

An Analysis of Contributing Mining Factors in Coal Workers' Pneumoconiosis Prevalence in the United States Coal Mines, 1986-2018

Younes Shekarian

Penn State: The Pennsylvania State University

Elham Rahimi

Penn State: The Pennsylvania State University

Naser Shekarian

University of Colorado Denver

Mohammad Rezaee

Penn State: The Pennsylvania State University

Pedram Roghanchi (✉ pedram.roghanchi@nmt.edu)

New Mexico Institute of Mining and Technology

Research

Keywords: Respirable coal mine dust, respiratory diseases, coal worker's pneumoconiosis (CWP), occupational exposure

Posted Date: May 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-506261/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at International Journal of Coal Science & Technology on October 13th, 2021. See the published version at <https://doi.org/10.1007/s40789-021-00464-y>.

Abstract

In the United States, an unexpected and severe increase in coal miners' lung diseases in the late 1990s prompted researchers to investigate the causes of the disease resurgence. This study aims to scrutinize the effects of various mining parameters, including coal rank, mine size, mining method, coal seam height, and geographical location on the prevalence of CWP in surface and underground coal mines. A comprehensive dataset was created using the U.S. Mine Safety and Health Administration (MSHA) Employment and Accident/Injury databases. The information was merged based on the mine ID by utilizing SQL data management software. A total number of 123,643 mine-year observations were included in the statistical analysis. Generalized Estimating Equation (GEE) model was used to conduct a statistical analysis on a total of 29,707, and 32,643 mine-year observations for underground and surface coal mines, respectively. The results of the econometrics approach revealed that coal workers in underground coal mines are at a greater risk of CWP comparing to those of surface coal operations. Furthermore, underground coal mines in the Appalachia and Interior regions are at a higher risk of CWP prevalence than the Western region. Surface coal mines in the Appalachian coal region are more susceptible to CWP than miners in the Western region. The analysis also indicated that coal workers working in smaller mines are more vulnerable to CWP than those in large mine sizes. Furthermore, coal workers in thin-seam underground mine operations are more likely to develop CWP.

1. Introduction

Mining operations such as drilling, cutting, crushing, blasting, and material handling are inherently associated with dust generation [1,2]. Dust negatively impacts the air quality in the mining area and causes severe health hazards such as lung disease [3,4]. Cumulative inhalation of RCMD[1] can lead to diseases such as coal worker's pneumoconiosis (CWP), silicosis, mixed dust pneumoconiosis, dust-related diffuse fibrosis (DDF), and progressive massive fibrosis (PMF) [3,6].

In the US, after the implementation of the interim coal mine dust standard of 3.0 mg/m³ in 1970, and the final standard of 2.0 mg/m³ in 1972, the prevalence of CWP and concentrations of coal mine dust declined substantially [1,5]. However, a resurgence in disease incidence has been observed since the mid-1990s [3]. A study conducted by NIOSH in 2005 reported a rapidly progressive disease among coal miners. This study revealed that among the miners with pneumoconiosis, 35% diagnosed with rapidly progressive disease including 15% with the most severe and often lethal form (i.e., PMF) [11]. Medical surveys further revealed that relatively young miners, who have spent their entire employment in the modern dust control regulations, are being affected [2,39]. Such rapidly progressive pneumoconiosis has been specifically observed in the Appalachian coal region [7-9].

The geographical clustering of pneumoconiosis has been speculated to be mainly due to mining thin-coal seams in this area, necessitating cutting a significant amount of roof and floor rock. These out of seam materials are mostly composed of silica (i.e., quartz) and silicate minerals (cyclosilicates, phyllosilicates, tectosilicates, etc.) [16], which have been associated with severe lung diseases [6]. Other potential factors such as duration and level of exposure, mining method, coal and rock-strata geological conditions, dust characteristics (i.e., size, shape, mineralogy, elemental content), dust mitigation techniques, mine size, and coal rank have also been suggested as contributing factors to the recent unexpected trend [9-16,40]

The recent resurgence of coal mine dust lung diseases has raised concern in the scientific and regulatory community. To address the concern, Mine Safety and Health Administration (MSHA) issued the 2014 dust rule changing the RCMD measurement technology, allowable limits, and sampling protocol for RCMD exposure [3]. However, many researchers have questioned whether the reduced limit would actually target the root problem [e.g., 15, 16]. A recent report convened by the National Academies of Sciences, Engineering, and Medicine (NAS) highlighted the need for research efforts with respect to respirable coal dust including characterization and deposition, sampling protocol, and monitoring strategies for controlling miner's exposure to RCMD [3].

This study was motivated by recommendations number 8,10, and 12 of the NAS report, which are associated with achieving a greater understanding of the relationships between mining activities and technology changes in coal extraction and increases in coal worker's diseases. The objective of this study is to determine the effect of several factors, namely, mine size, coal rank, mining method, geographical location, and coal seam thickness to the prevalence of CWP. To achieve this objective, a detailed statistical analysis of the relationship between the rate of CWP cases and the potential contributing factors was conducted.

2. Material And Methods

2.1 Data Acquisition

A comprehensive dataset was extracted from the MSHA accident/injury and MSHA employee/production. The number of CWP cases, mine ID (as the unique key for this data), mine operation type, miners' age, total/mine/job experience, and gender were obtained from the MSHA accident/injury. The number of employees (as an indicator for operation size), coal production rate, number of coal mines, coal seam thickness, geographic location, state, county, coal rank, and mine ID were collected from MSHA employee/production file. These groups of information were then merged based on the mine ID by utilizing SQL data management software. The summary report from SQL provided a total number of 123,589 mine-year observations (i.e., $\sum i.t$ (number of mines \times number of years)) from 1986 to 2018. These observations were obtained from 21,396 number of mine operations categorized as underground, surface, and others (i.e., augur, milling and preparation plant, office, culm banks, independent shops and yards, surface at underground mine), as follows:

1. Number of observations for underground mines: $X=5,477$ (mine-ID)*5.42(avg.) = 29,707
2. Number of observations for surface mines: $Y=5,759$ (mine-ID) *5.67(avg.) = 32,643
3. Number of observations for other operations: $Z=10,160$ (mine-ID) *6.03(avg.) = 61,239

Where avg. is the average number of observations for each category between 1986 to 2018.

All observations in the analysis included the following information:

- MSHA mine identification (mine ID)
- Mine operation (surface, underground, and others)
- Geographic location (Appalachian, Western, and Interior)[1]
- State and county
- The average number of employee in each mine per year (as an indicator for mine size)
- Total number of employee hours per year
- Number of reported CWP cases by mines per year
- Average of coal seam height in each mine
- Coal rank of each mine

In this study, the average number of employees was used as an indicator for mine size, classified in three categories as small, medium, and large with a number of employees less than 50, between 50 and 100, and more than 100, correspondingly [10,15,19]. Furthermore, coal seams were categorized into three different groups based on the average seam height as thin (less than 40 inches), medium (i.e., between 40 and 75 inches), and thick (more than 75 inches). Available data for coal rank was categorized based on two major coal ranks in the U.S., anthracite, and bituminous (both bituminous and sub-bituminous[2]) (Table 1).

