Rates and Costs of Respiratory Illness in Coal Mining: A Cross-Industry Comparative Analysis

George Van Houtven, PhD, W. Randy Reed, PhD, Elyce A. Biddle, PhD, John C. Volkwein, PhD, Laurel Clayton, BA, and Eric Finkelstein, PhD

Objective: To estimate the prevalence and costs of respiratory illness for workers in coal mining, compared with other US industries. **Methods:** Using 5 years of insurance claims data for an annual average of 96,240 adult males, we model the probability and costs of respiratory illness as a function of workers' industry and other factors. **Results:** Controlling for nonindustry factors, workers in coal mining had significantly higher rates of respiratory illness claims (by 2.1% to 3.3% points) compared with other mining, agriculture, construction, and manufacturing. For coal mining workers with respiratory illness, annual medical care costs for these claims were also significantly higher (by $111 to $289). Surprisingly, drug costs were mostly lower (by $17 to $268). **Conclusions:** Our findings underscore the continued importance and potential cost effectiveness of measures to protect miners from harmful occupational exposures, particularly to coal dust.

Accumulated empirical evidence from the past 40 years strongly indicates that coal miners are not only at relatively high risk of pneumoconiosis (ie, black lung disease) but also of more broadly defined respiratory illness. For example, in a review of epidemiological studies conducted before 1998, Coggon and Taylor[1] found consistent evidence that occupational exposures to coal mine dust cause reductions in lung function and lead to increased risk of chronic obstructive pulmonary disease (COPD). More recent studies of coal mine workers in Great Britain[2] and in Spain[3] have continued to confirm these findings.

Given this evidence, two directly related and equally important questions are 1) what do these incremental health risks imply for the costs of respiratory illness among coal miners, and 2) how different are these costs compared with other industries? For coal miners, the average costs of respiratory illness are likely to be different than for other workers, because they experience higher rates of illness and because the average severity of these conditions may be different. Studies of the US population find the following:

1. Estimates of per capita medical care costs attributable to respiratory conditions, including asthma, COPD, pneumoconiosis, and asbestosis, range from $1500 to $3800 per year (converted from 1996 to 2006 dollars using the Consumer Price Index for medical services).[4]
2. As much as 20% of these conditions among adults are attributable to occupational exposures.[5]

Nevertheless, there is much less evidence about how the costs of respiratory illness differ across industries and occupations.

From the RTI International (Dr Van Houtven), Research Triangle Park, NC; National Institute of Safety and Health (NIOSH) (Dr Reed, Dr Volkwein), Pittsburgh, Penn; NIOSH (Dr Biddle), Morgantown, Va; University of North Carolina (Ms Clayton) Chapel Hill, NC; Duke University (Dr Finkelstein), Singapore.

Address correspondence to: George Van Houtven, RTI International, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle Park, NC 27709; E-mail: gvh@rti.org.

Copyright © 2010 by American College of Occupational and Environmental Medicine

DOI: 10.1097/JOM.0b013e3181de47e3

The purpose of this study is to estimate the prevalence and average annual medical care costs of respiratory conditions among individuals in the coal mining industry and to compare them with costs for similar individuals in other industries. To address this research objective, we analyze data on private health insurance claims more than 5 years (2002 to 2006) for adult males in the United States with employment in five selected industry categories.

DATA

The data for our analysis were compiled and provided to us by a large private US health insurance organization. The data were selected to compare outcomes among coal mine workers, who are almost exclusively male, with those from other blue-collar industries. Data on insured individuals and their claims were drawn separately for 5 successive years, from 2002 to 2006. In each year, a sample was drawn that included all adult male policy holders from the underground bituminous coal mining industry, from other mining industry categories, and from the agricultural and construction industries, including individuals who were retired but continued to be covered by their former employers' plans. Industries were identified according to claimants' Standard Industrial Classification codes—1222 for underground bituminous coal mining, 1011 to 1499 (excluding 1222) for other mining, 0111 to 0971 for agriculture, and 1521 to 1799 for construction. Data on claimants' occupation within these industries were not available; therefore, for example, the coal and other mining categories may include some individuals who do not physically work in the mines. In addition, for each year, a random sample of males twice the size of the coal mining sample was drawn from insured individuals in the manufacturing industry (Standard Industrial Classification codes 2011 to 3999).

The resulting yearly samples include an average of 96,400 claimants, ranging from 90,176 in 2002 to 101,718 in 2004. Roughly 0.5% of the original samples were excluded from the analysis because of missing data on individual characteristics. Although the data set was not designed to track the same cohort of individuals more than 5 years, many individuals are nonetheless included in more than 1 year of the data. In our initial data summaries, we treat each person in each year as a separate "person-year" observation, meaning that each claimant is treated as a separate entity from year to year (ie, the claimants from one year may or may not be the same individuals the following year). Pooling the data in this way resulted in a total sample size of 481,199 person-years during the 5-year period. In the regression analyses that follow, we correct the standard error estimates using a robust cluster method to specifically account for individuals who appear in more than 1 year of the data.

Tables 1 and 2 provide summaries of the data and allow for initial comparisons of the industry subsamples. More detailed and rigorous analyses of these data are described in the Methods and Results sections later in the article. Table 1 summarizes and compares characteristics of the insured individuals across industries. Because of the regional focus of the insurance company, more than half of the individuals in the data resided in Pennsylvania (in

Copyright © American College of Occupational and Environmental Medicine. Unauthorized reproduction of this article is prohibited.

