EXPOSURE ESTIMATES FOR THE NATIONAL COAL STUDY: THE USE OF MSHA COMPLIANCE DATA FOR EPIDEMIOLOGIC RESEARCH

NOAH S. SEIXAS,* MS, CIH • Thomas G. Robins,* MD, MPH • Michael D. Attfield,† Ph.D. • Lawrence H. Moulton,‡ Ph.D. • Carol Rice,§ Ph.D. • James L. Weeks,* Ph.D.

*Department of Environmental and Industrial Health, University of Michigan, School of Public Health
†Division of Respiratory Disease Surveillance, National Institute for Occupational Safety and Health
‡Department of Biostatistics, University of Michigan, School of Public Health
§Department of Environmental Health, University of Cincinnati
•Health and Safety Department, United Mine Workers of America

INTRODUCTION

Quantitative estimates of exposure in epidemiology greatly improve a study's ability to detect low level effects, distinguish between competing effects of etiological agents and perhaps most importantly, establish exposure-response relationships that can serve as the basis for public health intervention. However, the usefulness of exposure estimates depends on their accuracy and precision. Systematic bias in exposure measures will result in biased estimates of the exposure response relationship while random errors in measurement or misclassification will bias the observed association toward the null.[1]

The usefulness of accurate exposure data has been exemplified by researcb on the health effects of respirable coal mine dust by the British Pneumoconiosis Field Research (PFR) started in 1952.[2] The PFR studies get much of their strength from the concurrent measurement of health status and coal dust exposure over many years. The studies have been able to define the separate effects of dust exposure and confounding exposures[3] and to quantitatively define exposure-response relationships.[4,5,6]

Until recently, detailed exposure-response studies have not been possible in the United States. With the advent of the Federal Coal Mine Health and Safety Act of 1969, the National Study of Coal Workers Pneumoconiosis (NSCWP) was initiated. The study was set up as a series of rounds of health data collection. Miners for the first round of the study were selected from a group of 31 mines. In the second and third rounds, miners from additional mines were added to the group. The cohort for the fourth round of the study, concluded this year, includes a subset of younger miners wbo participated in the first two rounds.

The usefulness of the NSCWP studies for defining the exposure response relationship has been limited. Dust exposure estimates for the NSCWP were originally intended to be made using measurements taken on each individual miner by the Mine Safety and Health Administration (MSHA) compliance program.[7] However, the first two rounds were unable to use this data because only a few years of exposure data had been collected at the time. Instead, these analyses relied on years worked or years underground as a surrogate for exposure.[8,9] The third round of the NSCWP made limited use of exposure data in the analysis of pulmonary function changes.[10]

The exposure data collected for the MSHA compliance monitoring program have produced one of the largest exposure data sets in existence. It now contains about 18 years worth of personal exposure monitoring from all U.S. underground coal mines. Used appropriately, these data may provide a very substantial basis for precise exposure estimates for the NSCWP.

The overall purpose of our study is to combine measurements of miners' lung function with existing historical measurements of exposure in order to develop a quantitative exposure-response relationship. The current report is limited to analysis of the exposure data. Descriptive analyses of the data will first be undertaken and potential sources of bias explored. An exposure matrix will be designed with mine, occupation and year as the three dimensions. The final estimates of the mean exposure level in each cell of the mine, occupation and year matrix will be made with a regression model. The mean exposures will then be merged with the work histories of the miners in the study cobort in order to estimate cumulative exposure for each participant. These cumulative exposures will then be used to estimate the exposure-response relationships for pulmonary function and radiographic indices.

COLLECTION OF EXPOSURE DATA

The exposure data comes from the MSHA coal mine health compliance program. The law enacted in 1969 requires coal mine operators to maintain respirable dust levels below specified levels. To ensure compliance with the standard, operators are required to periodically conduct air sampling and MSHA compliance inspectors visit each coal mine several times each year and collect additional samples.[11]

The protocols for air monitoring have changed over time. Operators were originally required to sample miners in highly exposed occupations, termed High Risk, or Designated Occupations, repeatedly and all other miners from two to four times each year. In 1980 these requirements were altered to reduce the operators' sampling burden. Occupations with low expected exposures were dropped from the personal monitoring requirement and area samples were employed instead.