The summary report of SQL data was used to conduct the statistical analysis, including descriptive statistics, correlation, and regression analysis. Available data were used to formulate hypotheses for the relationship between CWP rate and the contributing factors (i.e., mine operation, geographic location, mine size, coal rank, and coal seam thickness). Fig. 1 illustrates the schematic of the methodology steps.

2.2 Descriptive statistics

A total of 123,589 mine-year observations was included in the regression analysis. The dependent variable was the rate of CWP, as the number of CWP per total hours of employees in each year for each mine. Mine operation type, mine size, coal seam thickness, geographic location, and coal rank were considered as the independent variables. Table 1 summarizes the type, classification, and description of the classification of each variable. For regression modeling, each multiclass variable was broken into multiple dummy variables as binary variables. A total number of 14 binary variables were defined for the independent variables (i.e., 3 variables for mine operation type, 3 variables for mine size, 3 variables for geographic location, 3 variables for seam thickness, and 2 variables for coal rank).

Table 1 Description of variables involves in the statistical analysis for U.S. coal mines, 1986-2018.

Row	Variable	Categories	Description	Type
1	CWP rate	Number of CWP per total hours of employees	The rate of CWP per total hours in each mine-year is used as an indicator for CWP prevalence in coal mines	dependent
2	Mine operation type	Underground	If the mine operation is underground, then 1, otherwise 0	independent
3		Surface	If the mine operation is surface, then 1, otherwise 0	independent
4		Other	If the mine operation is not underground or surface, then 1, otherwise 0	independent
5	Mine size	small	The average number of employees per each coal mine is equal or less than 50 (i.e., $e \leq 50$), then 1, otherwise 0	independent
6		Medium size	The average number of employees per each coal mine is between 50 and 100 (i.e., $50 < e \leq 100$), then 1, otherwise 0	independent
7		Large size	The average number of employees per each coal mine is more than 100 (i.e., $e > 100$), then 1, otherwise 0	independent
8	Geographic location (according to EIA classification)	Western region	If the mine is located in the Western region, then 1, otherwise 0	independent
9		Interior region	If the mine is located in the Interior region, then 1, otherwise 0	independent
10		Appalachian region	If the mine is located in the Appalachian region, then 1, otherwise 0	independent
11	Seam thickness	Thin	If the average thickness of the coal seam in the mine is less than 40 inches (i.e., $t \leq 40$ "), then 1, otherwise 0	independent
12		Medium	If the average thickness of the coal seam in the mine is between 40-75 inches (i.e., $40 < t \leq 75$ "), then 1, otherwise 0	independent
13		Thick	If the average thickness of the coal seam in the mine is more than 75 inches (i.e., $t > 75$ "), then 1, otherwise 0	independent
14	Coal rank	Bituminous	If the coal rank in the mine is Bituminous or Sub-bituminous, then 1, otherwise 0	independent
15		Anthracite	If the rank of coal in the mine is Anthracite, then 1, otherwise 0	independent

(e: Average number of employees in each mine; t: thickness of coal seam)

2.3 Regression analysis

Because of the panel nature of data, linear regression model was utilized to analyze the relationship between the rate of CWP and the independent variables. Considering a dependent variable of Y , and the independent variable of X , the following equation provides the relationship:

$$(1) \quad Y_{i,t} = \beta X_{i,t} + u_i + \varepsilon_{i,t}$$

Where Y denotes the dependent variable, β is the coefficient, X is the independent variable, i stands for Mine-ID, t represents year, u is mine-specific unobserved heterogeneity (i.e., factors constant over time but unobserved to the econometrician), and ε is the error term (e.g., observation-specific error) [23, 24, 33].

In estimation models, the assumption is that errors are independent and equally distributed. However, in longitudinal data, such as those used in this study, this independency does not exist. In other words, observations for each unit (mines) are correlated (e.g., each

mine has different observations in different years). One possible solution is to include subject-specific random effects in the model fitting. This method is called generalized estimating equations (GEE), which is a nonparametric way to account for this issue. GEE estimates the marginal effect of covariates averaged across units [25,33]. Here, this can be interpreted as the overall effect of mining operation, mine size, geographic location, coal rank, and coal seam thickness on CWP disease among coal miners in the U. S. coal mines during 1986-2018. GEE has been widely used for panel estimation [26,27,33]. This method provided the best fit for our data. The models' covariance structure was specified as exchangeable, implying a shared correlation between observations within each mine. Furthermore, variance inflation factor (VIF), robust regression, and homoscedasticity analyses were performed to achieve accurate estimation results.

A comprehensive review was conducted to develop hypotheses and examine the relationship between CWP rate and mining factors in the U.S. coal mines from 1986 to 2018 [19]. This systematic review found several mine-, individual-, and dust-specific factors that may contribute to the prevalence of lung diseases [19]. The selected database allows for studying the effects of some of those factors, including mine operation type, geographic location, mine size, coal rank, seam height. Therefore, the model was utilized to test the following hypotheses:

H₁: Workers in underground coal mines are more susceptible to CWP.

H₂: The coal region contributes to CWP incidence rate.

H₃: Workers in smaller operations are more susceptible to lung diseases.

H₄: The coal rank contributes to the CWP incidence rate.

H₅: Workers in thin-seam mine operations are more susceptible to the CWP incidence.

3. Results And Discussion

3.1 Descriptive statistics

A description of the single variable, frequency and correlation coefficients is a standard technique for statistical analysis [32]. Descriptive statistical analysis of the data was conducted by measuring means, standard deviations, minimum, and maximum of all the variables (Table 2). These information was determined for each of the independent variables, in underground and surface mines, from the data of 1986 to 2018. As for underground mines, 93% of observations were in the Appalachian region, 3% in the Western region, and the rest in the Interior region. The majority (about 80%) of mines were small size. Additionally, the highest number of CWP cases in the observations was 946 (West Virginia, Boone county). In surface mines, approximately 86% of mines are in the Appalachian region, 9% in the Interior region, and the rest (5%) in the West region. Similar to underground mines, the majority of mines (86%) were classified as small mine. Additionally, the highest number of CWP cases in the observations was 175 (West Virginia, Logan county). The descriptive statistical analysis of the CWP rate for various independent variables is discussed in the following.

Table 2 Descriptive statistics for variables in underground and surface coal mines in the U.S., 1986-2018.

