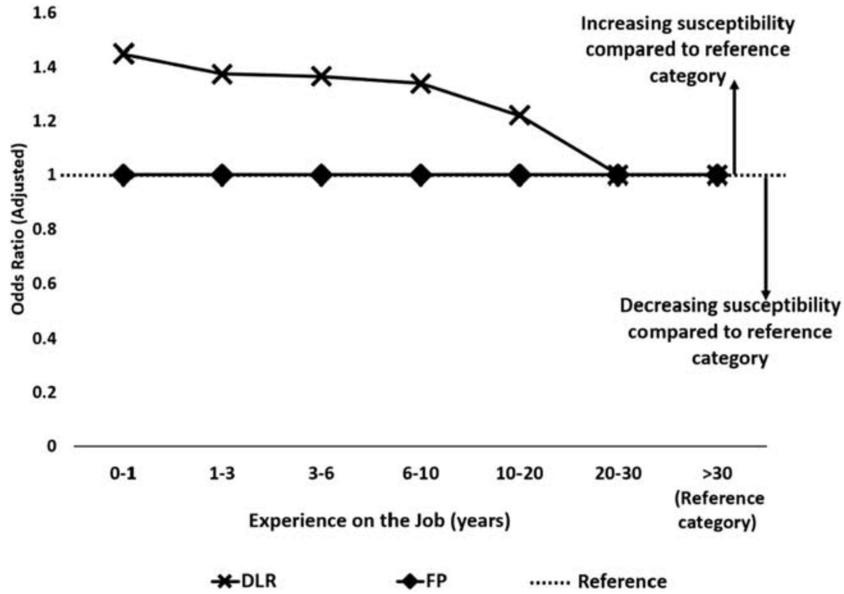

Fig. 6 Risk profile for experience on the job. Susceptibility to DLR generally decreases with an increase in experience on the job

are the most susceptible to DLR. This is not so with FP; none of the shifts shows higher susceptibility to FP. While operators are more susceptible to DLR than contractors, the converse is true for FP. Though non-coal miners are more susceptible to DLR than coal miners, they have about the same susceptibility to FP. Table 10 summarizes the most susceptible categories for each class of injury under consideration.

Though handling material is the most reported accident, as shown in Fig. 4, it does not prove to be the most susceptible incident to either DLR or FP, as Fig. 8 shows. Such detail is provided by logistic regression, which presents a more robust form of analysis. From the logistic regression analysis, we identify powered haulage and slip or fall as incidents that make a worker most susceptible to FP and DLR, respectively. This shows the applicability of logistic regression over basic statistics for injury analysis.

4.4 Surface vs. Underground

Considerable differences exist between the surface and underground operations regarding equipment type and design, schedule of operations, ambient conditions, and layout of the working environment. To this end, we split the dataset into the surface and underground incidents and conduct comparative analysis to assess the differences between them. Apart from work location (redundant for this purpose), all variables that have been

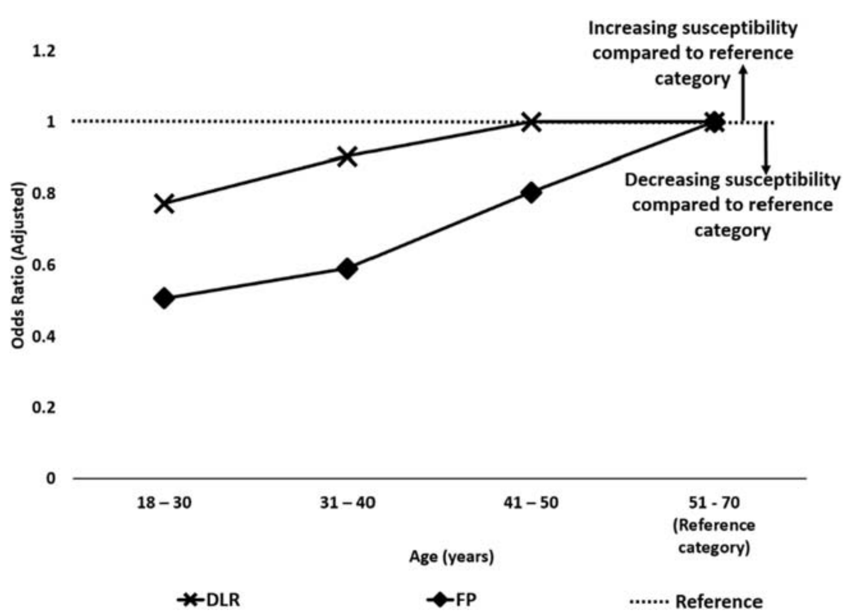

Fig. 7 Risk profile for age. Susceptibility to DLR and FP generally increases with an increase in age