TABLE 1. Summary of Sample Characteristics, by Industry Category

	Coal Mining	Other Mining	Manufacturing	Agriculture	Construction
Sample size (in person-yr)	91,241	71,043	182,539	10,090	126,286
Mean age	59.2	47.0	49.5	41.3	41.5
Percent married	85.3	62.3	65.7	45.6	55.4
Percent in HMO[a]	3.8	14.7	19.6	28.8	24.4
Percent in PPO[a]	66.6	66.3	58.1	63.7	70.8
Mean number of mo per year eligible for health insurance[b]	10.9	9.6	10.6	9.7	9.6
Mean value of median income in the individual's census tract	$31,233	$36,685	$42,214	$42,304	$42,930
Mean percentage of white residents in the individual's census tract	95.6	86.2	89.0	92.4	90.8
Percent in Northeast census region	24.9	34.7	48.3	92.0	73.7
Percent in South census region	54.2	56.2	23.8	7.2	12.7
Percent in Midwest census region	19.5	8.0	22.3	0.2	7.8
Percent in West census region	1.4	1.0	5.3	0.7	5.4

[a]The remaining percent (other than PPO and HMO) are predominantly in traditional fee-for-service plans.
[b]Number of mo that each individual policyholder is covered by their selected insurance plan in each year.

TABLE 2. Cross-Industry Comparison of Annual Rates and Average Costs of Respiratory Medical Claims

Respiratory Illness Definition	Coal Mining	Other Mining	Manufacturing	Agriculture	Construction
Claims for COPD					
Percent with 1 or more claims during year	13.7%	5.4%	5.0%	3.2%	3.2%
Average annual medical care costs[a,b]	$1330	$810	$920	$420	$400
Average annual increment in drug costs[c,d]	$1590	$1450	$760	$490	$630
Claims for RESP (COPD plus other respiratory conditions)					
Percent with 1 or more claims during year	17.3%	7.1%	6.7%	3.9%	4.0%
Average annual medical care costs[a,b]	$1480	$1050	$1130	$480	$600
Average annual increment in drug costs[c,d]	$1550	$1460	$820	$680	$740

[a]In 2006 dollars, adjusted using the consumer price index (CPI) for medical care services.
[b]Conditional on having at least one respiratory medical claim during the year.
[c]In 2006 dollars, adjusted using the CPI for medical care commodities.
[d]Difference in average annual drug costs between policy holders with and without respiratory illness claims.

the Northeast Census region) and West Virginia (in the South Census region); however, the data also include residents from all other 48 states. Demographic characteristics such as income and race were not available at the individual level; therefore, the median household income and percentage of white residents in the policy holder's Census tract (in 2000) were included instead.

Comparing these summary characteristics across industry categories reveals important differences, particularly for the individuals in the coal mining industry. For instance, the average age in the coal mining category (almost 60 years old) is substantially higher than in the other industries. These individuals are also more likely to be married and tend to live in areas with lower median income and a higher percentage of white residents. These differences across industries highlight the importance of including these characteristics as control variables in the statistical analysis. We do not necessarily have strong expectations regarding the effects of these characteristics on respiratory illness outcomes and costs; however, they provide the best available proxies (ie, controls) in our data for unobserved characteristics related to health status and behaviors. Unfortunately, no other individual-specific data on health status or behaviors (in particular smoking) were available for our sample. In the section "Discussion and Conclusions", however, we do discuss how separate data comparing smoking rates across industries can be used to interpret the role of these behaviors in our results.

For each person-year in the sample, data were also compiled describing the number, types, and costs of nondrug medical claims for selected conditions. It is important to note that these data only include claims that were submitted to the private insurance company. Claims that were only submitted to and fully paid by third parties, such as some claims to workers' compensation programs and Medicare, are not captured in these data. Claims for respiratory conditions were identified using International Classification of Diseases-ninth revision (ICD-9) system diagnosis codes. For our analysis, we defined the conditions in two ways. The narrower definition of respiratory condition (COPD) includes only the ICD-9 categories for COPD and allied conditions (ICD 490–496), which includes conditions such as bronchitis, emphysema, asthma, and other chronic airway obstructions. The broader definition (referred to in this article as RESP) is based on the respiratory illness definitions used in Petsonk and Parker,[6] Attfield et al,[7] and Jalloul and Banks. RESP included COPD plus claims for the following ICD-9 codes: 010–012 (primary tuberculosis infection, pulmonary tuberculosis, and other respiratory tuberculosis), 017.9 (tuberculosis of other specified organs), 031 (diseases caused by other mycobacteria), 415–416 (acute and chronic pulmonary heart disease), 500–506 (pneumoconioses and other lung disease caused by external agents, excluding pneumonitis [507] and respiratory conditions caused by other and unspecified external agents [508]), 511 (pleurisy), 515 (postinflammatory pulmonary fibrosis), 518.89 (other diseases of lung, not elsewhere classified), 519.8 (other diseases of the respiratory system, not elsewhere classified), and 714.81 (rheumatoid lung). For the analysis reported in this article,

© 2010 American College of Occupational and Environmental Medicine
Copyright © American College of Occupational and Environmental Medicine. Unauthorized reproduction of this article is prohibited.

we selected claims where the ICD-9 codes for COPD or RESP were listed as the primary diagnosis. We conducted additional analyses that also included claims whose secondary diagnosis was coded as respiratory illness (based on the selected ICD-9 categories). The fundamental findings of our analysis were similar under this alternative specification, and these results are available from the authors on request.

