American Journal of Epidemiology
Copyright © 2000 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved
Vol. 151, No. 10
Printed in U.S.A

# Characterization of Dust Exposure for the Study of Chronic Occupational Lung Disease: A Comparison of Different Exposure Assessment Strategies

Dick Heederik[1,2] and Michael Attfield[2]

Various exposure assessment strategies were compared in the study of the relation between dust exposure and 11-year lung function change in 1,172 miners with 36,824 concurrently measured personal dust samples available from the 1969–1981 US National Study of Coal Workers' Pneumoconiosis. A miner's average exposure was assessed by calculating average exposures based on dust samples taken from each individual and by using different job exposure matrices (JEMs) with different underlying exposure categorizations, based on occupational categories, job title, mine, and time, to obtain average exposure estimates. For each grouping procedure, intragroup and intergroup variances and the pooled standard error of the mean were calculated to assess relative efficiency. The results show that considerable variation in slopes of exposure-response relations was found using different exposure assessment strategies. Standard errors of the slopes of the exposure-response relations with exposure on an individual basis compared with JEMs. Exposure assessment on an individual basis was extremely sensitive to the number of exposure measurements per individual. The study demonstrates the advantages and disadvantages of different exposure assessment strategies and shows the need for explicit publication of exposure assessment strategies for epidemiologic studies. Careful assessment of the influence of misclassification error in the exposure assessment on exposure-response modeling is warranted. *Am J Epidemiol* 2000;151:982–90.

forced expiratory volume; lung diseases, obstructive; occupational diseases; respiratory function tests

Over the last decade, a wide range of exposure and dose assessment strategies has been proposed for use in retrospective and prospective epidemiologic studies (1). Some of these approaches, such as toxicokinetic modeling of exposure data and quantifying biomarkers of exposure, aim at monitoring biologically relevant exposure indices. Instead of using crude surrogates of exposure, there is a movement toward establishing indices of greater relevance to the etiology of the disease of interest and application of quantitative exposure data (2). Another issue relates to uncertainty in the process of assessing exposure-response relations. Uncertainty is inherent to all exposure assessment strategies and is caused by large random variation in exposure over time and in space. In general, this reduces the power of a study to detect an association between exposure and disease and can introduce bias into the exposure-response estimates (2). If the error is random, the bias is usually toward zero and referred to in the statistical literature as "attenuation." If exposure measurements have been taken for each individual, the magnitude of potential bias in the regression coefficient depends on the ratio of the intraindividual and interindividual components of variance and the number of repeated measurements, according to the formula (3):

$$b = \beta(1 + \lambda/k)^{-1}$$

where:

$b$ = observed value of the empirical regression coefficient of $Y$ on $X$, while $X$ measured with error

$\beta$ = true value of the regression coefficient of $Y$ on X

$\lambda = {}_wS_i^2/{}_bS_i^2$

${}_wS_i^2$ = estimate of intraindividual variance in exposure

${}_bS_i^2$ = estimate of interindividual variance in exposure

$k$ = number of repeated measurements per individual

Received for publication December 7, 1998, and accepted for publication June 25, 1999.

Abbreviations: $FEV_1$, forced expiratory volume in 1 second; JEM, job exposure matrix; MSHA, Mine Safety and Health Administration; SE, standard error.

[1] Environmental and Occupational Health Group, University of Utrecht, Utrecht, Netherlands.

[2] Division of Respiratory Disease Studies, Epidemiology Branch, National Institute for Occupational Safety and Health, Morgantown, WV.

Reprint requests to Dr. Dick J. J. Heederik, Department of Environmental Sciences, Environmental and Occupational Health Group, University of Wageningen, P.O. Box 238, 6700 AE Wageningen, Netherlands (e-mail: Dick.Heederik@Staff.eoh.wau.nl).

(Exposure data are often log-normally distributed and, in these cases, the variance components are obtained after log transformation of the data.)