MSHA inspector sampling follows a similar pattern although the number of samples is much lower. Inspectors sample each of the Designated Occupations although they exercise discretion in deciding which other occupations or miners to sample.

All samples used here are full shift personal samples collected with a 10 mm nylon cyclone-filter assembly. Results are determined gravimetrically by MSHA. The specified sampling flow rate is 2 liters per minute and a correction factor is applied to the result to make it comparable to the British MRE sampler upon which the federal exposure standard is based. The data provide an internally consistent measure of respirable coal mine dust and are therefore appropriate for investigation of chronic respiratory effects.

DATA SET DESCRIPTION

The exposure data used for this study was collected by MSHA and files containing the data were obtained from two sources. NIOSH provided data tapes which contain all exposure data for underground coal mines from 1970–1979. There is a total of 3,226,602 samples contained in the NIOSH data set.

The data covering 1979–1987 was obtained from the Bureau of Mines (BOM). The BOM supplied data tapes of all underground coal mine samples for the years 1979–1987, totalling 1,519,892. A breakdown by year and sample source is given in Table I. Note that the number of operator generated samples is significantly reduced after 1980 when the revised sampling protocols were adopted.

The information contained in both data sets include (a) date of sample, (b) MSHA mine number, (c) occupation code (d) concentration, and (e) social security number, along with various information about the type of sample, type of mine, sampling time, etc.

POTENTIAL LIMITATIONS OF THE EXPOSURE DATA

Several potential problems with the use of the data for epidemiologic analysis may be identified. Although the overall number of samples is very large, about 4.7 million, they are not equally distributed between mines, occupations and years. Most samples are taken in the High Risk, or Designated Occupations. Thus, in many cells of the mine, year and occupation matrix, there is no data at all.

The fact that the exposure data is based on a regulatory compliance program suggests that the way in which the samples are taken may lead to bias. MSHA compliance sampling continues for up to five days, depending on the results of the earlier samples. These sampling rules may lead to more sampling of jobs with higher exposures. In addition, it has been suggested[12] that adjustments in the work process or engineering controls could be made which would reduce exposures on subsequent days of a compliance inspection. Thus, use of an average of all inspector data, without respect to the day it was taken, might lead to a biased estimate of exposure.

Table I

Number of Underground Coal Mine Dust Samples in Bureau of Mines Data Set 1979–1987

YEAR	OPERATOR	INSPECTOR	(%)	TOTAL
1979	417,912	23,546	(5.3)	441,458
1980	269,092	19,642	(6.8)	288,734
1981	56,862	29,845	(34.4)	86,707
1982	39,755	27,342	(40.8)	67,097
1983	118,692	26,523	(18.3)	145,215
1984	109,929	26,686	(19.5)	136,615
1985	98,927	26,882	(21.4)	125,809
1986	93,479	24,767	(21.0)	118,246
1987	85,886	24,125	(21.9)	110,011
Total	1,290,534	229,358	(15.1)	1,519,892

For non-Designated Occupations, both inspector and operator sampling may concentrate on particular jobs with a history of high exposure. Thus, the data is likely to be weighted toward high exposures.

It has also been suggested[13,14] that operator data deviates from the lognormal distribution which one would generally expect in exposure data. They show that operator data has a large number of low exposure samples, yielding lower mean exposure levels than inspector data.

METHODOLOGY

The overall analytic strategy is to (a) choose a sample of mines, (b) conduct descriptive analyses, (c) assess bias, (d) reduce the number of categories for analysis, (e) specify the best model for prediction of mean exposures and (f) calculate estimates of exposure.