For these two definitions of respiratory illness, Table 2 compares annual rates of respiratory claims across industry categories. For the sample as a whole, the percentage of person-years with at least one claim for COPD is 6.2%. This estimate corresponds well with nationwide annual prevalence statistics, which indicate that ~6% of the adult population in 2000 had self-reported COPD.[9] For the broader definition of respiratory conditions (RESP), the annual rate for the full sample is 8%. Annual rates of RESP are 0.7% to 1.7% points higher than for COPD in the other mining, agriculture, construction, and manufacturing categories. A notable exception is coal mining where the rate is 3.6% points higher.

Comparing across industry categories, the annual rates of respiratory claims are systematically higher in coal mining than in all other categories, regardless of how respiratory conditions are defined. Annual rates of COPD claims are 13.7% in coal mining, with the next highest rates found in the other mining category (5.4%). For both definitions of respiratory illness, the lowest annual rates of claims are in the agriculture and construction industries. The magnitude and statistical significance of the cross-industry differences summarized in Table 2 are more systematically analyzed in the regression analyses that follow.

The data also provide detailed information on the costs per respiratory illness claim, including the amount paid by the insured individual (through copayment or deductible), the amount paid by the insurer, and the amount paid by third parties. Aggregating these individual claim costs in each year, we estimated the annual costs of respiratory illness claims for each person-year. The resulting values are also summarized in Table 2 for the five industry categories and the two respiratory condition definitions. Roughly 1% total annual medicals costs were equal to zero, because of adjusted or nonreimbursed claims. These zero values are excluded from the average annual cost estimates.

Comparing again across industry categories, the average annual costs of respiratory illness (conditional on at least one claim during the year) are also systematically higher in coal mining than in all other categories. Annual costs for COPD claims are $1330 in coal mining, compared with, for example, $920 in manufacturing and $400 in construction. Using the broader definition (RESP), the number and costs of claims are higher across all industry categories, but they show a similar pattern across industries. The annual costs for RESP claims are $1480 in coal mining, compared with $1130 in manufacturing and $600 in construction.

In addition to the nondrug medical claims information described above, data on annual prescription drug claims were separately acquired for the subsample of individuals with drug coverage during the selected years (319,107 person-years in total). In contrast to the medical claims database, the prescription drug database has unique limitations. The drug claims cannot be directly matched to ICD-9 respiratory illness diagnosis codes; therefore, the costs of all drug claims (regardless of health condition) were aggregated for each person-year. For each industry category, Table 2 reports the difference in average annual drug costs for individuals with and without at least one nondrug respiratory illness claim during the year. As expected, in all industries the average drug costs are higher for persons with at least one respiratory claim (ie, the differences are all positive).

As was found for the nondrug medical costs, the incremental drug costs associated with respiratory illness are higher in the coal mining industry. For instance, individuals in coal mining with at least one COPD claim have annual drug costs that are on average $1590 higher than those without these claims. In contrast, this difference is $630 in the construction industry and $760 in manufacturing.

METHODS

Although the summary statistics reported in Tables 2 provide evidence of systematic differences in rates and costs of respiratory illness across industries, we cannot conclude from these results that the observed differences are attributable to occupational exposures. Absent specific data on these types of exposures, we instead apply multivariate regression methods to approximate the effect of occupation on respiratory illness by controlling for other potentially influential variables (in particular, those summarized in Table 1).

To estimate the specific effects of coal mining on individuals' rates and costs of respiratory illness, we apply a two-part regression modeling approach (see, for example, Leigh and Fries[10] for a similar medical cost modeling approach). The first part uses a discrete dependent variable framework to model individuals' annual probability of treatment for respiratory illness. The second part uses a continuous ordinary least squares (OLS) framework to model individuals' annual costs of respiratory illness. Because of data differences, medical care costs and drug costs were analyzed separately in this second part of the two-part analysis.

Using the results of the second part models, we estimate the effect of coal mining on conditional expected costs (CEC) of respiratory illness. In other words, for individuals with respiratory illness, we estimate differences in average (per capita) annual costs of respiratory illness across industries, controlling for other individual characteristics such as age, marital status, and other factors listed in Table 1.

Combining results from the first and second part models, we then estimate the effect of coal mining on unconditional expected costs (UEC; ie, the average costs of respiratory illness across all individuals [including those without respiratory illness]). Examining unconditional costs provides a broader assessment of the effects of coal mining because it combines 1) its effect on the probability of respiratory illness and 2) its effect on conditional costs of respiratory illness.

First Part Model—Probability of Respiratory Illness

To be more specific, in the first part we model the discrete illness outcome (Y), which is set equal to 1 if an individual is treated for respiratory illness during the year (otherwise, $Y = 0$). Applying a probit statistical model to the full sample, we estimate how the probability of respiratory illness—$P_Y(X, Z)$—is affected by a vector of explanatory factors (X) plus a coal mining industry indicator ($Z = 1$). In particular, we estimate the incremental effect of being in the coal mining industry (relative to other industries) on an individual's probability of respiratory illness. These models were estimated using the "dprobit" command in Stata (use of the Stata data analysis and statistical software package does not imply endorsement). The main advantage of this model compared with a logistic regression model is that it provides coefficient estimates that can be directly interpreted as the marginal effect of each explanatory variable on the probability of a respiratory illness claim during the year. Otherwise, the statistical results were virtually identical to those from logistic models. Because certain individuals appear in the data during multiple years, the standard errors and reported significance levels were adjusted to account for within cluster (ie, within individual) correlation across years, using the robust cluster estimator in Stata.[11]

© 2010 American College of Occupational and Environmental Medicine
Copyright © American College of Occupational and Environmental Medicine. Unauthorized reproduction of this article is prohibited.