Variable	Underground mines (# of Observations = 29,707)				Surface mines (# of Observations = 32,643)			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
	Number of CWP	0.18	1.52	0	101	0.02	0.40	0
Number of CWP per Total hour	0.00	0.09	0	10	0.00	0.03	0	5
Appalachian region	0.93	0.25	0	1	0.86	0.34	0	1
Interior region	0.03	0.18	0	1	0.09	0.28	0	1
Western region	0.03	0.18	0	1	0.05	0.22	0	1
Small size	0.79	0.41	0	1	0.86	0.35	0	1
Medium size	0.09	0.29	0	1	0.07	0.26	0	1
Large size	0.12	0.33	0	1	0.07	0.26	0	1
Thin seam height ($\leq 40''$)	0.27	0.45	0	1	0.49	0.50	0	1
Medium seam height ($>40 & \leq 75''$)	0.59	0.49	0	1	0.36	0.48	0	1
Thick seam height ($>75''$)	0.14	0.35	0	1	0.15	0.36	0	1
Bituminous coal rank	0.96	0.20	0	1	0.93	0.25	0	1
Anthracite coal rank	0.04	0.20	0	1	0.07	0.25	0	1

Mine operation type: Mining method is a significant determinant of RCMD exposure [3]. RCMD compositions at surface and underground differ from each other, due to the different operations in these mining types [6,37]. Consequently, underground coal workers are at a higher risk of CWP than surface mines due to a confined space and limitations in artificial ventilation systems. Analysis of CWP by mine operation shows that, out of 7,337 CWP cases, the majority (i.e., 76%) was reported in underground mines. Approximately 11% of the CWP cases were reported in surface mine operation. The rest (i.e., 13%) of cases were workers in other mine operations, mainly mill or preparation plant (Fig. 2a). As shown by chronological data of CWP cases in the U.S. coal mines (Fig. 2b), there was a decrease in the rate of CWP by implication of permissible exposure limit (PEL). However, there was an increase in the prevalence of CWP in the 1990s.

Geographic location: The geographic location of coal mines is an important factor in assessing the RCMD health risk. Regional variations in dust characteristics exist due to the geographical clustering of coal mines in the U.S. In Central Appalachia, for instance, mines may have more rock strata sourced dust compared to other regions [14]. Amandus' study showed that coal workers in the eastern region of Appalachian coal field, including West Virginia and Pennsylvania, are at a higher risk of CWP than U.S. western mines [38]. The findings of Sarver's study supported this hypothesis that RCMD characteristics differ substantially among mining regions. Understanding the difference in mineral and elemental compositions, as well as particle size distributions of RCMD among geographic locations sheds light on the recent CWP resurgence [14].

From 1986 to 2018, a total number of 106 counties from 16 states across the country reported coal miners with CWP disease. The number and distribution of CWP prevalence among coal miners in different states and counties are demonstrated in Fig. 3. The hot spot area, including West Virginia, Kentucky, Virginia, and Pennsylvania reported a higher number of CWP cases.

Mine size: Several studies has identified mine size (identified by number of underground miners employed) as a predictor of CWP risk among US underground coal miners [1,5,19,41,42]. These studies indicated workers in small mines are associated with an increased risk of CWP, but it was unknown if abnormal lung functions are also linked to the mine size. Blackley et al. (2014) showed that mine size significantly affects the CWP prevalence and lung function abnormality [10]. The spirometry and radiographic analysis among 3770 coal miners in their study showed that there is a higher risk of abnormal spirometry (18.5% vs. 13.8%, $p < 0.01$), CWP (10.8% vs. 5.2%, $p < 0.01$), and progressive massive fibrosis (2.4% vs. 1.1%, $p < 0.01$) in miners working in small mine operations compared to that of large operations. They also concluded that coal workers in small mines (i.e., the number of employees is less than 50) in Kentucky,

Virginia, or West Virginia are at a higher risk of CWP prevalence than those in large mines. Suarathana et al. (2012) found an association between decreasing mine size and prevalence of CWP and PMF among coal miners in the U.S. One possible explanation is that smaller mines may have fewer health and safety resources than larger operations [1]. Moreover, previous investigations indicated that the average concentrations of RCMD in small mines are higher [1,10-12].

Size analysis of coal mines in the U.S. revealed that most of underground and surface mines are in a small size (Fig. 4a). The distribution of CWP per mine size indicated that the number of CWP in the large mines are more than that in small and medium mines. However, the total number of CWP in each mine size does not necessarily show that people in the large mine size are at a higher risk of CWP. Therefore, the rate of CWP per number of employees was calculated to compare the prevalence of CWP per mine size. The data showed that the rate of CWP in the small mines are more than that in medium and large mines (Fig 4b).

Coal seam thickness: Coal seam thickness is one of the potential contributing factors that could influence the prevalence of CWP among coal miners [9,10,19]. Seam height in coal mines vary based on the coal reserves' geographic location and geological properties. Suarathana et al. (2012) reported that the average coal seam thickness for central Appalachia mines is lower than that in other regions. Further to this review, it was concluded that CWP and abnormal lung function prevalence were likely associated with the low seam height and small mine size in the U.S [9,19]

Distribution of surface and underground mines in three coal seam thickness classes was performed (Fig S1 in SI). Only less than 11% of mines had thick coal seams, and the majority of mines had either thin or medium size coal seam heights. In underground mines, the medium seam height was dominant (i.e., 44%), while the majority (i.e., 50%) of surface mines operated thin coal seams.

The distribution of CWP per coal seam height and mine size for both underground and surface coal mines was subsequently studied. The results showed that CWP rate is more prevalent in underground mines operating medium seams than that of mines operating thin and thick coal seams (Fig. 5a). However, the rate of CWP in surface mines indicated a higher rate percentage for thick seams than thin and medium seams (Fig. 5b). Regardless of coal seam thickness, the majority of CWP cases in underground and surface mines was reported in the small mine size.

Coal rank: Several studies confirmed that there is a higher risk of CWP for higher coal rank, even at the same level of RCMD concentrations [11,17]. Gamble et al., (2011) proposed higher rank coal as a plausible factor for CWP prevalence within Appalachian region [17,18]. In many bituminous coal mines, the higher prevalence of CWP have been also linked to a higher quartz content in respirable dust [8]. Previous studies indicated that an apparent link between the coal rank and CWP may be attributed to the particle surface charge and mineralogical composition of RCMD [34,35]. However, coal rank causal effects have not been exclusively investigated.

Distribution of U.S. surface and underground coal mines based on the coal rank showed that bituminous mines account for about 95% of the coal operations (Fig. 6a). Compared to surface anthracite mines (59 mines), only a few active underground anthracite mines (9 mines) existed in 2018. The distribution of CWP by coal rank indicated that the bituminous coal rank contributes to about 95% CWP rate. (Figs. 6 b and c).

3.2 Regression Analysis

Analysis was performed (at three significance levels of 0.01, 0.05, and 0.1) to determine the presence and strength of correlations among the variables considered in this study. The correlation study indicated no strong correlation between the independent variables (Tables S.1, and S.2 in Supplemental information). Thus, these variables can be used in multivariate regression modeling. Regression analysis of the relationship between CWP rate and the identified contributing factors was carried out utilizing GEE model. Table 3 shows the main results of the GEE analysis which utilized for testing the hypotheses of relationship between CWP and the independent variables (described in section 3.3).

H_1 : It was hypothesized that workers in underground mines are more susceptible to developing CWP than those in other operations. The regression analysis showed that coal workers in underground coal mines are at a higher risk of CWP than surface coal miners ($\beta=4.010$, $p<0.01$). It also showed that coal workers in other mine operation (including milling and preparation plant) are at a higher risk of CWP than workers at surface mines ($\beta=2.706$, $p<0.05$). Therefore, **H_1 is supported**, and mine operation type is a significant factor contributing to the CWP prevalence.