In practice, the accuracy (validity and precision) of a chosen index of exposure affects the results of an exposure-response analysis. This leads to the paradoxic observation that a crude surrogate of exposure can sometimes correlate better with the response than does a biomarker, because it is measured with greater precision (4). In practice, it is difficult to assess measures of exposure because an exposure "gold standard" does not exist, forcing researchers to compare different indices of exposure (2, 5). This is unsatisfactory because chance alone could be responsible for the best fitting model, while differences in performance of several exposure indices in the analysis can seldom be explained in a satisfactory way (5, 6). A better approach, perhaps, would be to get insight into the accuracy of various alternative exposure indices and to develop criteria for optimal exposure indices based on that information. In this way, the choice of an exposure assessment strategy would be based on objective information on the exposure variables instead of the observed behavior of exposure indices in an epidemiologic analysis. Unfortunately, few studies have compared different approaches to estimating mean exposure in this way. In an attempt to fill this gap, we assessed and compared various exposure indices for a cohort of underground miners who participated in a longitudinal epidemiologic study, the National Study of Coal Workers' Pneumoconiosis (7). The aims of this study were to evaluate different ways of calculating the long-term average exposure in an exposure-response analysis. The exposure assessment strategies to be examined were chosen on the basis of their underlying exposure variance components. Estimates of average exposure were based on 36,824 measurements of respirable coal mine dust concentration taken over approximately a 10-year period. The average exposure was calculated using the same database of exposure measurements but different strategies. Because the time of exposure was constant for each individual, confounding by time-related variables was absent. This has been a major drawback of some earlier multicomparison studies (5, 6). In addition, we calculated correlations between the different exposure indices and derived variance components of intraindividual, interindividual, intragroup, and intergroup variance, which might give a priori indications of the effectiveness of a certain measurement strategy.

## MATERIALS AND METHODS

### Exposure assessments

The exposure measurements used were collected by mine operators and inspectors under the auspices of the Mine Safety and Health Administration (MSHA) between 1970 and 1979. The rationale for the collection of dust samples differs from a strategy that would be a priori developed for epidemiologic purposes. In one respect, it is a compliance program, in which frequent sampling is undertaken on high-risk workers. In another respect, however, it provides data appropriate for epidemiologic investigations, since all other workers were sampled periodically at lower frequency, more or less at random (8).

All respirable dust samples were collected using a 10-mm nylon cyclone preseparator. The resulting data were converted by multiplication using a constant factor to make them equivalent to concentrations that would have been obtained using the British Mining Research Establishment sampler (8). Dust data were available for all workers who participated in two medical examinations undertaken in 1969–1971 (round 1) and 1977–1981 (round 3) and who had complete lung function data, smoking history, occupational history, and dust exposure data. The dust measurements for this population were extracted by Social Security Number from the MSHA records, giving 36,824 samples. Over 98 percent of these measurements were obtained by mine operators. Very few measurements were obtained by mine inspectors, which is in agreement with data presented by others (9, 10). Within occupational groups, the data were strongly skewed and could roughly be described by a log-normal distribution.

Occupational histories at round 3 were coded according to the standard MSHA occupation code classification scheme of 1984 (MSHA form 2000-169, July 1984). This enabled each worker's occupational history to be linked with the MSHA exposure measurements. Although the year in which a miner changed occupations was known, the month and day were not known. Hence, it was not possible to link dust measurements in that year precisely to the work history. Therefore, none of the measurements taken in a year in which a miner changed his occupation was used in the analysis. For some analyses, this led to a reduction of available miner-years of exposure data of 12–14 percent.

The long-term average exposure for each miner was calculated in the following ways.

*Method I* (by calculating an individual arithmetic mean exposure per person using all measurements available per worker and using random samples of these measurements (sample sizes of 3, 6, 9, and 12 measurements per worker)). These analyses were performed including miners with at least three exposure measurements available ($n$ = 1,105).

*Method II* (by creating several job exposure matrices (JEMs) that used average exposures based on the

36,824 measurements available by occupational category and merging this information with individual work history data to calculate an average dust exposure). The exposure categories of these matrices were made by aggregating workers with job titles with similar exposures. Alternative grouping strategies were evaluated based on a priori grouping on the basis of similarity in processes and tasks, but this yielded less distinct exposure categorizations. Six exposure categories were distinguished: 1) exposure of <0.5 mg/m$^3$, 2) 0.5–1.0 mg/m$^3$, 3) >1.0–1.5 mg/m$^3$, 4) >1.5–2.0 mg/m$^3$, 5) >2.0–2.5 mg/m$^3$, and 6) >2.5 mg/m$^3$.

*Method III* (by using the above-mentioned JEM (method II) but augmenting the matrix with a time axis and a mine axis, allowing for trends over time in exposure and for differences in exposure between mines). This led to matrices with more exposure categories than in method II above. Temporal reductions in exposure were expected as the compliance limit was reduced from 3 to 2 mg/m$^3$ in 1972 (11). All occupations showed a similar pattern in decline in exposure over time. In one model we distinguished the first 3 years with the clearest decrease in dust exposure over time on the time axis (four cells for time) of the matrix. In another model we took into account each year during the follow-up period.