First, a sample of mines will be chosen. All members of the round four cohort originally worked in one of 36 mines but most have also worked in other mines. Since the 36 mines constitute a substantial proportion of the total time spent in mining, it is reasonable to use these mines as a sample with which initial exploratory analyses and development of our final analytic strategy will be done. Using this model, the actual exposure estimation will then be done for all mines in the cohort members' occupational histories.

Second, descriptive analyses will include distributions of the number of data points and means within each mine, year and occupation category. The degree to which the data conforms to the expected lognormal distribution will then be assessed. In order to avoid analyzing the distribution as a combination of heterogeneous distributions, the data will first be log-transformed and then standardized (subtract mean, divide by standard deviation) within each three-way matrix cell. The data will then be analyzed as a single distribution, or stratified on pertinent categories such as occupation. The log-transformed data will be examined for skewness and kurtosis and histograms will be generated to visually assess normality.

Third, we will examine the issues of potential bias to determine the best use of the data. Since we do not have any standard by which to measure bias in different parts of the data, we must analyze the data internally, and use the information we have to select the most reliable data and estimate the magnitude of the potential bias.

Potential sources of bias will be examined to see if they make a substantial difference in the exposure estimates. For instance, the means of inspector versus operator samples will be compared and samples from the first day of a compliance inspection will be compared to subsequent days. If substantial differences are observed, we will consider adjusting the data to account for the bias. If adjustments are not possible, we will use the data which most closely reflect representative exposure conditions. The magnitude of the potential bias will be estimated and reported with the final results.

Fourth, we will explore reducing the number of cells in the mine, year, occupation matrix by examining the differences between mines and by attempting to collapse the number of occupational categories. These reductions will help both in simplifying the final analytic model adopted and by grouping data across similar categories in which data may be very sparse.

The difference between mines will be assessed first. If no substantial difference can be observed, we will remove the mine effect from the dimensions of the matrix. If differences do exist, the effect of individual mines or mine groups will necessarily be included in subsequent analyses.

Second, we will reduce the number of occupations from the 197 MSHA codes to a smaller number of homogeneous exposure categories. Means for each occupation will be calculated and rank ordered. Grouping of occupations will then be done to identify occupations with mean exposures within, for example, 0.2 mg/m^3. Analysis of variance will then be conducted within each group, controlling for mine and year, to assess the validity and effectiveness of the grouping.

Finally, after the potential biases are examined, the final most valid data is selected and the categories for analysis are defined, a regression model will be developed to estimate means and variances of exposure in a mine, exposure category and year matrix. This part of the analysis will be conducted on a larger data set incorporating all mines which appear in the work histories of the NSCWP cohort.

The analysis will be done in a two stage process. First the effect of year for each occupation group and mine will be estimated. Then these "offsets" can be used in the regression model to estimate the exposure category and mine effects. The model will generally be of the form:

$$C_i = a + \sum^{m}\beta_m k_{m\,i} + \sum^{x}\beta_x k_{x\,i} + Y_i + e_i$$

where C_i is the sample result, β_x and β_m are the estimated regression coefficients for each exposure category and mine, respectively, k_x and k_m are the dummy variable indicators for exposure category and mine respectively and Y_i are the offsets calculated for each year within individual mine and exposure categories. The parameter estimates will then be used to calculate the mean and standard deviation of exposure in each category of the matrix.

PRELIMINARY ANALYSIS OF TWO SAMPLE MINES

Two of the original NSCWP study mines were chosen for this preliminary analysis. Table II shows the number of samples in each of the two mines by source of collection. Only a small percentage of the overall data is collected by inspectors.

Table II

Number of Underground Coal Mine Dust Samples in Bureau of Mines Data Set 1979-1987

	INSPECTOR	(%)	OPERATOR	(%)	TOTAL
Mine 1	285	(2.1)	13,018	(97.9)	13,303
Mine 2	913	(4.1)	21,269	(95.9)	22,182
Total	1198	(3.5)	24,287	(96.5)	35,501

Figure 1 shows the number of samples collected in each mine by year. It is evident that the overall number of personal samples was greatly reduced after 1980.