H₂: Geographical location was hypothesized to be a contributing factor to the prevalence of CWP. The regression analysis showed a significant positive coefficient for both Appalachia (including West Virginia, Kentucky, Pennsylvania, Virginia, Alabama, Tennessee, Maryland, and Ohio) and Interior (including Illinois and Indiana) regions compared to the Western (including Wyoming, Texas, New Mexico, Oklahoma, Utah, and Colorado) geographic region. The statistical analysis showed that, compared to the Western region, underground coal workers in both Appalachia ($\beta=4.407, p<0.01$), and Interior ($\beta=3.750, p<0.01$) geographic regions are at a greater risk of CWP. Therefore, **H₂ is supported for underground coal mines**. The result of regression for surface mines, with Western region as a reference, showed that surface coal workers in the Appalachian region are at a higher risk of CWP ($\beta=5.101, p<0.01$). The outcome of the regression model for Interior region was not statistically significant. Therefore, **H₂ is supported only for Appalachia vs. Western surface coal mines**.

H₃: The third hypothesis investigated how the size of mine could influence the prevalence of CWP among coal miners. We categorized the mine size based on the average number of employees in each mine (small: less than 50; medium: between 50 and 100; large: more than 100). The results of the statistical analysis indicated that underground coal workers in small mines are at a higher risk of CWP in comparison with workers at medium ($\beta=-1.961, p<0.01$) and large ($\beta=-1.879, p<0.01$) mines. Therefore, **H₃ is supported for the underground mines**. In surface mines, coal workers in small mines are at a higher risk of CWP in comparison with medium size mine workers ($\beta=-1.277, p<0.1$). The results were not statistically significant for large operations. Therefore, **H₃ is supported only for the medium vs. small surface coal mines**.

H₄: It was hypothesized that coal rank contributes to the CWP incidence rate. For the coal rank, the statistical analysis showed a significant relationship between CWP rates and bituminous coal rank in underground mines. It indicated that underground bituminous coal miners are at a higher risk of CWP than anthracite coal miners ($\beta=7.383, p<0.01$). Therefore, **H₄ is supported for underground mines**. On the other hand, surface anthracite coal miners are at a higher risk of CWP than anthracite coal miners ($\beta=-1.476, p<0.01$). Therefore, **H₄ is supported for the surface mines**. It should be noted that the MSHA database classifies coal rank only as bituminous and anthracite. Only 0.3% of coal production in 2018 came from anthracite [19].

H₅: Finally, a hypothesis examined how the coal thickness could influence the prevalence of CWP among coal miners. The seam thickness was categorized into three groups based on the average of seam thickness in each mine (thin: Seam height ≤ 40 "; medium: $40 < \text{Seam height} \leq 75$ "; and thick: Seam height >75 inches). The GEE result indicated that coal workers in the underground mines operating thin ($\beta=1.416, p<0.05$) and medium ($\beta=1.397, p<0.01$) seams are at a higher risk of CWP prevalence, compared with those working in thick-seam underground operations. Therefore, **H₅ is supported for underground coal mines**. The result of regression for surface mines shows that we cannot make a conclusion for the coal workers in the thin-seam surface operations, but coal workers in the medium-seam operations ($\beta=-1.969, p<0.01$) are at a lower risk of CWP prevalence in comparison with workers for thick-seam mines. Therefore, **H₅ is not supported for the medium vs. thick-seam surface coal mines**. The results of hypothesis testing are summarized in Table 4.

Table 3 GEE Estimation Results. The values represent coefficient. The standard errors are included in the parentheses.

Category	Variables	Underground	Surface
		CWP-rate	CWP-rate
Mine operation	Surface (reference)	4.010**	
	Underground	(0.73)	
	Other	2.706**	
Geographic location	Western (reference)		
	Appalachia	4.407***	4.407***
	Interior	(0.493)	(0.493)
Mine size	Small (reference)		
	Medium	3.750***	3.750***
	Large	(0.670)	(0.670)
Coal rank	Anthracite coal (reference)		
	Bituminous coal	-1.961***	-1.961***
		(0.639)	(0.639)
Coal seam thickness	Seam height >75" (reference)		
	Seam height ≤40"	-1.879***	-1.879***
	Seam height >40" and ≤75"	(0.563)	(0.563)
Model Parameters	Constant	7.383***	7.383***
		(1.075)	(1.075)
	Observations	1.416*	1.416*
	Year	1.397**	1.397**
	Wald Chi ²	(0.554)	(0.554)
	Constant	-21.027***	-21.027***
	Observations	(1.554)	(1.554)
	Year	29,707	29,707
	Wald Chi ²	1986-2018	1986-2018
		996.48***	996.48***

*** p<0.01, ** p<0.05, * p<0.1

Table 4 Summary of hypothesis testing

Hypothesis	Underground mine	Surface mine
H1: Workers in underground coal mines are more susceptible to CWP	Supported	
H2: The coal region contributes to CWP incidence rate	Supported	Supported only for Appalachia vs. Western
H3: Workers in smaller operations are more susceptible to lung diseases	Supported	Supported only for medium vs. small
H4: The coal rank contributes to the CWP incidence rate	Supported	Supported
H5: Workers in thin-seam mine operations are more susceptible to the CWP incidence	Supported	Supported

In order to examine the accuracy of the regression model results, VIF and homoscedasticity analyses were conducted. VIF identifies multicollinearity in the regression models. Multicollinearity exists when there is a correlation among independent variables in a regression model. The presence of multicollinearity will negatively influence the results of the estimation. This connection is, in other words, a concern since independent variables should be independent. Each of the VIF scores for the dataset were less than 5 (mean score of 1.55) (Table S3), indicating that lack of multicollinearity has been met.

The homogeneity of variance of the residuals is one of the main assumptions of GEE. For all predicted dependent variables, the variance of residuals is roughly equal. This makes the prediction of regression unbiased, consistent, and accurate [19, 30]. The existence of homoscedasticity was tested using Breusch-Pagan test (Table 5). The p-value was statistically significant at significance level of 0.01; therefore, the null hypothesis, which is the existence of homoscedasticity, was rejected. Table 5 shows the result of Breusch-Pagan's test. To account for heteroscedasticity, the robust standard error was used in GEE model [29, 36].