*Method IV* (by modeling the exposure in an analysis of variance with each occupation and mine as explanatory variables and using predicted average exposure values for all occupation and mine combinations to calculate long-term average exposures). This analysis was performed on the original scale, and thus exposure data were not log-transformed.

*Method V* (by applying a recently developed external JEM for epidemiologic studies in the US mining industry containing an occupation, mine, and time axis (12)). This job exposure matrix made use of more than 300,000 personal respirable dust exposure measurements from the Mine Safety and Health Administration database for all US coal mines. The matrix comprised more than 12,000 strata of occupation, mine, and year three-way and two-way combinations.

The number of individuals included in the analyses differed for analysis (method I), since only miners with at least three exposure measurements were included. Similar results were obtained for analyses under methods II–V if the same workers were included as under method I.

### Medical surveys

The data considered were drawn from medical surveys taken at 31 nationally distributed mines at rounds 1 and 3. Because some of the mines closed between the surveys, and because in some cases too few miners attended both surveys at some mines, data from 24 mines were available. The methods used were similar in both surveys (7). In the first surveys, the final forced expiratory volume in 1 second ($FEV_1$) recorded was the largest value of up to five blows (including at least two practice blows). In the later survey, up to 10 blows were permitted but, for reasons of comparability with the first survey, only the first five were considered. At both surveys, an Ohio rolling seal spirometer (Ohio Medical Products Division, Airco, Inc., Madison, Wisconsin) was used. The technician differed between the two occasions. Smoking habits were classified, using a method that has been applied before in this cohort (7), into categories of current smokers, nonsmokers, exsmokers, and intermittent smokers. Lung function change between the two surveys was standardized to an 11-year period and expressed as annual change in lung function as described earlier (7).

### Statistical analyses

One-way analyses of variance were applied on log-transformed repeated exposure measurements to derive estimates of the intraindividual and interindividual variance in exposure, ${}_wS_i^2$ and ${}_bS_i^2$, respectively. For strategies based on various grouping procedures, the intergroup, intragroup, and intraindividual variance components were calculated by a two-way analysis of variance with category and worker (nested within a category) in the model as described earlier (13). As an estimate of the overall precision of the group average exposure, the median of the standard errors of the mean exposure per exposure group was calculated. These variance components and the median standard error could not be calculated for the external JEM because the original data used for this matrix were not available. As an approximation, intragroup and intergroup variances were calculated using the 36,824 exposure measurements available for this study with occupation and mine in the two-way analysis of variance model. The time axis was omitted because of the larger number of empty cells that occurred for this model.

The relations between lung function change and other variables were calculated using least squares regression analysis. Change in $FEV_1$ was chosen as the dependent variable with dust exposure, age, and standing height at the beginning of the survey and smoking habits as predictors (7).

## RESULTS

The complete data set consisted of 1,172 coal miners. The population characteristics at round 1 are shown in table 1. The arithmetic mean exposure

**TABLE 1. Population characteristics of 1,172 underground coal miners at round 1 of the medical survey, National Study of Coal Workers' Pneumoconiosis, 1969–1971**

| Variable | Mean |
|---|---|
| $\Delta FEV_1$* (liters) | −0.50 (0.45)† |
| Age (years) | 39.5 (10.2) |
| Standing height (cm) | 176.3 (6.6) |
| Smokers (%) | 39.7 |
| Exsmokers (%) | 19.4 |
| Nonsmokers (%) | 19.0 |
| Intermittent smokers (%) | 21.9 |

* $\Delta FEV_1$, change of lung function (forced expiratory volume in 1 second) over an 11-year period.
† Numbers in parentheses, standard deviation.

between rounds 1 and 3, broken down by year, is presented in table 2. The table shows that the mean exposure decreased over time, especially during the years just after completion of round 1 (1970–1974). Dust levels declined from 2.2 mg/m$^3$ in 1970 to 2.1 mg/m$^3$ in 1971, 1.7 mg/m$^3$ in 1972, and 1.4 mg/m$^3$ in 1973. A similar trend of decreasing exposure was observed when the population was broken down into specific underground and surface jobs. After this initial 4-year period, average exposure levels remained relatively stable over time. The number of measurements taken varied considerably over the period of observation, increasing from 905 in 1970 to a maximum of more than 5,300 in 1973 and declining afterward to approximately 2,000 measurements in 1979. The number of exposure measurements per person-year was on average approximately 2.9 and had a skewed distribution ranging from 1 to 10 measurements per person.