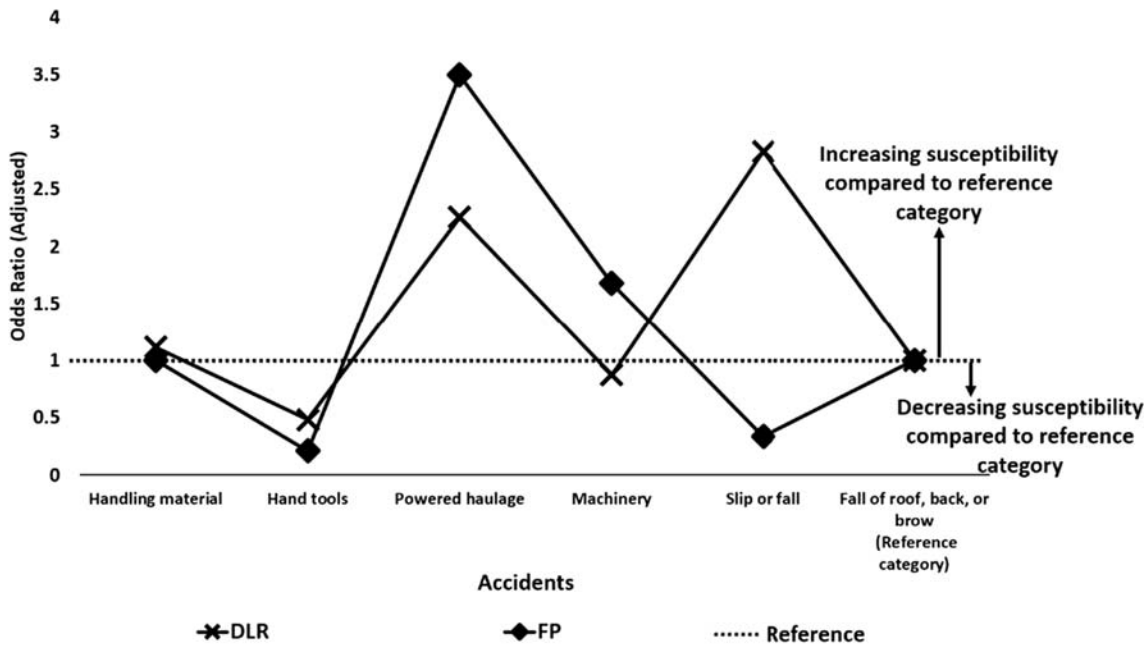

Fig. 8 Risk profile for accidents

considered in the final model of the general case are also considered for both the surface and underground scenarios.

From the analysis of the surface incidents, we realize that hours at work prior to injury occurrence do not have a significant impact on the model. For the underground incidents, the following variables do not significantly impact the model:

1. Experience on the job
2. Employment type (operator vs. contractor)
3. Mine type (coal vs. non-coal)
4. Hours at work prior to injury occurrence

Table 11 summarizes the respective variables that have had a significant impact on the surface and underground injury occurrences over the period under study. Figures 10 and 11 show the risk profiles for experience on the job and age variables, respectively, for the surface scenario. There is a decreasing trend in susceptibility to DLR as experience on the job increases, while a change in experience on the job does not cause a change in susceptibility to FP (Fig. 10). This is similar to the results for the general case (Fig. 6). Figure 11 shows an increase in susceptibility to FP as age increases. DLR susceptibility increases with age until 40 years and levels off