Second Part Model 1—Medical Care Costs

We then model the continuous outcome (C)—the annual nondrug medical care cost of treating the individual's respiratory illness—using OLS regression. To account for the rightward skew in the distribution of these annual costs, they were converted to logarithmic form. Because positive costs are only observed for individuals with treatment claims ($Y = 1$), we estimate this model using the subsample for whom $C > 0$. With this model, we can estimate conditional expected medical care costs—$E(C \mid Y, X, Z)$—for individuals with respiratory illness ($Y = 1$), controlling for industry (Z), and other characteristics (X). The incremental effect of coal mining on these CEC can be expressed as:

$$\Delta \text{CEC} = E(C \mid Y = 1, X, Z = 1) - E(C \mid Y = 1, X, Z = 0). \quad (1)$$

To appropriately "retransform" the model-predicted logarithmic costs into dollar terms, we exponentiate the log prediction and apply a smearing factor to the model-predicted values.[12] The smearing factor is equal to the average of the exponentiated residuals from the OLS estimation.

By combining the results from the probit and OLS models, we then estimate the incremental effect of coal mining on UEC, which can be expressed as:

$$\Delta \text{UEC} = P_Y(X, Z = 1) \times E(C \mid Y = 1, X, Z = 1) - P_Y(X, Z = 0) \times E(C \mid Y = 1, X, Z = 0). \quad (2)$$

Second Part Model 2—Drug Costs

Because of differences in the data structure, a somewhat modified approach is required to estimate the effect of coal mining on drug costs for respiratory illness. Information on drug claims and costs is only available for individuals with drug insurance coverage, and, as noted above, they cannot be directly linked to respiratory illness. Instead, the total number and costs of drug claims (regardless of condition or diagnosis) were aggregated for each insured individual in each year.

To account for policy holders who had drug insurance coverage but did not submit drug claims during the year (ie, zero drug costs), we embedded another two-part model within the second part analysis of drug costs. We first apply a probit regression to model the probability of at least one drug claim ($D = 1$) of any type during the year, $P_D(Y, X, Z)$. For those with $D = 1$, we then apply an OLS framework to model the magnitude of total annual drug costs (DC). Both parts include indicators of respiratory illness (Y) and coal mining industry (Z) as explanatory variables.

Combining results from this probit and OLS model, the conditional expected drug costs (CEDC) attributable to respiratory illness can be expressed as

$$\text{CEDC}(X, Z) = P_D(Y = 1, X, Z) \times E(\text{DC} \mid D = 1, Y = 1, X, Z) - P_D(Y = 0, X, Z) \times E(\text{DC} \mid D = 1, Y = 0, X, Z). \quad (3)$$

In other words, CEDC is calculated as the increment in expected total drug costs between individuals with and without respiratory illness (controlling for other characteristics X and Z).

The next question is whether these drug costs attributable to respiratory illness differ across industries. The effect of coal mining on these conditional drug costs can be expressed as follows:

$$\Delta \text{CEDC} = \text{CEDC}(X, Z = 1) - \text{CEDC}(X, Z = 0). \quad (4)$$

This "difference-in-difference" expression measures how the increase in expected annual drug costs because of respiratory illness (CEDC) differs for coal mining compared with other industries ($Z = 1$ vs $Z = 0$). To account for the rightward skew in the distribution of these annual costs, the dependent variables in the OLS models of DC were also expressed in logarithmic terms.

Finally, this second part model for drug costs can also be combined with the first part model of respiratory illness claims to estimate the incremental effect of coal mining on unconditional expected drug costs (UEDC):

$$\Delta \text{UEDC} = P_Y(X, Z = 1) \times \text{CEDC}(X, Z = 1) - P_Y(X, Z = 0) \times \text{CEDC}(X, Z = 0). \quad (5)$$

As before, unconditional costs are equal to conditional costs multiplied by the probability of illness.

RESULTS

Table 3 reports the results of the first part probit regressions using the two different definitions of respiratory illness for the dichotomous dependent variable. The probit models include four industry categorical variables, leaving the coal mining industry as the excluded reference category. All model results indicate that, compared with the other four industries, the coal mining industry has a positive and statistically significant effect on the probability of respiratory illness (ie, the change in P_Y is positive and significant). For instance, relative to the construction industry, coal mining carries annual risks that are higher by 2.6% to 3.2% points, and compared with other mining, it has risks that are higher by 1.7% to 2.1% points.

Other control variables are also found to have significant effects on the probability of respiratory illness claims. As expected, aging increases the probability of illness, and the positive coefficient on the age squared term indicates that it does so at an increasing rate. Those living in higher income areas tend to have lower rates of illness, whereas those in areas with a higher percentage of white residents have higher rates. Individuals in preferred provider organization (PPO) and health maintenance organization (HMO) plans also tend to have significantly lower rates compared with those in traditional indemnity plans (the omitted category).

Corresponding to the two first stage probit models, Table 3 also reports results for two second stage OLS models. The dependent variables are the annual costs of medical care services for respiratory illness. The data used to estimate these models only included person-years with 1) at least one respiratory illness claim during the year—(as defined by the COPD and RESP)—and 2) positive annual costs associated with each of these claim definitions. To analyze variation in annual medical costs, the OLS models use the same set of explanatory variables as the probit models. As in the probit models, the reported significance levels for the coefficient estimates were adjusted to account for within-individual correlation.