Table 3 gives the relation between change in $FEV_1$ and average dust exposure. The relation between change in $FEV_1$ over 11 years and dust exposure was corrected for age at the start of the survey, standing height, and smoking habits. As expected, an increase

**TABLE 2. Number of measurements, average respirable dust exposure, and geometric standard deviation (GSD) between surveys of rounds 1 (1969–1971) and 3 (1977–1981), per year for a population of 1,172 coal miners, National Study of Coal Workers' Pneumoconiosis**

| Year | No. of measurements | Mean (mg/m$^3$) | GSD |
|---|---|---|---|
| 1970 | 905 | 2.2 | 2.7 |
| 1971 | 3,455 | 2.1 | 3.1 |
| 1972 | 4,769 | 1.7 | 3.4 |
| 1973 | 5,326 | 1.4 | 3.7 |
| 1974 | 4,856 | 1.4 | 3.7 |
| 1975 | 4,641 | 1.4 | 3.6 |
| 1976 | 4,634 | 1.5 | 3.2 |
| 1977 | 3,249 | 1.1 | 3.4 |
| 1978 | 3,004 | 1.2 | 3.3 |
| 1979 | 1,985 | 1.3 | 3.2 |

**TABLE 3. Regression analysis of the relation between 11-year individual mean exposure to respirable coal dust and 11-year change of lung function ($FEV_1$) for 1,105 coal miners allowing for age, standing height, and pack-years of tobacco smoked, National Study of Coal Workers' Pneumoconiosis, 1969–1981**

| Samples/ worker | $FEV_1$ coefficient (ml per mg/m$^3$) | SE* (ml per mg/m$^3$) |
|---|---|---|
| All samples | −4.5 | 1.5 |
| 15 | −3.8 | 1.4 |
| 12 | −3.6 | 1.4 |
| 9 | −3.2 | 1.3 |
| 6 | −2.5 | 1.3 |
| 3 | −1.8 | 1.1 |

* SE, standard error.

in age was associated with a faster decrease in $FEV_1$ over 11 years of approximately 9.6 ml per year (standard error (SE), 13 ml; $p < 0.001$). An increase in standing height was also associated with an increased but smaller drop in $FEV_1$ of 6.7 ml per cm (SE, 2.0 ml; $p < 0.05$). The $FEV_1$ of smokers and intermittent smokers decreased faster compared with that of nonsmokers (−90 ml (SE, 35 ml; $p < 0.05$) and −47 ml (SE, 40 ml; $p = 0.24$), respectively). Exsmokers experienced a slight, statistically nonsignificant increase in $FEV_1$ over the 11-year period of 31 ml (SE, 41 ml; $p = 0.45$). These coefficients remained similar in all analyses. The top row of the table gives the relation between change in $FEV_1$ over 11 years and dust exposure when all dust exposure measurements available for each individual were used to calculate an average dust exposure over the period. As fewer measurements were used, the slope of lung function change on dust exposure became progressively smaller and changed from statistically highly significant to statistically nonsignificant. The slope of the exposure-response relation based on three exposure measurements was about one third of that found when all measurements were used. This occurred despite a total of more than 3,000 measurements being available and used in the former analysis.

The analysis of variance revealed that the intraindividual variability (day-to-day variability) was three times the interindividual variability. This indicates that a considerable number of measurements per individual must be available to distinguish reliably if average exposure levels between individuals differ and, thus, to provide a reliable estimate of a worker's long-term average exposure. Given a variance ratio of three and the average number (36,824 / 1,172 ≈ 31) of measurements per worker available to estimate individual average exposures, the observed exposure-response relation would be attenuated by less than 10 percent (based on the formula $b = \beta(1 + \lambda / k)^{-1}$). If only one mea-

surement per worker was available, the observed regression coefficient would have led to a severely underestimated regression coefficient and would have been approximately 25 percent of the true regression coefficient. The latter calculations are in close agreement with the results presented in the table. Because underestimation of the exposure-response relation was the least when all measurements available were being used, and because the theoretically expected underestimation was small when all measurements were used, this index of exposure was considered the "gold standard" in further analyses.