Fig. 9 Risk profile for districts

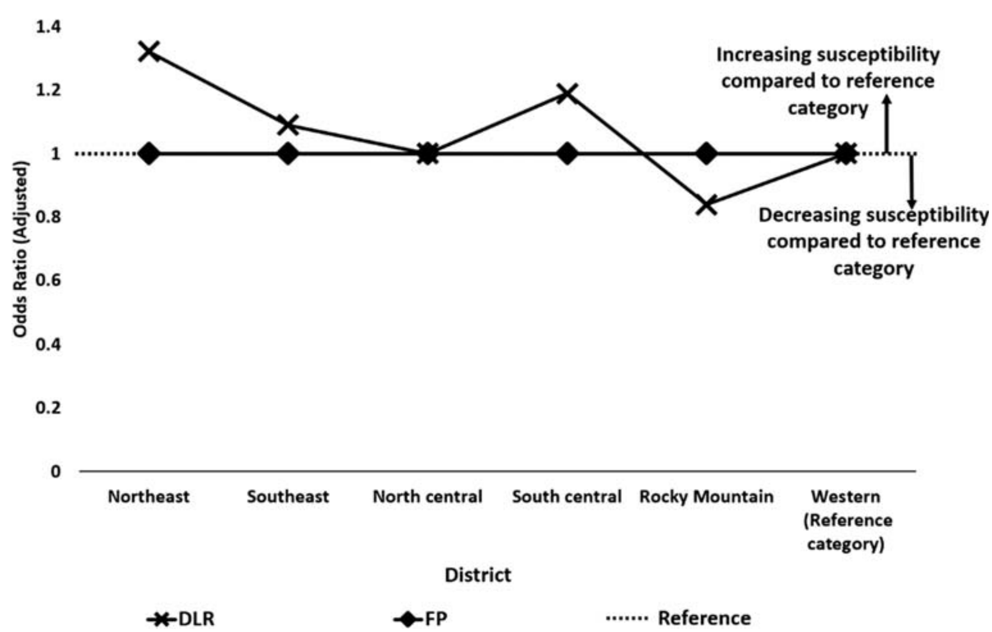

Table 10 Most susceptible categories

Predictor	DLR	FP
District	Northeast	*
Experience on the job (years)	0–1	*
Shift start time	Shift 2 (peak start time, 3 pm)	*
Age (years)	51–70	51–70
Employment type	Operator	Contractor
Mine type	Non-coal	*
Accident type	Slip or fall	Powered haulage

*Implies all categories of the predictor have equivalent injury susceptibility

thereafter. Figure 12 shows the risk profile for age in the underground scenario. An increase in age generally comes with an increase in susceptibility to both DLR and FP.

Table 12 summarizes the differences between the surface and underground incidents in respect of the most susceptible groups of miners. There are a few instances where the results for the surface and underground are similar. For DLR injuries, both the surface and underground have miners aged 41–50 and 51–70 as the most susceptible categories. Likewise, they share slip or fall as the most susceptible accident to DLR. Aside from these two variables (i.e., age and accident type), they have varying results for DLR. For FP, both the surface and underground incidents have miners aged 51–70 as the most susceptible category. Another similarity is that no district or shift stands out as the most susceptible category for either surface or underground incidents.

5 Discussion

The results of this study identify the risk factors associated with three injury classes, namely, fatal and total permanent or partial permanent disability (FP), non-fatal with days lost and/or days of restricted work activity (DLR), and non-fatal with no days lost or restricted activity (NDLR). The age variable for both DLR vs. NDLR (Table 8) and FP vs. NDLR (Table 9) identifies the older miners, i.e., 51–70 years of age (reference category), as the most susceptible group. This may be as a result of an increase in the population of older workers in the workforce in recent years [26]. Slip or fall accidents have the highest susceptibility to DLR injuries. This aligns with a study by Nowrouzi-Kia et al. [27] who identify slip and fall accidents as a major contributor to lost-time injuries (defined in this study as DLR) in the mining industry. Contractors are more likely to sustain FP injuries than operators, and this has some corroboration with the study by Muzaffar et al. [13] on MSHA data, where the authors observe that contractors are more susceptible to fatalities than operators. In consequence, the results of our study and that of Muzaffar et al. [13] suggest that contractors have been more susceptible to the severest form of injury over the past two decades.