Compared with other industries and controlling for other factors such as age, the coal mining industry is found to have significantly larger per capita annual medical costs for individuals with respiratory illness. The smallest differences are found with respect to other mining. As for the control variables, aging is found to increase the average annual costs of respiratory illness, whereas being married or living in an area with a higher percentage of white residents is associated with lower annual costs. Compared with traditional indemnity plans, individuals with respiratory illness claims who are in HMO plans (and, to a lesser extent, those in PPO plans) tend to have significantly lower annual costs, which suggests that these nontraditional plans are relatively effective with cost containment for treatment of respiratory illness.

To directly compare the magnitudes of medical costs across industries, we used the OLS model results to first predict conditional annual medical costs (CEC) for individuals with specific characteristics. These estimates are reported in Table 4, and the values used for the independent variables in these predictions are

Copyright © American College of Occupational and Environmental Medicine. Unauthorized reproduction of this article is prohibited.

TABLE 3. Factors Explaining Annual Probability of Respiratory Illness (Probit) and Annual Medical Care Costs (OLS)[a]

	Probit Model (for P_Y)		OLS Model (for ln(C))[c]	
Explanatory Variables[b]	**COPD**	**RESP**	**COPD**	**RESP**
Industry (reference = coal mining)				
Construction	−0.026**	−0.032**	−0.189**	−0.204**
Manufacturing	−0.028**	−0.033**	−0.187**	−0.225**
Agriculture	−0.02**	−0.030**	−0.250**	−0.326**
Other mining	−0.017**	−0.021**	−0.159**	−0.118**
Year (reference = 2002)				
2003	0.004**	0.005**	0.033	0.034
2004	0.003**	0.004**	0.086**	0.092**
2005	0.007**	0.009**	0.094**	0.135**
2006	0.002*	0.005**	0.146**	0.146**
Demographic				
Age	0.001**	0.001**	0.011**	0.031**
Age squared	7.7e–06**	9.8e–06*	3.0e–05	−1.8e–04**
Married	0.001	0.002	−0.065*	−0.057
Median income (census tract)[d]	−3.9e–04**	−4.0e–04**	−9.1e–04	−4.6e–05
Percent white (census tract)	0.021**	0.021**	−0.279**	−0.148
Insurance (reference = traditional)				
HMO	−0.015**	−0.017**	−0.661**	−0.554**
PPO	−0.014**	−0.015**	−0.195**	−0.174**
Months of eligibility	0.005**	0.006**	0.010*	0.014**
Number of observations	481,199	481,199	29,381	38,013
R-squared	n/a	n/a	0.068	0.040

[a]Statistical significance, two-tailed test: *$P < 0.05$; **$P < 0.01$.
[b]Categorical variables for the 14 states with the most observations were also included in model but not reported here.
[c]Semi-log functional form: dependent variable is natural logarithm of annual costs for respiratory illness.
[d]In thousands of 2006 dollars.

TABLE 4. Predicted Annual Medical Care Costs of Respiratory Illness[a,b]

Respiratory Illness Definition	Coal Mining	Other Mining	Manufacturing	Agriculture	Construction
CEC: conditional on having at least one claim in year					
COPD	$759 (687–839)	$648 (588–713)	$630 (576–689)	$592 (507–682)	$629 (582–681)
RESP	$1041 (942–1150)	$925 (850–1013)	$831 (764–906)	$752 (647–866)	$849 (785–924)
UEC: unconditional					
COPD	$58 (52–64)	$34 (30–37)	$26 (24–29)	$26 (21–31)	$27 (24–30)
RESP	$103 (93–113)	$64 (58–72)	$48 (44–53)	$41 (34–49)	$48 (44–53)

[a]In 2006 dollars. Predictions are based on individuals with the following selected characteristics: age = 50 yr, year = 2006, state = Pennsylvania, married = 0.66, median income = $40,000, percent white = 90%, insurance = PPO, and mo of eligibility = 12.
[b]Bootstrapped 95% confidence intervals in parentheses.

reported in the footnotes of the Table 4. For an individual with at least one respiratory illness claim, the predicted annual medical costs for COPD are $759 in the coal mining industry, compared with as little as $592 in agriculture and $629 in the construction and manufacturing industries. Predicted annual costs are substantially higher in all industries for the broader definition of respiratory illness (RESP), but again they are highest in coal mining ($1041) and lowest in agriculture, manufacturing, and construction ($752 to $849).

Table 4 also reports the predicted unconditional annual medical costs (UEC), which multiply the CEC estimates by corresponding predicted probability estimates based on the probit models shown in Table 3. For individuals in the coal mining industry (with the selected characteristics defined in the table footnotes), expected annual costs of COPD are $58. Because of both lower rates of illness and lower conditional costs, these unconditional costs are considerably lower (between $26 and $34) in the manufacturing, agriculture, construction, and other mining categories. A similar pattern emerges for RESP, with UEC of $103 per year in coal mining compared with $41 to $64 in the other four industries. Annual unconditional costs in the other mining industry generally fall between coal mining and the other three industries.

The analysis of drug-related costs reveals a pattern across industries that is different from the one observed for medical

Copyright © American College of Occupational and Environmental Medicine. Unauthorized reproduction of this article is prohibited.