Only some of the occupations have had substantial monitoring of exposure. Out of the 197 possible occupation codes, samples were actually taken in only 76 and 108 of the categories in the two mines, respectively. Thus, many of the occupation codes had no data in them at all although some of these occupations may not exist in these two mines.

Figure 2A and B shows an example of the exposure data distribution, in this case for all years in one mine and a single, highly monitored occupation, continuous miner operator. The untransformed data (Figure 2A) is highly skewed, while the log-transformed values (Figure 2B) are much closer to normal, with a skewness coefficient of -0.43. There appears to be an over representation of sample values of 0.1 mg/m^3. These low values occur mainly in the operator collected data and may represent some bias.

Finally, Figure 3 shows the arithmetic mean concentration of respirable dust based on both operator and inspector data for continuous miner operator in the two mines. Although there is a visually observed downward trend, the year to year variability in the means is also apparent.

CONCLUSION

The data collected by the MSHA coal mine compliance program forms a very substantial basis for relatively precise exposure estimates for the fourth and subsequent rounds of the NIOSH National Study of Coal Workers Pneumoconiosis. These estimates should greatly enhance the power of the NSCWP to detect any possible effects of the relatively low dust exposure levels present after enactment of the MSHA law. The estimates will also help establish the exposure-

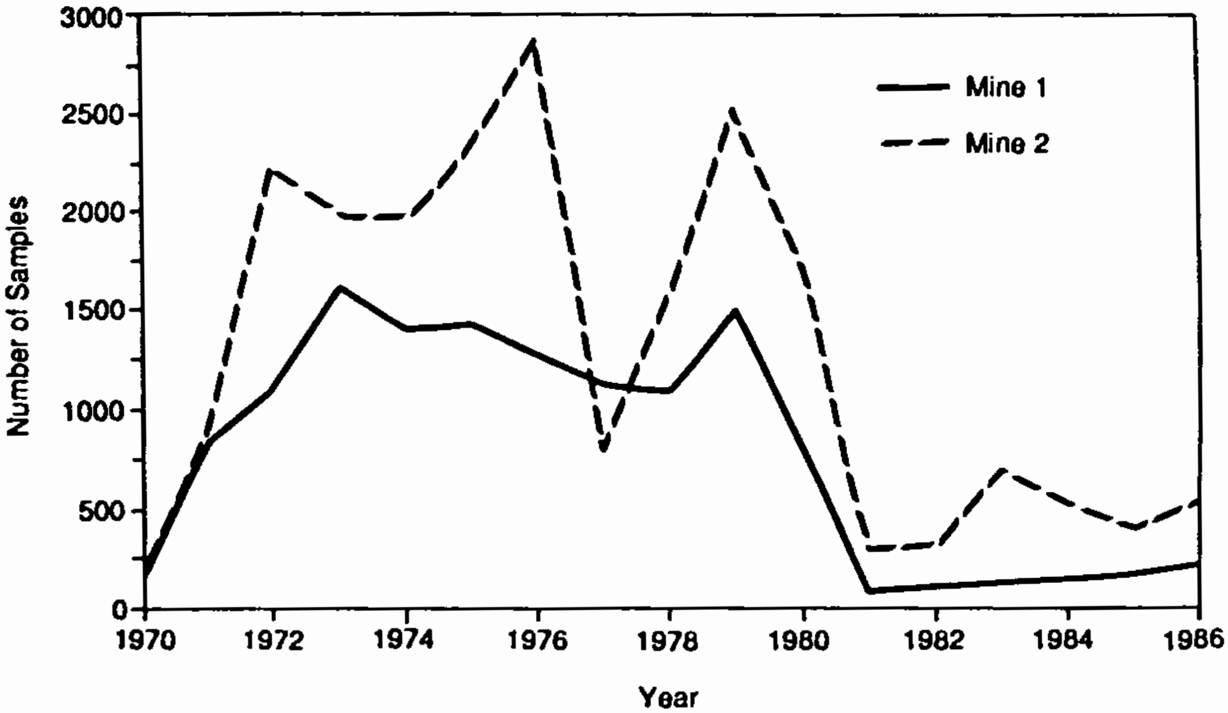

Figure 1. Number of samples by year for two selected mines.