Results of the regression of $FEV_1$ on average exposure calculated in various ways are given in table 4. The relations found were all negative. The exposure-response coefficient is the greatest for the average exposure, which was estimated by a JEM using a combination of six exposure categories and mine. This regression coefficient was somewhat greater than that obtained using the individual average exposures but within the confidence interval of this regression coefficient. All other estimates of the relation between average exposure and lung function change based on JEMs show smaller coefficients, although confidence intervals overlap. When standard errors are examined, the one from the regression using individual average exposures was the lowest and had the highest associated $t$ value and statistical significance. The regression coefficients had greater standard errors and often became statistically insignificant when JEMs were used or if the exposures were estimated by using the predicted values of an analysis of variance model with "mine" and "occupation" as main effects. The modeled group exposure estimates demonstrated slightly more bias but a lower standard error compared with the comparable nonmodeled grouping strategy. Only models based on individual average exposures, JEMs based on a breakdown by exposure category and mine, and the model predicting an average exposure by exposure category and mine had statistically significant regression coefficients of $FEV_1$ on exposure. The middle column of the table gives the correlations of exposures calculated by using the JEMs with the individual average exposure. For instance, the average exposure calculated using a JEM based on six exposure categories by mine had a correlation of 0.53 with the individual average exposure. Generally, exposure indices from JEMs gave greater slopes and had the higher correlations with the individual average exposure. Average 11-year exposure values derived using the different strategies are also given in table 4 and are similar. However, the standard deviation of the distribution of cumulative exposures is remarkably different depending on the strategy chosen. That for the individual exposure assessment strategy is greater than those derived from the JEM-based strategies. Overall, it appears that the standard deviation of the cumulative exposure is related to the number of cells of the matrix.

To study the effect of reducing the measurement effort when a JEM is being used, we took a random sample of 3,315 exposure measurements. This is the same subset used in the exposure-response analysis of $FEV_1$ on individual average exposure using three dust samples per worker from table 3. The 3,315 samples were used to calculate group average exposures for the JEM with six exposure categories by mine. The derived regression coefficient was −5.6 (SE, 2.2; $p < 0.05$), in close accord with that derived when more than 11 times more exposure measurements were used.

A comparison of the variance components of intergroup and intragroup variability in exposure and precision for the different JEMs and some additional strategies can be found in table 5. The simplest JEM, using a distinction among six occupational exposure cate-

**TABLE 4. Regression analysis of the relation between average exposure to coal dust and 11-year change of lung function ($FEV_1$) for 1,172 coal miners, corrected for age, standing height, and pack-years of tobacco smoked for different exposure assessment strategies, and estimated average cumulative exposure and standard deviation (mg x year/m³), National Study of Coal Workers' Pneumoconiosis, 1969–1981**

| | $FEV_1$ coefficient (ml per mg/m³) | SE* (ml per mg/m³) | Correlation† | Average exposure | SD* |
|---|---|---|---|---|---|
| Individual | −4.4 | 1.6 | 1.0 | 10.4 | 7.4 |
| Six categories | −2.2 | 3.2 | 0.39 | 10.0 | 3.9 |
| Six categories + time (10)‡ | −3.2 | 3.4 | 0.39 | 10.3 | 3.8 |
| Six categories + time (4) | −2.0 | 3.1 | 0.39 | 10.4 | 3.9 |
| Six categories + mine | −5.9 | 2.4 | 0.53 | 10.0 | 5.2 |
| Model category + mine | −5.0 | 1.9 | 0.51 | 10.0 | 4.9 |
| External matrix | −2.0 | 2.4 | 0.47 | 10.2 | 5.0 |

* SE, standard error of the regression coefficient; SD, standard deviation.

† Pearson's correlation between average individual exposures using all exposure measurements and the exposure generated with a particular grouping strategy.

‡ Numbers in parentheses, number of categories.