TABLE 5. Factors Explaining Annual Probability of Drug Claims (Probit) and Total Annual Costs of Drug Claims (OLS)[a]

	Probit Model (for P_D)		OLS Model (for ln(DC))[c]	
Explanatory Variables[b]	**COPD**	**RESP**	**COPD**	**RESP**
Respiratory Illness				
Reference category (coal mining)	0.253**	0.253**	0.641**	0.619**
Respiratory illness* construction[d]	0.020	0.008	0.146**	0.193**
Respiratory illness* manufacturing[d]	−0.021	−0.028*	0.163**	0.178**
Respiratory illness* agriculture[d]	0.054	0.048	0.068	0.111
Respiratory illness* other mining[d]	0.110**	0.102**	0.098*	0.102*
Industry (reference = coal mining)				
Construction	−0.104**	−0.102**	−0.244**	−0.246**
Manufacturing	−0.072**	−0.069**	−0.146**	−0.146**
Agriculture	−0.084**	−0.082**	−0.084*	−0.081*
Other mining	−0.035**	−0.034**	−0.026	−0.026
Year (reference = 2002)				
2003	0.074**	0.074**	0.0236*	0.023*
2004	0.066**	0.066**	0.0916**	0.090**
2005	0.095**	0.095**	−0.042**	−0.042**
2006	0.170**	0.170**	0.0982**	0.097**
Demographic				
Age	0.009**	0.009**	0.107**	0.108**
Age squared	−2.6e–05**	−2.9e–05**	−5.8e–04**	−5.9e–04**
Married	0.067**	0.067**	0.002	0.003
Median income (census tract)[e]	2.8e–04**	2.9e–04**	1.6e–03**	1.6e–03**
Percent white (census tract)	0.088**	0.088**	0.225**	0.226**
Insurance (reference = traditional)				
HMO	0.163**	0.164**	−0.422**	−0.421**
PPO	0.056**	0.057**	−0.404**	−0.405**
Months of eligibility	0.029**	0.028**	0.092**	0.091**
Number of observations	319,107	319,107	195,336	195,336
R-squared	n/a	n/a	0.275	0.277

[a]Statistical significance, two-tailed test: * = $P < 0.05$; ** = $P < 0.01$.
[b]Categorical variables for the 14 states with the most observations were also included in model but not reported here.
[c]Semi-log functional form: dependent variable is natural logarithm of annual costs for all drug claims by the insured individual.
[d]Interaction term, multiplying respiratory illness indicator by the industry category dummy variable, which captures the incremental effect of respiratory illness in the industry relative to the incremental effect in coal mining (the reference category).
[e]In thousands of 2006 dollars.

service costs. Drug costs were analyzed using the subsample of individuals in each year with a drug coverage plan. To estimate the portion of total drug costs that are attributable to respiratory illness, we first analyzed the determinants of a dichotomous outcome (ie, did the individual have at least one drug claim for any illness during the year?) and then analyzed the continuous outcome (ie, if so, what were total drug costs for the year?). In both cases, we specifically investigate the effect of respiratory illness on the outcomes.

Table 5 reports the results of two probit analyses of the probability of at least one drug claim (P_D). The only difference between these two models is how respiratory illness is defined (COPD or RESP). As expected, controlling for other explanatory factors, the presence of respiratory illness has a strong, positive and statistically significant effect on the probability of submitting at least one drug claim during the year. The effect size in the coal mining industry (the reference category) is estimated to be a 25.3% point increase in probability for both COPD and RESP.

By including interaction terms that multiply the respiratory illness indicator by each of the other four industry categorical variables, the probit models also examine whether there are cross-industry differences in the effect of respiratory illness on the probability of a drug claim. In comparison with coal mining, the effect of respiratory illness on the probability of a drug claim is not statistically different in the three industries of construction, manufacturing, and agriculture; however, the effect of respiratory illness on the probability of a drug claim is found to be statistically significantly larger (by 10% to 11% points) for other mining in comparison with coal mining. This result indicates that there are other factors—not included in the regression but associated with other mining—that influence this effect, but it is not known what these factors are.

The other control variables show that there are significant differences across industries in rates of at least one drug claim of any type during the year. Coal mining increases the rate of at least one drug claim by 3.4% to 3.5% points compared with other mining and by more than 10% points compared with construction. Age, marital status, median income, percent white, insurance type, and duration of eligibility are also consistently found to have positive and significant effects on the probability of a drug claim during the year.

Table 5 also reports the results of two continuous OLS models, which analyze total annual drug costs (in logarithmic terms) and the effects of respiratory illness (COPD and RESP) on

Copyright © American College of Occupational and Environmental Medicine. Unauthorized reproduction of this article is prohibited.

TABLE 6. Predicted Incremental Annual Drug Costs Associated with Respiratory Illness, by Industry Category[a]

Respiratory Illness Definition	Coal Mining	Other Mining	Manufacturing	Agriculture	Construction
CEDC: conditional on having at least one respiratory medical claim during the year					
COPD	$1194 (1128–1268)	$1462 (1306–1596)	$1331 (1208–1462)	$1291 (918–1686)	$1211 (1111–1333)
RESP	$1125 (1067–1190)	$1388 (1266–1504)	$1280 (1178–1391)	$1304 (959–1663)	$1216 (1127–1325)
UEDC: unconditional					
COPD	$91 (83–98)	$76 (67–84)	$55 (50–61)	$56 (40–77)	$51 (46–57)
RESP	$111 (103–119)	$97 (88–105)	$74 (68–81)	$71 (52–94)	$69 (63–76)

[a]In 2006 dollars. Predictions are based on individuals with the following selected characteristics: age = 50 yr, year = 2006, state = Pennsylvania, married = 0.66, median income = $40,000, percent white = 90%, insurance = PPO, and mo of eligibility = 12.
[b]Bootstrapped 95% confidence intervals in parentheses.

these costs. As in the probit models, respiratory illness is found to have a strong, positive, and statistically significant effect. For individuals in the coal mining industry (reference category) with at least one drug claim during the year, respiratory illness has a positive and statistically significant effect on annual drug costs; however, the incremental effect of respiratory illness on annual total drug costs is significantly higher for individuals in manufacturing, construction and other mining.