Table 5 Homoscedasticity test results

Test	Chi-square	Pr > ChiSq	Variables
Breusch-Pagan	192025.76	< 0.0001	Cross of all variables

4. Conclusions

In the United States, the increase in the rate of CWP in the mid-1990s has renewed the urge among medical and science researchers to investigate the primary root causes of the problem. No study has investigated the effect of all of the available mining factors in the prevalence of CWP among coal miners in a multivariable model. This study attempted to conduct such a comprehensive study using the longitudinal data collected from multiple credible sources such as MSHA accident/injury, and MSHA employee/production. First, a 33-years panel data analysis on the U.S. coal mines was conducted to determine the relationship between the CWP rate and mining parameters including mining operation type, geographic location, mine size, coal rank, and coal seam thickness. Second, for each type of mining operation, five hypotheses were developed to determine the relationship between the CWP rate and the contributing factors using a multiple linear regression model. The results of the GEE regression supported all of the hypotheses for underground coal mines. More specifically, mine operation, geographic location, mine size, coal seam thickness, and coal rank contribute to the prevalence of CWP among coal miners. In surface mines, mine size (only medium vs small size), geographic location (only Appalachia vs Western region), coal rank, and seam height (only medium-thickness vs thick-seam) contribute to the prevalence of CWP.

Declarations

Acknowledgments

This study is funded by the National Institute for Occupational Safety and Health (NIOSH) [75D30119C06390]. The authors would like to thank Dr. Arvin Ebrahimkhanlou for all precise technical comments related to statistical analysis.

References

1. Suarathana, E., A. S. Laney, E. Storey, J. M. Hale, and M. D. Attfield. (201). Coal workers' pneumoconiosis in the United States: Regional differences 40 years after implementation of the 1969 Federal Coal Mine Health and Safety Act. *Occupational and Environmental Medicine* 68(12):908-913.
2. Perret, J.L., Plush, B., Lachapelle, P., Hinks, T., Walter, C., Clarke, P., Irving, L., Brady, P. Dharmagei, S.C., Stewart, A. (2017). Coal mine dust lung disease in the modern era. *Asian Pacific Society of Respiriology*, 22, 662–670. doi: 10.1111/resp.13034.
3. National Academies of Sciences, Engineering and Medicine (NAS). (2018). Monitoring and sampling approaches to assess underground coal mine dust exposures. The National Academies Press. Washington DC, USA, (1-150). doi: <https://doi.org/10.17226/25111>.
4. Onder, M., Yigit, E. (2009). Assessment of respirable dust exposures in an opencast coal mine. *Environ Monit Assess* 152, 393. <https://doi.org/10.1007/s10661-008-0324-4>
5. Mine Safety and Health Administration (MSHA). (2014). Federal Register, Rules and Regulations, 79 Fed. Reg. 84, May 1, 2014.
6. Schatzel SJ. (2009). Identifying sources of respirable quartz and silica dust in underground coal mines in southern West Virginia, western Virginia, and eastern Kentucky. National Institute for Occupational Safety and Health, Pittsburgh Research Laboratory, Pittsburgh, USA.
7. Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R.W., Morozoff, C., Shirude, S., Naghavi, M., et al. (2017). Trends and patterns of differences in chronic respiratory disease mortality among US counties, 1980-2014. *Jama*. 318(12):1136–49.
8. Gamble J.F., Reger, R.B., Glenn, R.E. (2012). A Critical Review of Coal Workers Pneumoconiosis (CWP) and Coal Rank for Evaluation of Safe Exposure Levels in Coal Mining. *J Clinic Toxicol* S1:009. doi:10.4172/2161-0495.S1-009.
9. Laney, A.S., Petsonk E.L., Attfield, M.D. (2010). Pneumoconiosis among underground bituminous coal miners in the United States: Is silicosis becoming more frequent? *Occupational and Environmental Medicine* 67(10):652-656.
10. Blackley, D.J., Halldin, C.N., Wang, M.L., Laney, A.S. (2014). Small mine size is associated with lung function abnormality and pneumoconiosis among underground coal miners in Kentucky, Virginia and West Virginia. *Occupational Environment Medicine*. 71, 690–694.
11. Antao, V.C., Petsonk, E.L., Sokolow, L.Z., Wolfe, A.L., Pinheiro, G.A., Attfield, M.D. (2005). Rapidly progressive coal workers' pneumoconiosis in the United States: Geographic clustering and other factors. *Occupational and Environmental Medicine* 62(10):670-674. doi:10.1136/oem.2004.019679.
12. Antao, V., Petsonk, E.L., Attfield, M.D. (2006). Advanced cases of coal workers' pneumoconiosis—Two counties, Virginia, 2006. *Morbidity and Mortality Weekly Report* 55(33):909-913.
13. Sarver, E., Kelesa, C., Rezaee, M. (2019). Beyond conventional metrics: Comprehensive characterization of respirable coal mine dust. *International Journal of Coal Geology*. 207, 84-95. <https://doi.org/10.1016/j.coal.2019.03.015>.
14. Sarver, E., Keles, C., Lowers, H., Zulfikar, R., Zell-Baran, L., et al. (2020). Mineralogic Analysis of Respirable Dust from 24 Underground Coal Mines in Four Geographic Regions of the United States. *American Journal of Respiratory and Critical Care Medicine* 2020; 201:A2635. https://www.atsjournals.org/doi/abs/10.1164/ajrccm.conference.2020.201.1_MeetingAbstracts.A2635.
15. Laney, A.S., Attfield, M.D. (2010). Coal workers' pneumoconiosis and progressive massive fibrosis are increasingly more prevalent among workers in small underground coal mines in the United States. *Occupational and Environmental Medicine* 67(6):428-431.
16. Johann-Essex, V., Keles, C., Sarver, E. (2017). A computer-controlled SEM-EDX routine for characterizing respirable coal mine dust. *Minerals*, 7 (15). doi:10.3390/min7010015.
17. Gamble, J.F., Reger, R.B., Glenn, R.E. (2011). Rapidly progressing coal workers pneumoconiosis as a confounding risk factor in assessing coal mine dust safe exposure levels. *Journal of Clinical Toxicology*. S1:003. doi: 10.4172/2161-0495.S1-003.
18. (2008). Work-Related Lung Disease Surveillance System (eWoRLD) 2008-171. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Respiratory Health Division, Morgantown, WV [online]. Available at https://www.cdc.gov/eworld/Data/Respirable_coal_mine_dust_Geometric_mean_exposures_and_percent_exceeding_designated_occupational_exposure_limits_by_type_of_mine_MSHA_inspector_and_mine_operator_samples_1979-2003/171 (accessed February 22, 2017).
19. Shekarian, Y. (2020). An Investigation of the Effects of Mining Parameters on the Prevalence of Coal Worker's Pneumoconiosis (CWP) Risks among the US Coal Miners. Master's Thesis, New Mexico Institute of Mining and Technology.