The foregoing analysis culminates in a useful tool for identifying areas of risk in mining operations. This can be applied to historic data from a single mine to enable safety officers to obtain detailed information about aspects of operations that require immediate risk or hazard mitigation measures. Corporations and government institutions with stakes in mine health and safety can apply this tool to historic data from a cluster of mines, with the aim of identifying common areas of risk. This will, subsequently, enhance the ability of such institutions to provide the necessary support and guidance towards targeted risk mitigation on the concerned mine sites. This information can benefit the industry by identifying specific training areas based on the operation's demographics, for example.

The usefulness of this analytic tool can be further enhanced if the data quality is improved. Issues with data quality include missing records, inconsistency, and subjectiveness. For instance, apprentices and trainees are used interchangeably when referencing the over 200 job

Table 11 Significant predictors (surface vs. underground)

Significant predictors	Surface	Underground
District	✓	✓
Experience on the job (years)	✓	
Shift start time	✓	✓
Age (years)	✓	✓
Employment type	✓	
Mine type	✓	
Accident type	✓	✓
Hours at work prior to injury occurrence		

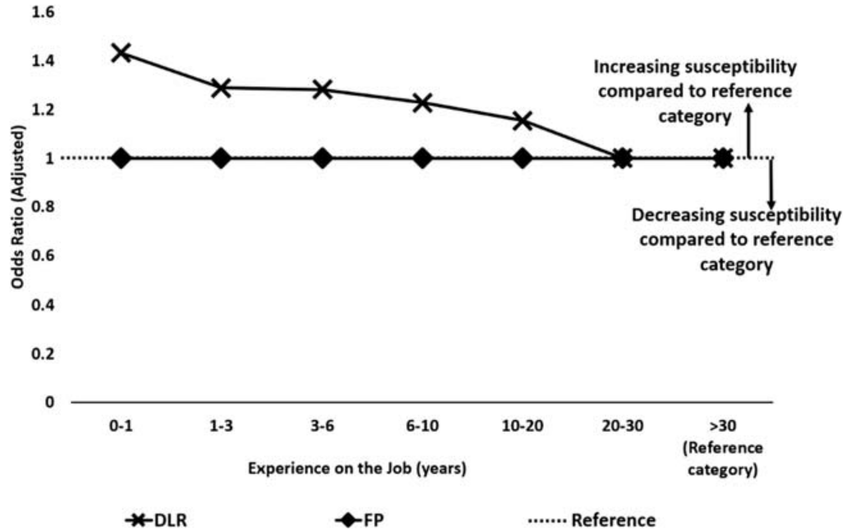

Fig. 10 Risk profile for experience on the job (surface). Susceptibility to DLR generally decreases with an increase in experience

titles found in the database. Aggregation of these titles into related job roles will enhance data analysis that incorporates the effect of job roles. To this end, MSHA can introduce a new variable in the database that represents an aggregated form of the existing job titles. Furthermore, safety supervisors can review accident and injury reports for missing data before onward submission to MSHA. This will contribute to a more comprehensive and robust data analysis.

It must be noted that logistic regression does not provide all of the necessary analysis to have a complete understanding of safety accidents. It can be used to identify trends and to look beyond industry standard safety statistics, e.g., incident rates. We use logistic regression in this study to identify trends that are not easy, if even possible, to distinguish using standard safety statistics. We, however, note that there is a limitation to the results of the analysis, and that is, the study does not consider the impact of case rate (number of incidents over a given population, e.g., hours worked or number of mines per type). Thus, results developed for the mine type variable may be biased since the analysis only considers the number of incidents. For instance, we realize from the results that workers in non-coal mines have about the same susceptibility to FP as workers in coal mines. This conflicts with the conventional knowledge that there are more non-coal miners in the USA than coal miners and that both categories of miners recorded about the same number of incidents during the period under study (Fig. 5), which