The other control variables show that there are again significant differences across industries. The coefficients for the industry variables measure cross-industry differences for individuals without respiratory illness claims, and they indicate at these drug costs are higher in coal and other mining compared with the other three categories. Of the demographic variables, age and areas with higher average incomes and more white residents have consistently significant (positive) effects on annual drug costs. HMO and PPO have significantly lower annual drug costs compared with traditional plans, which again suggests that these nontraditional plans are relatively effective with cost containment for treatment of respiratory illness.

Combining the probit and OLS model results reported in Table 5, we can also predict and compare annual drug costs across industries for individuals with the same selected characteristics. The interaction terms in these models were used to separately predict, for each industry category, the incremental effect of respiratory illness on annual drug costs. These predictions are summarized in Table 6. The presence of respiratory illness is estimated to increase average annual drug costs by $1125 and $1194 in the coal mining industry for RESP and COPD, respectively, compared with, for example, $1280 and $1331 in manufacturing and $1216 and $1211 in construction for RESP and COPD, respectively. Unlike the differences found for the medical service costs associated with respiratory illness (reported in Table 4), the annual drug costs for respiratory illness are lowest in the coal mining industry.

Combining these results with the probit model results from Table 3 (ie, probability of respiratory illness in each industry), we can also predict and compare unconditional expected annual drug costs attributable to respiratory conditions. For an individual in coal mining (with other characteristics as defined in the footnotes of Table 6), these unconditional drug costs are estimated to be $91 and $111 for COPD and RESP, respectively. In contrast, in agriculture, construction, and manufacturing, these UEC are all lower, varying from $61 to $97 for RESP and from $51 to $76 for COPD. In other words, the lower probabilities of respiratory illness in these industries more than offset their higher conditional drug costs.

DISCUSSION AND CONCLUSIONS

To appropriately design and evaluate policies that protect coal miners from respiratory illnesses associated with occupational exposures, it is useful to understand the relative risks and health costs faced by these workers. This study adds to this knowledge in several respects.

First, it provides additional evidence that even with the advances that have been made and current protections in place, workers in the coal mining industry continue to face significantly higher risks of COPD (and related respiratory illnesses) than similar workers in other industries. Using a large and unique data set on private health insurance claims for male workers, and controlling for individual characteristics, we find that being in coal mining increases the annual rates of at least one health insurance claim for COPD respiratory illness by 1.7% to 2.8% points, compared with manufacturing, construction, agriculture, and other mining. For respiratory illness defined as RESP, rates in the coal mining industry are between 2.1% and 3.3% points higher. Compared with the observed rates in the noncoal mining sample (3% to 5% for COPD and 4% to 7% for RESP as shown in Table 2), these differences are substantial.

Second, it provides new evidence regarding both the absolute and relative magnitudes of medical costs because of respiratory illness. Tables 4 and 6 present the predicted annual medical care cost of respiratory illness and predicted incremental annual drug cost associated with respiratory illness, respectively. Controlling again for individual characteristics, we find that the average annual medical costs attributable to respiratory illness are also higher in coal mining. Using Table 4, for medical care services, we estimate annual costs for an individual in coal mining (conditional on having COPD) that are $129 to $167 higher than a comparable individual in manufacturing, agriculture, or construction. For individuals with at least one RESP claim, the costs are $190 to $289 higher; however, using Table 6, we estimate that annual drug costs for an individual in coal mining are $17 to $137 lower for COPD and $91 to $179 lower for RESP when compared with the manufacturing, agriculture, or construction industries. The other mining category tends to have lower medical costs ($111 and $116 for COPD and RESP, respectively) but higher drug costs ($268 and $263 for COPD and RESP, respectively).

The precise reasons for the reverse in ranking of the conditional drug costs compared with the medical costs are difficult to disentangle with these data; however, it must be emphasized that the drug costs attributable to respiratory illness are estimated with much less precision than the medical costs. Although nondrug medical claims are directly linked to respiratory illness through ICD-9 codes, the attribution of drug costs to respiratory illness must be done indirectly by looking at differences in total drug costs between those with and without medical claims for respiratory illness. The cross-industry ordering of nondrug medical costs alone suggests that the average severity of respiratory conditions (conditional on

 © 2010 American College of Occupational and Environmental Medicine
Copyright © American College of Occupational and Environmental Medicine. Unauthorized reproduction of this article is prohibited.

having COPD or other related illnesses) may be higher in coal mining than in the other industries, resulting in higher medical costs.

An inevitable question arising from this analysis is whether and to what extent the higher rates and costs of illness are attributable to workplace exposures (ie, to coal mine dust) as opposed to other characteristics or activities prevalent among workers in the coal mining industry. Unfortunately, the data do not allow us to directly measure and control for workplace exposures. Nevertheless, the analysis does control for several observable factors such as age, marital status, and socioeconomic characteristics in the workers' area of residence.

One potential confounding factor that cannot be directly controlled for in the analysis is smoking behaviors by individual policy holders; however, a more general comparison of male smoking rates across industries suggests that these differences cannot fully explain the observed differences in respiratory illness rates. For example, the largest differences in respiratory illness rates relative to coal mining were found with respect to the construction and manufacturing industries. Nevertheless, according to 2003 data from the Tobacco Use Supplement of the Current Population Survey, smoking prevalence among males in the coal mining industry is 30.5% (with a 20.1% to 40.8% 95% confidence interval) compared with 30.8% in construction (29.8% to 31.9%) and 23.7% in manufacturing (22.9% to 24.5%). The smallest differences in respiratory illness rates were found with respect to other mining, where the smoking prevalence is 23.4% (17.5% to 29.3%). The range of these summary statistics do not rule out smoking as a contributing factor to observed differences in respiratory illness rates, but they do suggest that smoking is not the main factor.