doi: 13140/RG.2.2.22358.27205

20. Shekarian, M. (2019). An Examination of the Relationship among Network Design, Disruption, and Innovation in Airline Industry. Master's Thesis, North Carolina Agricultural and Technical State University.
21. Wackerly, D.D., Mendenhall, W., Scheaffer, R.L. (2008). *Mathematical statistics with applications*. Thomson Higher Education . 7th edition. Pages 939.
22. Frost, j. (2019). *Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries* Paperback. Statistics By Jim, PA, 16801, ISBN 978-1-7354311-0-9.
23. Benfratello, L. (2014). Random Effects Regression for Panel Data. In: Michalos A.C. (eds) *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_2402.
24. Bruin, J. (2006). *newtest: command to compute new test*. UCLA: Statistical Consulting Group. <https://stats.idre.ucla.edu/stata/ado/analysis/>.
25. Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2008). Generalized Estimating Equations for Longitudinal Data Analysis. In *Longitudinal data analysis* (pp. 57-92). Chapman and Hall/CRC.
26. Pang, M. S., (2017). "Politics and Information Technology Investments in the US Federal Government in 2003–2016," *Information Systems Research* 28(1), pp.33-45.
27. Shah, R., Ball, P. G., and Netessine, S. (2017). Plant Operations and Product Recalls in the Automotive Industry: An Empirical Investigation. *Management Science* (63:8), pp. 2439-2459.
28. Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*, 39(2), 81-90.
29. Pitselis G.A (2013). Review on robust estimators applied to regression credibility *Journal of Computational and Applied Mathematics*, 239 (2013), pp. 231-249. Available from: <http://dx.doi.org/10.1016/j.cam.2012.09.009>
30. Salkind, N. J. (2007). *Encyclopedia of measurement and statistics*. SAGE Publications, Inc. pp.1-416.
31. Shekarian, M., & Mellat Parast, M. (2020). An Integrative approach to supply chain disruption risk and resilience management: a literature review. *International Journal of Logistics Research and Applications*, 1-29.
32. Rahimi, E., Maghsoudi, A., & Hezarkhani, A. (2016). Geochemical investigation and statistical analysis on rare earth elements in Lakehsiyah deposit, Bafq district. *Journal of African Earth Sciences*, 124, 139-150.
33. Shekarian, N., Ramirez, R., Khuntia, J. (2020). The Impact of Data Analytics on Hospital Performance. *AMCIS 2020 Proceedings*. 3. https://aisel.aisnet.org/amcis2020/data_science_analytics_for_decision_support/data_science_analytics_for_decision_support/3/
34. Sellaro, R., Sarver, E. (2014). Characterization of respirable dust in an underground coal mine in central Appalachia. *Transactions of the Society for Mining, Metallurgy, and Exploration*. 336, 457-466.
35. Landen, D., Wassel, A., McWilliams, L., Patel, J. (2011). Coal dust exposure and mortality from ischemic heart disease among a cohort of U.S. coal miners. *American Journal of Industrial Medicine*, 54(10), 727-733.
36. Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach*. Nelson Education. fifth edition. pp.1-910.
37. Thakur, P. (2019). *Advanced mine ventilation: Respirable coal dust, combustible gas and mine fire control*. *Advanced Mine Ventilation*. 1st Edition, Elsevier Science, 105-136. doi.org/10.1016/B978-0-0.
38. Amandus, H. E., and G. Piacitelli. (1987). Dust exposures at U.S. surface coal mines in 1982-1983. *Archives of Environmental Health* 42(6):374-381
39. Graber, J. M., G. Harris, K. S. Almborg , C. S. Rose, E. L. Petsonk, R. A. Cohen. (2017). Increasing Severity of Pneumoconiosis Among Younger Former US Coal Miners Working Exclusively Under Modern Dust-Control Regulations. *Journal of Occupational Environmental Medicine* 59(6):e105–e111.
40. Sarver, E., Keles, C., & Rezaee, M. (2019). Characteristics of respirable dust in eight appalachian coal mines: A dataset including particle size and mineralogy distributions, and metal and trace element mass concentrations. *Data in Brief*, 25, [104032]. <https://doi.org/10.1016/j.dib.2019.104032>
41. Laney AS, Petsonk EL, Hale JM, et al. Potential determinants of coal workers' pneumoconiosis, advanced pneumoconiosis, and progressive massive fibrosis among underground coal miners in the United States, 2005–2009. *Am J Public Health* 2012;102:S279–83. 12
42. National Institute for Occupational Safety and Health. (2011). *Enhanced Coal Workers' Health Surveillance Program (ECWHSP)*.