**TABLE 5. Number of strata in the job exposure matrices, intragroup and intergroup variance in exposure, and median standard error of the mean, as calculated with an analysis of variance on log-transformed exposure measurements, National Study of Coal Workers' Pneumoconiosis, 1969–1981**

| Job exposure matrix | No. of strata | Variance components | | Median standard error |
|---|---|---|---|---|
| | | Intergroup | Intragroup | |
| Six categories | 6 | 0.22 | 0.27 | 0.06 |
| Six categories + time (10)* | 60 | 0.20 | 0.35 | 0.14 |
| Six categories + time (4) | 24 | 0.23 | 0.30 | 0.12 |
| Six categories + mine | 134 | 0.23 | 0.18 | 0.17 |
| Mine | 23 | 0.08 | 0.30 | 0.11 |
| Job | 103 | 0.16 | 0.29 | 0.25 |
| Job + mine | 628 | 0.28 | 0.15 | 0.28 |

* Numbers in parentheses, number of categories.

gories only, showed the smallest pooled standard error of 0.06. As expected, and as is clear from the table, the standard error increases as the number of cells involved in a JEM increases. For instance, the JEM using a classification by job and mine, with the largest number of categories, has a standard error that is almost five times greater, at 0.28, than that for the JEM with six categories only. The increase of the median standard error is accompanied by a slight increase of the intergroup variability. For instance, the simplest JEM with six exposure categories has an intergroup variance of 0.22, while the JEM of job by mine, with 628 exposure categories, had an intergroup variance of 0.28, indicating that an increase of the number of cells led to only a small increase in contrast between the categories of the exposure. The intragroup variance of 0.27 for the six categories was almost halved to 0.15 for the JEMs by mine and job, suggesting that, as expected, the groups became more homogeneous as the number of categories increased. The results for the other JEMs were in between the ones discussed. For instance, the JEM of six occupational categories by mine had intragroup and intergroup variances that were close to the strategy with a categorization by job and mine, but the median standard error seemed much smaller (SE, 0.17). This makes sense because the number of categories is five times smaller than that for the JEM with the finer grouping structure (134 cells compared with 628 cells; see table 5). Inclusion of a time axis led only to a small improvement compared with the JEM with six categories without a time axis. The intergroup variance remained almost constant while the intragroup variance unexpectedly became smaller, and the median standard error of the mean increased by a factor of more than two. A time axis in which only the first 3 years were distinguished, which were the years in which the exposure most clearly decreased, did not result in any improvement either. The categorization based on mine and six occupational categories had intergroup and intragroup variances that were almost equal to the categorization based on mine and job, while the median standard error of the mean was much smaller. This suggested that the JEM with 134 cells was almost as informative as the one based on 628 cells with regard to the homogeneity and contrast between exposure categories, with the advantage of smaller standard errors around the cell mean exposure levels.

## DISCUSSION

Results of this analysis clearly demonstrated exposure-response relations between average respirable coal dust exposure and lung function change. These results confirmed previous analyses of survey data for coal workers in the United States and other countries in the world. Clear differences were found between slopes of exposure-response relations, depending on the way the average exposure was calculated. For instance, a slope of –4.5 was found when $FEV_1$ was related to individual long-term average exposure using all measurements. When fewer measurements were used to calculate the individual mean exposure, a considerably lower slope was observed between $FEV_1$ and exposure, reducing by up to a factor of three when only three exposure measurements per worker were used.

The fact that measurement effort and exposure level were correlated to some extent (owing to the increased sampling undertaken on high-risk jobs) may have influenced the results from this study to some degree. In the main, it probably led to a better improvement in the attenuation with increasing sample number than might be observed in other studies lacking any correlation between sample number and exposure level. In other studies, the increased exposure measurement sample size would apply to all jobs, regardless of exposure level. In these cases, and where the variability in exposure increases with the mean exposure level, this added measurement effort is being inefficiently applied, since the jobs with low intrajob variability are

having their variances lowered further. High variance jobs are also having their variances lowered by the additional sampling, but by applying additional sampling effort to them instead of to the low variance jobs, more benefit could be obtained. In our study, this is what happened to some extent. High exposure level jobs (high variance jobs) received additional sampling, thereby giving more reliable mean values than if the extra sampling effort had been applied over all jobs equally.

It should be realized that the attenuated exposure-response relation found in this study for three measurements/worker was derived from more than 3,000 measurements. Few examples of epidemiologic studies exist in which such a measurement effort was realized. The attenuation in this study of coal workers was considerable when fewer measurements were used, because the intraindividual variability in exposure was three times larger than the interindividual variability in exposure (variance ratio of three), meaning that the "noise" variance was three times as large as the "signal." The influence of this noise became almost negligible when (on average) all 31 measurements per worker were used. Similar variance ratios have been found in settings with strongly varying exposure as in agriculture (14). Clearly smaller variance ratios have been found as well (15). An important difference between this study and others (14, 15) is that, although variance ratios are similar, the individual variance components are relatively small; for example, the differences in exposure from day to day between workers and within and between groups are small, most likely because of the effects of control technology in this industry.