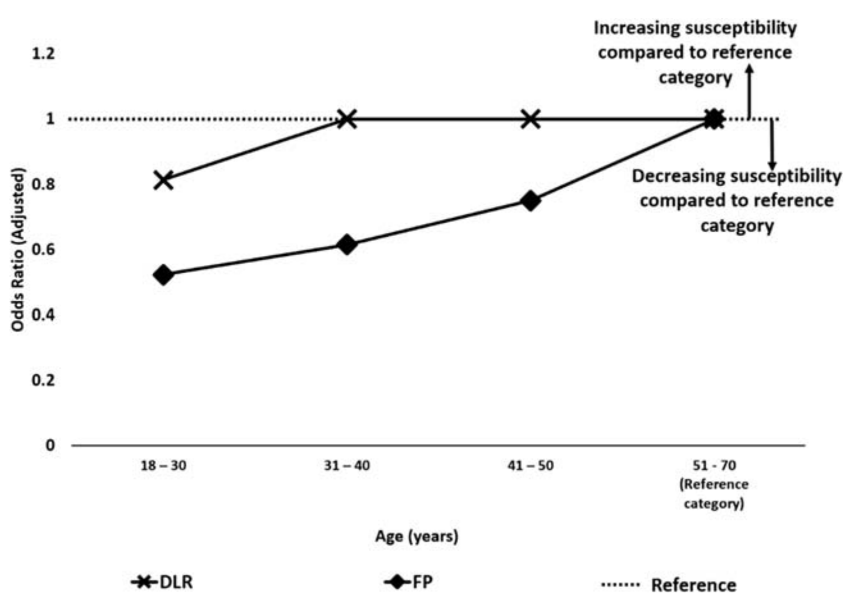

Fig. 11 Risk profile for age (surface). Susceptibility to FP increases with an increase in age

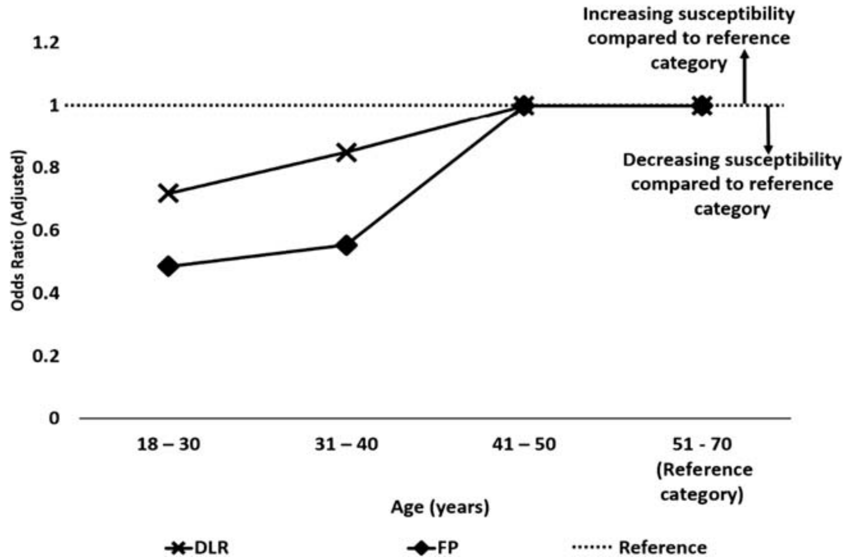

Fig. 12 Risk profile for age (underground). Susceptibility to DLR and FP generally increases with an increase in age

suggests that coal miners are rather more susceptible than non-coal miners. Interestingly, this convention is also in contrast with Muzaffar et al.'s [13] study, where results show there is no significant difference in susceptibility to fatality between the coal and non-coal miners ($p = 0.68$) over the period 1998–2007. Further analysis by the authors, based on work location (surface vs. underground), shows that non-coal miners are more susceptible to fatality in either work location.

6 Conclusions

We analyze injury data collected by the Mine Safety and Health Administration for the period 2008 to 2017 using multiclass logistic regression. As a result, miner's age, mine type (coal vs. non-coal), experience on the current job (years), shift start time, employment type (operator vs. contractor), mining district, and accident type have been identified as safety risk factors. We show how interactions between these variables contribute to a miner's susceptibility to the following injury classes: non-fatal with no days lost or restricted activity (NDLR), non-fatal with days lost and/or days of restricted work activity (DLR), and fatal and total permanent or partial permanent disability (FP). Powered haulage and slip or fall are the incidents that make miners most susceptible to FP and DLR, respectively. Contractors are more susceptible to FP than operators, while operators are more susceptible to DLR than contractors. Miners in shift two (with a peak start time of 3 pm) are the most susceptible to DLR, compared to miners in other shifts. Miners aged 51–70 are the most susceptible age group to both DLR and FP. The researchers identify miners