Figures

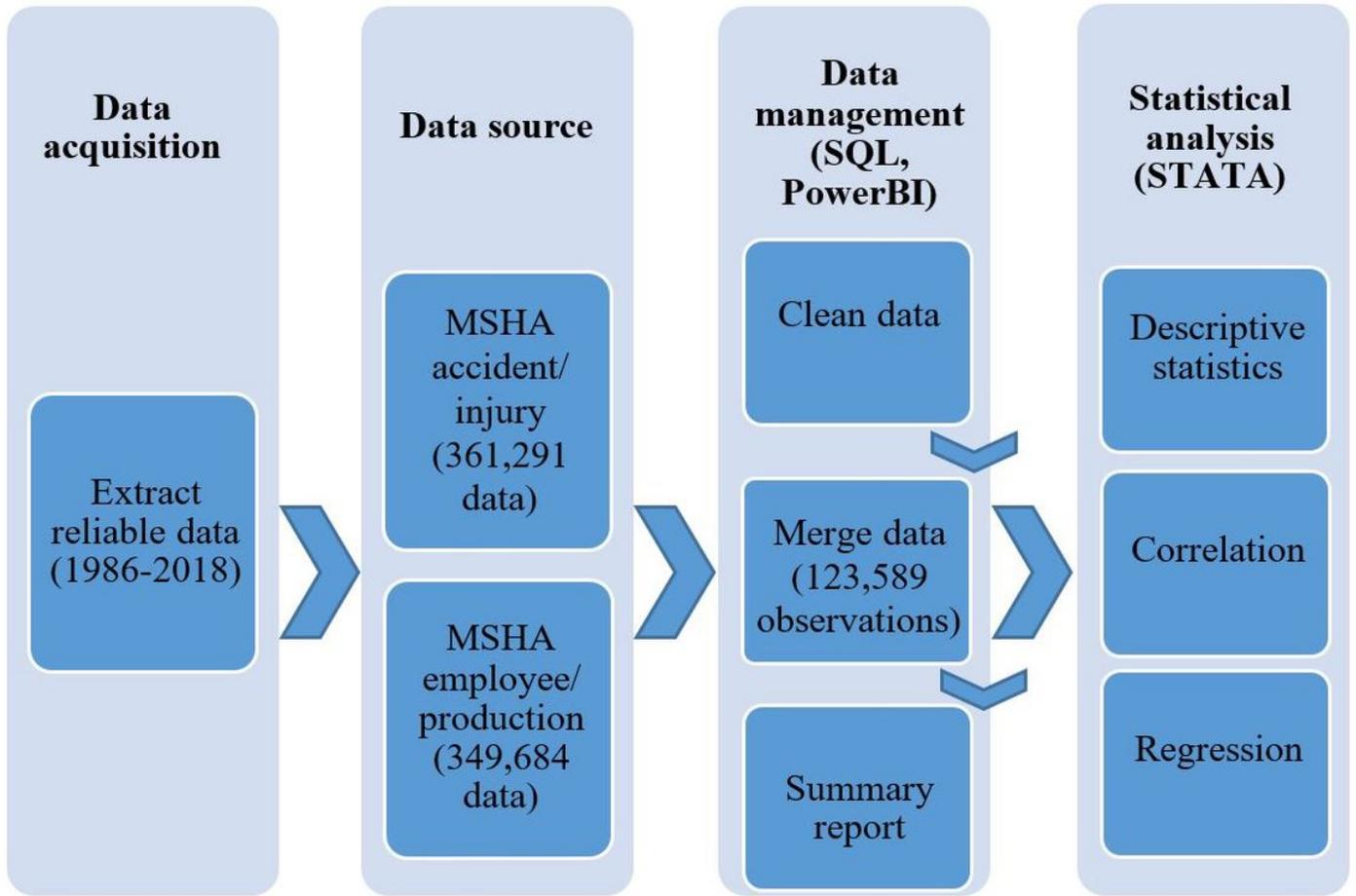


Figure 1

Summary of data management and methodology for statistical analysis

■ Underground mines ■ Surface mines ■ Other

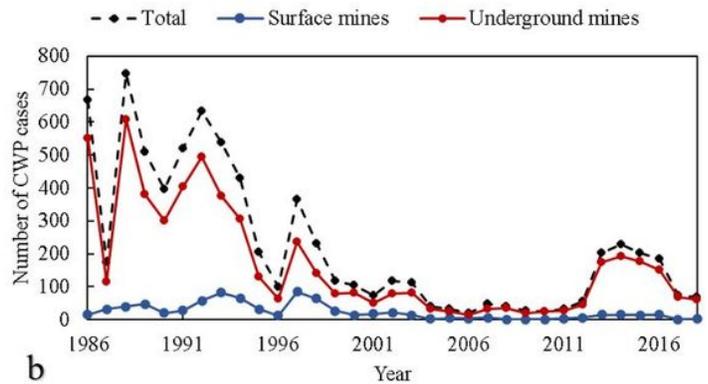
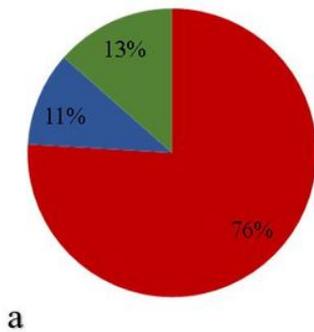


Figure 2

Number of CWP cases as function of (a) mine operation type and (b) year in Unites States during 1986-2018 (total number of CWP is 7337)

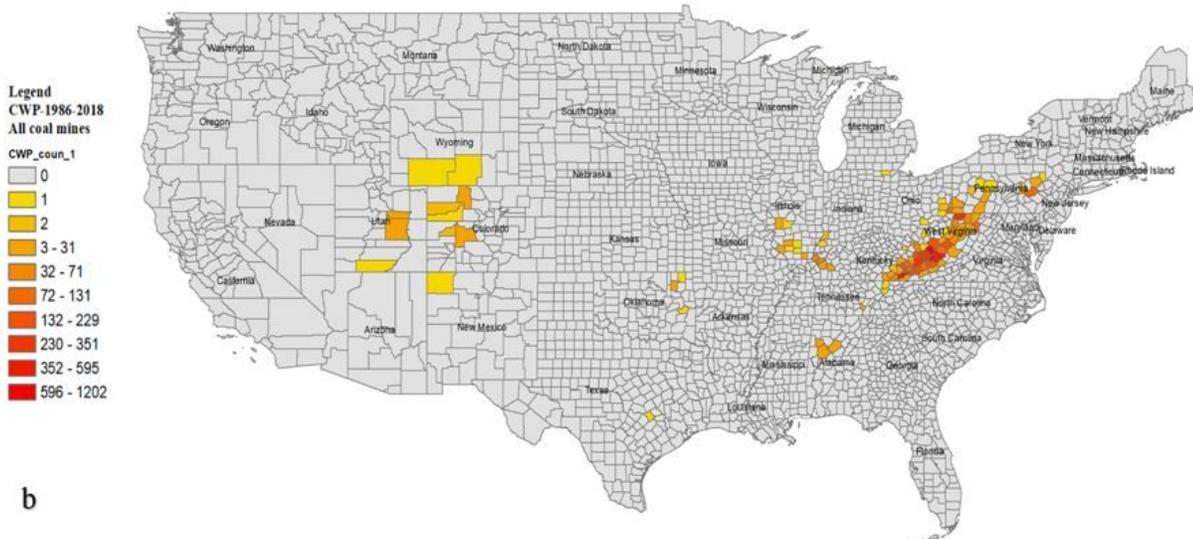
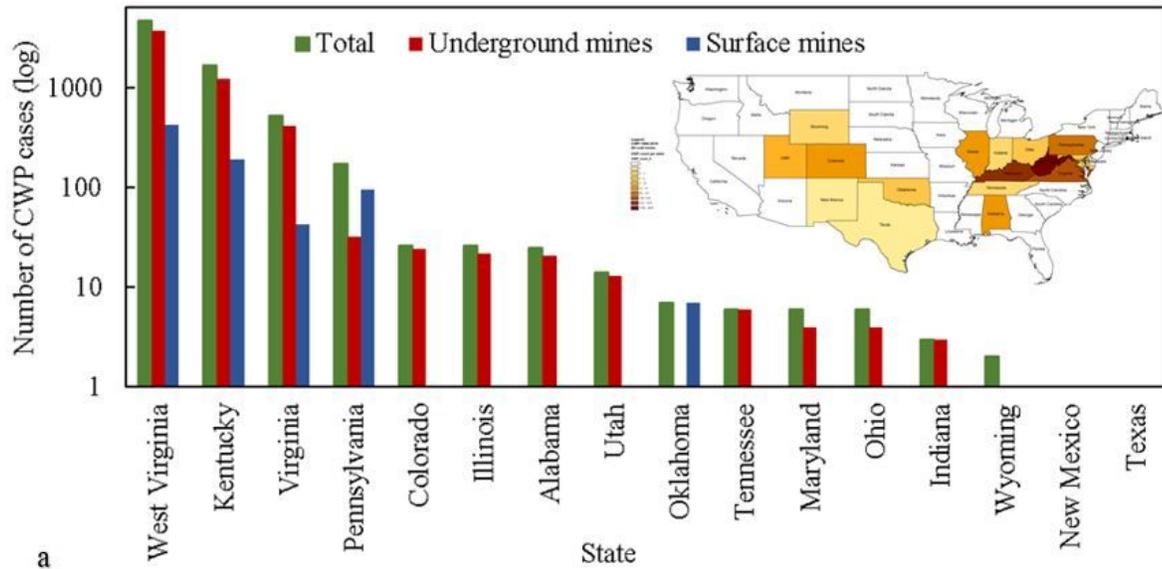


Figure 3

Distribution of CWP per state (a), and county (b) for underground, surface, and total data during 1986-2018 in the U.S. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