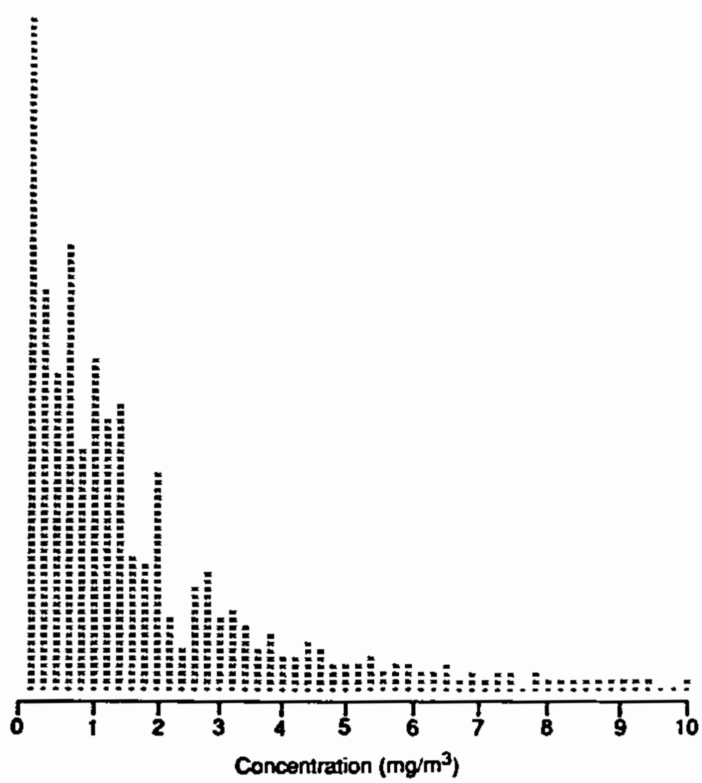

Figure 2A. Histogram of untransformed sample values for continuous miner operator, 1970–1986. (Arithmetic mean = 1.6 mg/m^3, skewness = 1.8, kurtosis = 3.8.)

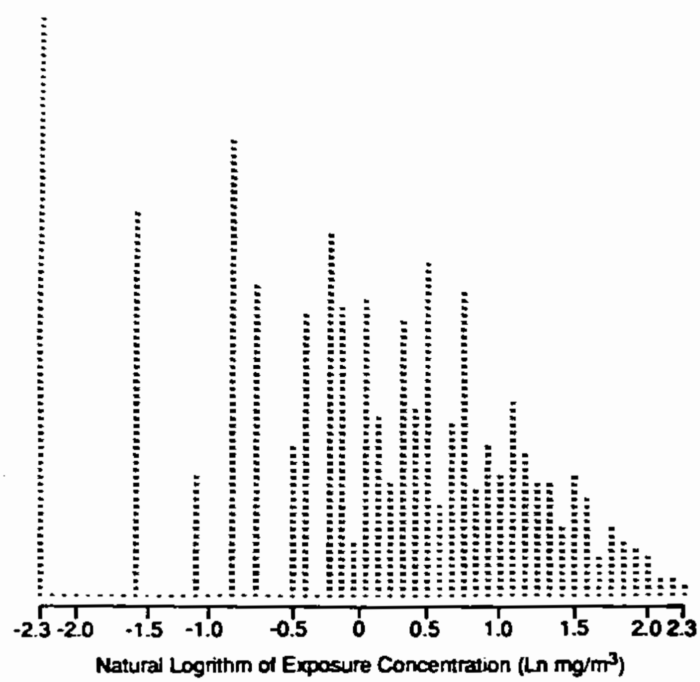

Figure 2B. Histogram of log transformed values for continuous miner operator, 1970–1986. (Geometric mean = 0.95, geometric standard deviation = 3.1, skewness = −0.43, kurtosis = −0.48.)