Table 12 Most susceptible categories (surface vs. underground)

Predictor	DLR		FP	
	Surface	Underground	Surface	Underground
District	Southeast	Northeast	*	*
Experience on the job (years)	0–1	N/A	*	N/A
Shift start time	*	Shift 2 (peak start time, 3 pm)	*	*
Age (years)	31–40, 41–50, 51–70	41–50, 51–70	51–70	41–50, 51–70
Employment type	Operator	N/A	Contractor	N/A
Mine type	Non-coal	N/A	*	N/A
Accident type	Slip or fall	Slip or fall	Machinery	Powered haulage

N/A refers to predictors that have no significant impact on underground incidents; they are subsequently dropped from the model

*Implies all categories of the predictor have equivalent injury susceptibility with up to a year's experience on the job as the most susceptible to DLR, which is perhaps the most actionable takeaway from this analysis. Regardless of the overall mining experience of the individual, the initial year in a new position incurs the highest DLR risk.

This method does have a limitation, as seen when comparing the FP susceptibility of the coal and non-coal miners. The results show that both injury classes have about the same susceptibility, yet we know that there are more non-coal mines in the USA than coal mines and that both mine types recorded about the same number of incidents over the period. Therefore, the analysis is not correctly weighted for mine type (coal vs. non-coal).

We further show the differences and similarities between the surface and underground mine incidents. Miner's age, mine type (coal vs. non-coal), experience on the current job (years), shift start time, employment type (operator vs. contractor), mining district, and accident type are the risk factors associated with susceptibility to surface injuries. Those associated with susceptibility to underground injuries are miner's age, shift start time, mining district, and accident type. Both the surface and underground cases have slip or fall as the accident type that makes miners most susceptible to DLR. Regarding FP, machinery and powered haulage are the accident types that make miners most susceptible in the surface and underground, respectively.

To this end, multiclass logistic regression proves to be a tool that can be used beyond basic statistics in providing robust mine accident and injury analysis. Using this tool, safety managers can identify areas that need prioritized training or attention. Periodic analysis with this tool and taking pragmatic measures, thereafter, will promote a strong safety culture and risk mitigation in the mine environment.

It must be acknowledged that a significant proportion of the original data has incomplete records. A total of 16.70% of the records are missing one or more of the variables considered in this study; therefore, that data was not included in this analysis. We recommend that safety supervisors review accident and injury reports for missing data before onward submission to MSHA. Having a more complete dataset will contribute to a more robust analysis by researchers and safety personnel. Additionally, incorporating the distribution of coal to non-coal mine operations and surface to underground operations would provide more realistic results by taking into account the number of operations and incidents in each category.

Plans for future work include evaluating different dependent variables, e.g., accident type, and aggregation strategies to provide more insight into injury and accident occurrence on mine sites. With the aggregation of the degrees of injury, variations of aggregation methods should be evaluated in addition to varying the dependent variable. Different supervised machine learning techniques could be evaluated to compare results. While this study identifies risk factors, it does not provide the why. Identifying the root cause of increased susceptibility to a given injury class would require additional data and analysis.