In this study, the JEM using six occupational categories and mine showed the largest slope of lung function change on dust exposure. It also provided results comparable to those found using the means of individual exposures. Other JEMs, including the one recently proposed and based on the MSHA job-coding scheme by mine and year (12), were less successful than the simpler approach studied here. The slope of the exposure-response relation using other JEMs was smaller compared with the one found for the individual exposure assessment strategy. Interestingly, the standard error of the slope of the exposure-response relations was smaller when an individual average exposure was calculated than when exposure group average exposures were calculated with JEMs. This can be explained by unexplained variance in lung function within an occupational category, related to interindividual variability in exposure. This phenomenon had been described as early as the 1950s by Prais and Aitchison (16) and is a result of a Berkson type error. Another reason for the small standard error is the wider range in individual average exposures compared with the grouped exposure.

This study shows clearly the advantages and disadvantages of exposure assessment strategies based on grouping into homogeneous exposure categories compared with individual exposure assessment strategies. The latter can require a considerable measurement effort but can yield exposure-response relations with the smallest possible standard error if a larger population of coal workers could be studied. Grouping strategies will in most cases lead to reasonably unbiased estimates of the exposure-response relation, because the systematic overestimation or underestimation of the exposure of some of the group members leads to an error that is of the Berkson type (3, 16, 17). This is also illustrated by the analysis of the exposure-response relation of $FEV_1$ of dust level, calculated on the basis of a grouping by the six categories and mine, but using only 3,315 exposure measurements. The results of this analysis are in close range with results for the analysis with all measurements.

Taking all grouping strategies into account, we found that using job and mine seemed to lead to the most homogeneous exposure categories and most contrast among groups, although the precision was least of all others. In comparison, the external job exposure matrix did not perform as well, despite the effort that went into its derivation and the novel methods used to minimize random variation in the data. In this strategy, detailed three-way (mine/job/time) groupings were used for specificity, but broad groupings had to be used in addition, such as two-way classifications of occupation-year, mine-year, or broad occupational category groupings, to provide data to missing cells. A special formula was derived in order to incorporate the broader group data with three-way detailed classifications in order to minimize the variance of the mean exposure for strata with very few samples. Data from the external matrix demonstrated a relatively high correlation with the individual exposure data, suggesting that this approach was reasonably specific. However, the regression coefficient was the joint lowest of the seven listed there, while the standard error of the regression coefficient fell in the middle of the range in the table. Hence, it appears that, though the steps taken may have succeeded in improving the precision of the estimates of an average exposure, they did not succeed in reducing the attenuation in the regression coefficient. Overall, the advantages of a very specific grouping by job, mine, and time are traded off by a low precision of the estimates as generated by the matrix. This observation is supported by a recently conducted simulation study (18).

Most occupational hygienists and epidemiologists apply some rules of thumb principles that have been mentioned in the statistical literature earlier to obtain the optimal grouping structure (3, 15, 16, 19–21). This body of literature suggests that the performance of a grouping strategy in an epidemiologic study is determined by the intragroup variance (the homogeneity), the between group variance (the contrast), and the standard error of the mean exposure in a category (the precision) and the measurement effort (available number of exposure measurements per worker). Theoretically a grouping is optimal, if each exposure category is as homogeneous as possible with regard to the exposure, if as much contrast as possible exists between the categories, and if sufficient measurements per category have been taken so that the estimate of the average exposure in that category is relatively precise. This will be obtained by minimizing the interindividual variability within an exposure category (intragroup variability), maximizing the contrast between groups (intergroup variability), and minimizing the standard error of the mean in each exposure category. These requirements may be conflicting because there always are only a limited number of measurements available. Making categories as homogeneous as possible will generally lead to a large number of small categories of workers with similar occupations or who perform similar tasks. This, in turn, reduces the number of measurements per exposure category and thus the precision of the estimate of the average exposure. This apparent conflict among different requirements can only be solved by using quantitative expression for grouping strategies as already exists for individual-based exposure assessment strategies. Some formulas have recently been published for balanced data sets, assuming a constant variance over and within exposure categories (19, 20). Expressions applicable to a wide range of situations to assess the optimal quantitative strategy for exposure assessment in a specific situation are not yet available.