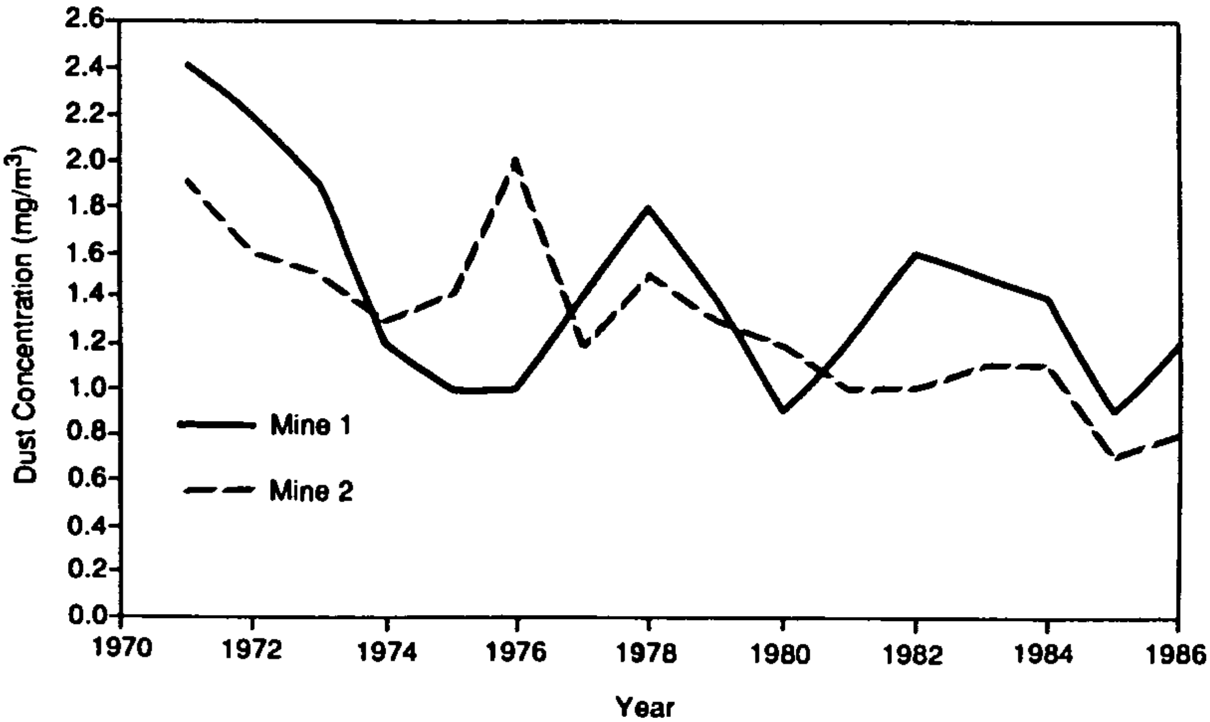

Figure 3. Arithmetic mean exposures by year for continuous miner operator in two selected mines.

response relationship and separate the effect of dust from other confounding exposures.

By grouping similar occupations into exposure categories, and by using regression modelling techniques, the problem of data sparseness will be substantially limited. Although there may be some bias in the data, the magnitude of the error is unlikely to greatly affect the observed exposure-response relationship.

In addition to enhancing the analysis of the current NSCWP studies, development of quantitative exposure estimates across the mining industry will also substantially help future research into various parameters of coal mine dust induced disease.