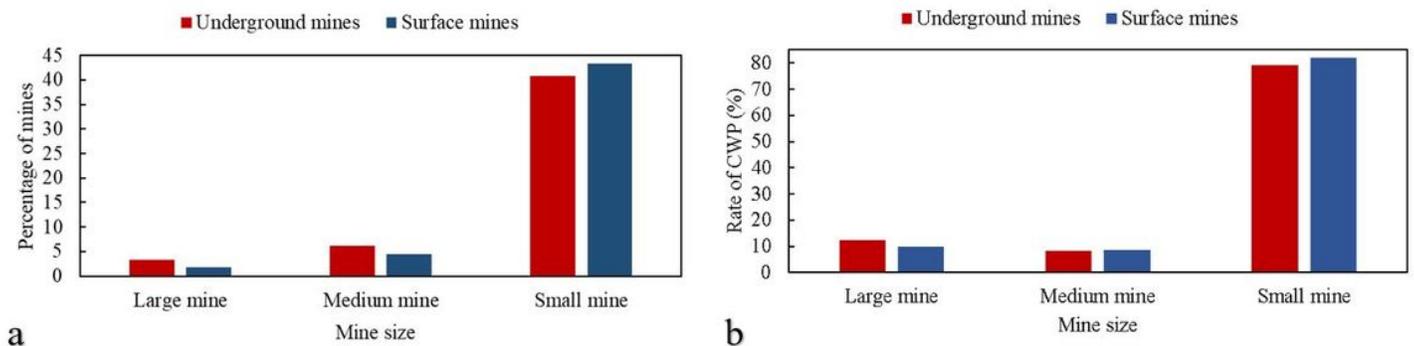


Figure 4

Percentage of underground and surface mines (a) and the percentage of CWP rate (b) per mine size in the U.S. during 1986-2018.

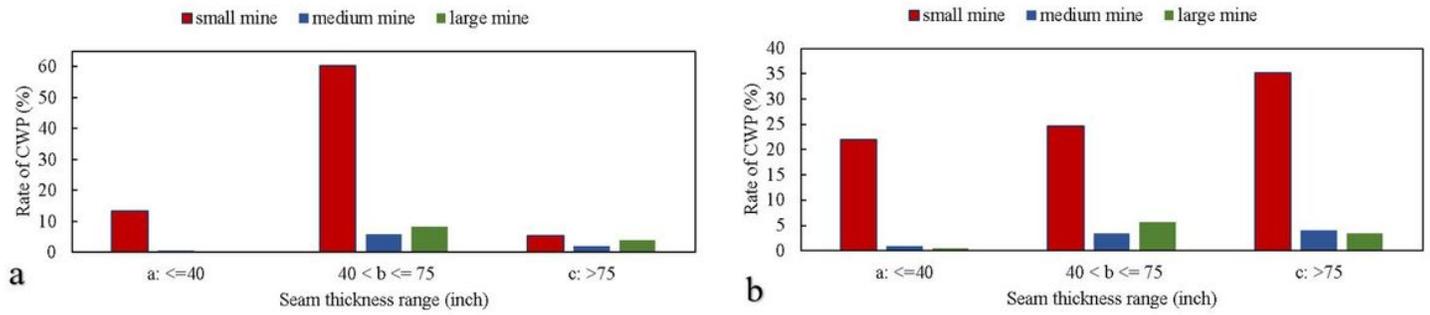


Figure 5

Percentage of CWP rate by seam thickness and mine size in the U.S. underground (a) and surface (b) mines during 1986-2018.

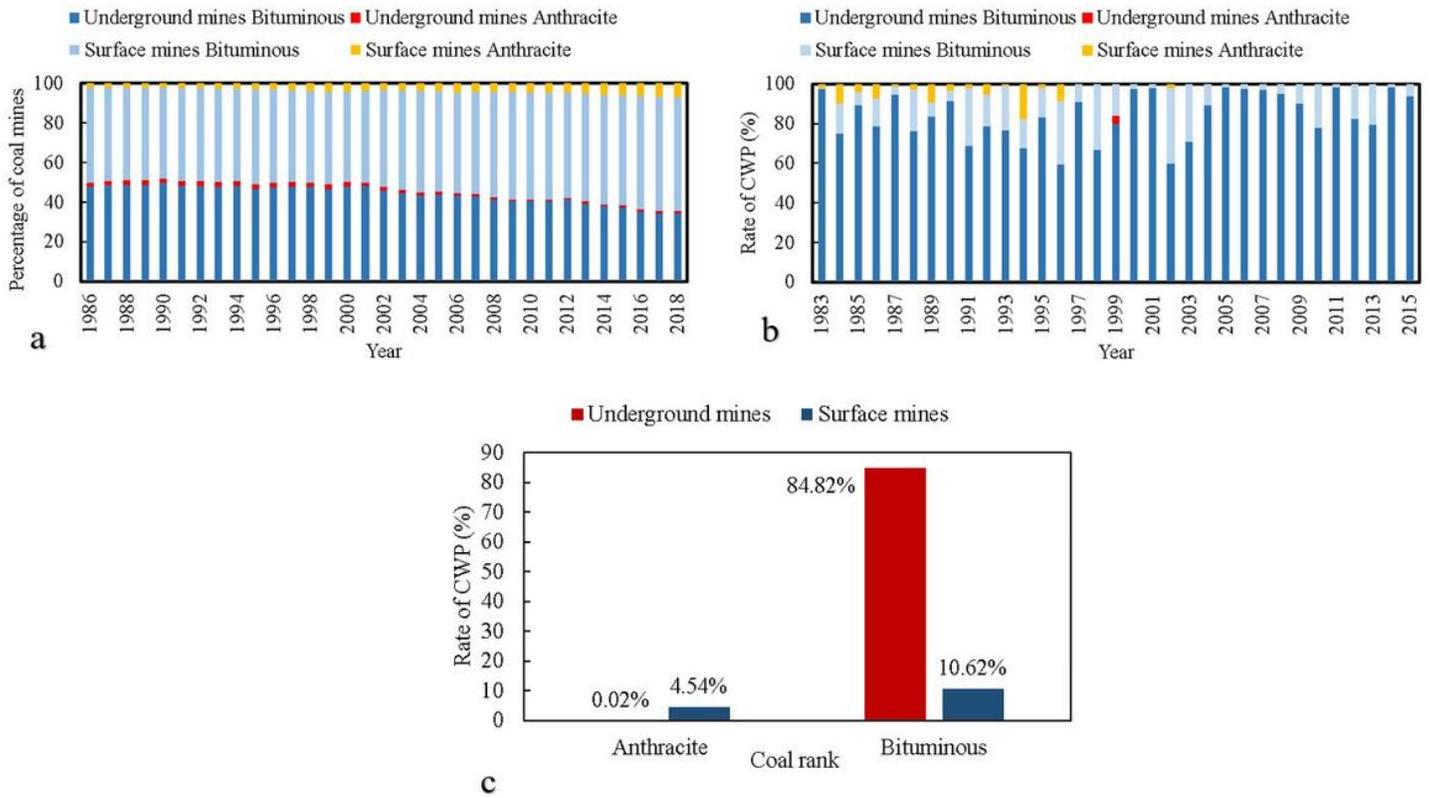


Figure 6

Number of underground and surface mines (a), and percentage of CWP rate per coal rank (b,c) in the U.S. during 1986-2018.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalInformation.docx](#)