Funding We thank the National Institute for Occupational Safety and Health (NIOSH) for funding this research under grant 0000HCCR-2019-36404.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Onder M, Onder S, Adiguzel E (2014) Applying hierarchical loglinear models to nonfatal underground coal mine accidents for safety management. Int J Occup Saf Ergon 20:239–248
2. Karra VK (2005) Analysis of non-fatal and fatal injury rates for mine operator and contractor employees and the influence of work location. J Saf Res 36:413–421
3. Friedman LS, Almberg KS, Cohen RA (2019) Injuries associated with long working hours among employees in the US mining industry: risk factors and adverse outcomes. Occup Environ Med 76: 389–395
4. Cullen ET, Camm T, Jenkins M, Mallett L (2006) Getting to zero: the human side of mining. Information Circular 9484. National Institute for Occupational Safety and Health (NIOSH), Spokane Research Laboratory, Spokane, WA
5. MSHA (2020) MSHA reports fatal mining accidents dropped in 2019. In: Min Eng. https://me.smenet.org/webContent.cfm?webarticleid=2956. Accessed 20 Apr 2020
6. Nieto A, Duerksen A (2008) The effects of mine safety legislation on mining technology in the USA. Int J Min Miner Process Eng 1: 95–103
7. NIOSH (2016) Section 8 Coding Manual. https://www.cdc.gov/niosh/mining/UserFiles/data/codes.pdf. Accessed 19 Aug 2020
8. Grosan C, Abraham A (2011) Machine learning. Intell Syst Ref Libr. https://doi.org/10.1007/978-3-642-21004-4_10
9. Carranza EJM, Hale M (2001) Logistic regression for geologically constrained mapping of gold potential, Baguio district, Philippines. Explor Min Geol 10:165–175
10. Palei SK, Das SK (2009) Logistic regression model for prediction of roof fall risks in bord and pillar workings in coal mines: an approach. Saf Sci 47:88–96
11. Peng CYJ, Lee KL, Ingersoll GM (2002) An introduction to logistic regression analysis and reporting. J Educ Res 96:3–14
12. Bennett JD, Passmore DL (1985) Multinomial logit analysis of injury severity in U.S. underground bituminous coal mines, 1975-1982. Accid Anal Prev 17:399–408
13. Muzaffar S, Cummings K, Hobbs G, Allison P, Kreiss K (2013) Factors associated with fatal mining injuries among contractors and operators. J Occup Environ Med 55:1337–1344
14. Maiti J, Bhattacherjee A (2001) Predicting accident susceptibility: a logistic regression analysis of underground coal mine workers. J South Afr Inst Min Metall 101:203–208
15. Ajith MM, Ghosh AK, Jansz J (2020) Risk factors for the number of sustained injuries in artisanal and small-scale mining operation. Saf Health Work 11:50–60
16. Chau N, Mur JM, Benamghar L, Siegfried C, Dangelzer JL, Français M, Jacquin R, Sourdot A (2004) Relationships between certain individual characteristics and occupational injuries for

various jobs in the construction industry: a case-control study. Am J Ind Med 45:84–92
17. Zhang KI, Hassan M (2019) Injury severity analysis of nighttime work zone crashes. ICTIS 2019 - 5th Int Conf Transp Inf Saf 1301–1308
18. Robin P (2014) Use on multinomial logistic regression in work zone crash analysis for Missouri work zones. MS Thesis. Missouri University of Science and Technology
19. Akboga Kale Ö, Baradan S (2020) Identifying factors that contribute to severity of construction injuries using logistic regression model. Tek Dergi. https://doi.org/10.18400/tekderg.470633
20. MSHA (2018) Mine Injury and Worktime, Yearly Report. https://arlweb.msha.gov/Stats/Part50/WQ/2018/MIWQReportCY2018.pdf. Accessed 22 Apr 2020
21. SPSS Software (2020) IBM SPSS software. https://www.ibm.com/analytics/spss-statistics-software. Accessed 22 Apr 2020
22. Stockburger DW (1996) Introductory Statistics: Concepts, Models, and Applications. Missouri State University
23. Heumann C, Schomaker M, Shalabh (2017) Introduction to statistics and data analysis: With exercises, solutions and applications in R. Springer Int Publ. https://doi.org/10.1007/978-3-319-46162-5
24. Minitab (2020) Multicollinearity in regression - Minitab. https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/model-assumptions/multicollinearity-in-regression/. Accessed 22 Apr 2020
25. Hosmer DW, Lemeshow S (2000) Applied Logistic Regression, 2nd ed. John Wiley & Sons
26. Smith S, Pegula S (2020) Fatal occupational injuries to older workers. Mon Labor Rev. https://doi.org/10.21916/mlr.2020.2
27. Nowrouzi-Kia B, Sharma B, Dignard C, Kerekes Z, Dumond J, Li A, Larivière M (2017) Systematic review: lost-time injuries in the US mining industry. Occup Med (Chic Ill) 67:442–447