The results reported in this article must also be interpreted with a few caveats in mind. First, using industry codes to proxy for differences in occupational exposures is limited by underlying variation in these exposures within each industry. For example, some of the individuals included from the coal mining industry may not work in the mines themselves. A combination of industry and occupational codes would provide a better set of controls; however, this type of information is not available in our data.

Second, we use annual private insurance claims and ICD-9 codes to identify COPD and RESP cases and to calculate annual rates and costs of illness. These measures are imperfect and most likely understate actual rates of illness because they do not include individuals who are ill but do not seek medical care or those whose claims were submitted to and entirely handled by other insurance (eg, Medicare or workers' compensation) or those who died. Similarly, the medical costs estimates are also likely to be underestimates because they do not capture the costs of unreimbursed claims or claims submitted to and entirely handled by other insurance. It is important to note, however, that the potential underestimation of illness rates applies to all industries; therefore, the estimated differences in rates across industries are not necessarily biased.

Third, the estimated drug costs for respiratory illness may overstate the amount that is truly attributable to COPD or RESP. Although the regression analyses of total annual drug costs control for several individual factors, the respiratory illness indicators used in the analysis may be correlated with unobserved factors that also contribute to increased drug costs (ie, comorbidities), and the estimated regression coefficients may be capturing some costs attributable to these other factors. This type of overestimation is much less of a concern for other costs because nondrug medical claims for respiratory illness are directly identified through ICD-9 codes.

Fourth, we caution against combining the medical cost and drug cost estimates to obtain a total cost of respiratory illness across the five industries. As previously discussed, the methods and data used to derive the two categories of cost estimates are different. In particular, the ability to directly link medical claims to respiratory illness through ICD-9 codes implies that these cost estimates for respiratory illness (and the cross-industry differences in these costs) are subject to considerably less uncertainty than the corresponding drug cost estimates.

Finally, the data used in our analysis do not include any measures of indirect costs of respiratory illness (eg, lost wages or productivity because of illness, absenteeism, or mortality). These costs may be significant, and equally important, and may vary considerably across industries. Therefore, identifying data sources and expanding the analysis to compare indirect costs of respiratory illness across industries is a potentially fruitful and important area for future research, especially for quantifying the benefits that may result from successful intervention strategies.

The results of our analysis point to the need for continued efforts to reduce risks of respiratory illness, especially in the coal mining industry. The relatively high rates of illness in the coal mining industry, coupled with the finding of relatively high medical costs of respiratory illness and drug costs, suggest that the gains and potential cost effectiveness of measures to reduce coal mine exposures are particularly high. Even comparatively expensive interventions to reduce exposures may result in overall cost savings.

ACKNOWLEDGMENTS

The authors are grateful for contributions from Matthew Farrelly and Charles Feagan, and the helpful review and comments on earlier drafts provided by several individuals including J. Paul Leigh, Christopher Mark, and Patricia Gladowski, and an anonymous referee. Funding for this research was provided by the Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health through contract number 200-2002-00776.

REFERENCES

1. Coggon D, Taylor AN. Coal mining and chronic obstructive pulmonary disease: a review of the evidence. *Thorax*. 1998;53:398–407.
2. Cowie HA, Miller BG, Rawbone RG, Soutar CA. Dust related risks of clinically relevant lung functional deficits. *Occup Environ Med*. 2006;63: 320–325.
3. Montes II, Fernández GR, Reguero J, et al. Respiratory disease in a cohort of 2,579 coal miners followed up over a 20-year period. *Chest*. 2004;126: 622–629.
4. Yelin E, Trupin L, Cisternas M, Eisner M, Katz P, Blanc P. A national study of medical care expenditures for respiratory conditions. *Eur Respir J*. 2002;19:414–421.
5. Trupin L, Earnest G, San Pedro M, et al. The occupational burden of chronic obstructive pulmonary disease. *Eur Respir J*. 2003;22:462–469.
6. Petsonk EL, Parker JE. Coal workers' lung diseases and silicosis. In: Fishman AP, ed. *Fishman's Pulmonary Diseases and Disorders*, 4th ed. New York: McGraw-Hill; 2008:967–980.
7. Attfield MD, Castranova V, Wagner GR. Respiratory disease in coal miners. In: Rom WN, ed. *Environmental and Occupational Medicine*, 4th ed. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins; 2007, chapter 22.
8. Jalloul AS, Banks DE. The health effects of silica exposure. In: Rom WN, ed. *Environmental and Occupational Medicine*, 4th ed. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins; 2007, chapter 23.
9. Mannino D. COPD: epidemiology, prevalence, morbidity and mortality, and disease heterogeneity. *Chest*. 2002;121:1215–1265.
10. Leigh JP, Fries JF. Health habits, health care use and costs in a sample of retirees. *Inquiry*. 1992;29:44–54.
11. Rogers WH. Regression standard errors in clustered samples. *Stata Technical Bulletin.* 1993;13:19–23 (Reprinted in Stata Technical Bulletin Reprints 3:88–94).
12. Duan N. Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc*. 1983;78:605–610.

Copyright © American College of Occupational and Environmental Medicine. Unauthorized reproduction of this article is prohibited.