REFERENCES

1. Cochran, W.G.: Errors in Measurement in Statistics. *Technometrics* 10:637-666 (1968).
2. Rogan, J.M., Rae, S.: The National Coal Board's Pneumoconiosis Field Research—An Interim Review. In: *Inhaled Particles and Vapours*, pp. 493-508. C.N. Davies, Ed. Pergamon Press, London (1967).
3. Jacobsen, M., Burns, J., Attfield, M.D.: Smoking and Coalworkers' Simple Pneumoconiosis. In: *Inhaled Particles IV*, pp. 759-772. W.H. Walton, Ed. Unwin Bros., London (1977).
4. Jacobsen, M., Rae, S., Walton, W.H., Rogan, J.H.: The Relation Between Pneumoconiosis and Dust Exposure in British Coal Mines. In: *Inhaled Particles III*. Vol. 2, pp. 903-919. W.H. Walton, Ed. Unwin Brothers, London (1971).
5. Rogan, J.M., Attfield, M.D., Jacobsen, W.H.: Role of Dust in the Working Environment in Development of Chronic Bronchitis in British Coal Miners. *Br. J. Ind. Med.* 30:217-226 (1973).
6. Soutar, C.A., Hurley, J.F.: Relation Between Dust Exposure and Lung Function in Miners and Ex-Miners. *Br. J. Ind. Med.* 43:307-320 (1986).
7. Jacobson, M.: Personal communication, July 12, 1988
8. Attfield, M., Reger, R., Glenn, R.: The Incidence and Progression of Pneumoconiosis Over Nine years in US Coal Miners: II. Relationship With Dust Exposure and Other Potential Causative Factors. *Am. J. Ind. Med.* 6:417-425 (1984).
9. Morgan, W.K.C., Handelsman, L., Kibelstis, J., et al.: Ventilatory Capacity and Lung Volumes of US Coal Miners. *Arch. Environ. Health* 28:182-189 (1974).
10. Attfield, M.: Longitudinal Decline in FEV_1 in United States Coalminers. *Thorax* 40:132-137 (1985).
11. US Congress.: Public Law 91-173: Federal Coal Mine Health and Safety Act of 1969. *US Statutes at Large* 83:742-804 (1970).
12. Sworn Testimony of November 9, 1982, p. 198, Consolidated Coal Company vs. Secretary of Labor, No. WEVA 82-245 (Federal Mine Safety and Health Review Commission) (Mr. Sutherland is Chief, Division of Health, Coal Mine Safety and Health, MSHA).
13. Boden, L.I., Gold, M.: The Accuracy of Self-Reported Regulatory Data: The Case of Coal Mine Dust. *Am. J. Ind. Med.* 6:427-440 (1984).
14. Sharp, G.: Dust Monitoring and Control in the Underground Coal Mines of Eastern Kentucky. unpublished paper, Dept. of Anthropology, University of Kentucky (1978).

Proceedings of the VIIth International Pneumoconioses Conference | *Part*
Transactions de la VIIe Conférence Internationale sur les Pneumoconioses | *Tome*
Transaciones de la VIIa Conferencia Internacional sobre las Neumoconiosis | *Parte*

I

NIOSH - ILO

Pittsburgh 1988

BOM - MSHA - OSHA

Pittsburgh, Pennsylvania, USA—August 23-26, 1988
Pittsburgh, Pennsylvanie, Etats-Unis—23-26 aout 1988
Pittsburgh, Pennsylvania EE. UU—23-26 de agosto de 1988

DEPARTMENT OF HEALTH & HUMAN SERVICES · USA

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control
National Institute for Occupational Safety and Health

CDC
CENTERS FOR DISEASE CONTROL

Sponsors

International Labour Office (ILO)
National Institute for Occupational Safety and Health (NIOSH)
Mine Safety and Health Administration (MSHA)
Occupational Safety and Health Administration (OSHA)
Bureau of Mines (BOM)

September 1990

DISCLAIMER

DHHS (NIOSH) Publication No. 90-108 Part I