In this study, we applied compliance data for epidemiologic purposes. The possibility of bias in these data, especially suspicions about tampering, has led to uncertainties in interpretation of earlier derived epidemiologic findings. However, a recent study concludes that, although there is evidence of bias in the exposure assessment, the magnitude of this bias will be of the order of only 10–15 percent for those mines included in this study (8). This order of magnitude is considerably smaller than bias introduced by using fewer samples when using an individual exposure assessment strategy or by calculating group average exposures in different ways.

The results of this study confirm general guidelines found in the literature and suggest that exposure assessment strategies can be designed and optimized to some extent a priori. Information about the variability of the exposure and its structure in terms of variance components by individual workers or occupational categories should be present, for instance, by performing limited scale pilot studies in which repeated measurements are being taken. The findings illustrate the need to undertake epidemiologic analysis of exposure-response relations, applying multiple exposure assessment strategies and analyzing the variability of the exposure at the same time. The results of this study also warrant explicit publication of exposure assessment strategies for epidemiologic studies. The results can be used to optimize study designs by influencing intraindividual, interindividual, intragroup, or intergroup variance components. The values of these variance components can be influenced by changing the number of exposure groups or by including high or low exposed populations.

## REFERENCES


1. Stewart P, Herrick B. Issues in performing retrospective exposure assessment. Appl Occup Environ Hyg 1991;6:421–7.
2. Armstrong BG. The effects of measurement error on relative risk regressions. Am J Epidemiol 1990;132:1176–84.
3. Liu K, Stamler J, Dyer A, et al. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. J Chronic Dis 1978;31:399–418.
4. Rappaport SM, Symanski E, Yager JW, et al. The relationship between environmental monitoring and biological markers in exposure assessment. Environ Health Perspect 1995;103:49–53.
5. Lee-Feldstein A. A comparison of several measures of exposure to arsenic. Matched case-control study of copper smelter employees. Am J Epidemiol 1989;129:112–24.
6. Blair A, Stewart PA. Do quantitative exposure assessments improve risk estimates in occupational studies of cancer? Am J Ind Med 1992;21:53–6.
7. Attfield MD. Longitudinal decline in $FEV_1$ in United States coalminers. Thorax 1985;40:132–7.
8. Attfield MD, Hearl FJ. Application of data on compliance to epidemiological assessment of exposure-response: the case of data on exposure of United States coal miners. Occup Hyg 1996;3:177–84.
9. Boden LI. Evaluating dust exposure using operator-reported data. Ann Am Conf Gov Ind Hyg 1986;14:521–6.
10. Seixas NS, Robins TG, Rice CH, et al. Assessment of biases in the application of MSHA compliance data to an epidemiological study. Am Ind Hyg Assoc J 1990;51:534–40.
11. Attfield MD, Morring K. The derivation of estimated dust exposures for US coal miners working before 1970. Am Ind Hyg Assoc J 1992;53:248–55.
12. Seixas NS, Moulton LH, Robins TG, et al. Estimation of cumulative exposures for the National Study of Coal Workers' Pneumoconiosis. Appl Occup Environ Hyg 1991;6:1032–41.
13. Kromhout H, Heederik D. Occupational epidemiology in the rubber industry: implications of exposure variability. Am J Ind Med 1995;27:171–85.
14. Preller L, Kromhout H, Heederik D, et al. Modeling long-term average exposure in occupational exposure-response analysis. Scand J Work Environ Health 1995;21:504–12.
15. Heederik D, Boleij JSM, Kromhout H, et al. Use and analysis

of exposure monitoring data in occupational epidemiology: an example of an epidemiological study in the Dutch animal food industry. Appl Occup Hyg 1991;6:458–64.
16. Prais SJ, Aitchison J. The grouping of observations in regression analysis. Rev Int Stat Inst 1954;1:1–22.
17. Durbin J. Errors in variables. Rev Int Stat Inst 1954;1:23–32.
18. Heederik D, Braun W, Kromhout H. The influence of random exposure estimation error on the exposure-response relationship when grouping into homogeneous exposure categories. Occup Hyg 1997;3:229–41.
19. Ashford JR. The design of a long-term sampling program to measure the hazard associated with an industrial environment. J R Stat Soc (A) 1958;121:333–47.
20. Tielemans E, Kupper LL, Kromhout H, et al. Individual based and group based exposure assessment: some equations to evaluate different strategies. Ann Occup Hyg 1998;42:115–19.
21. Oldham PD, Roach SA. A sampling procedure for measuring industrial dust exposure. Br J Ind Med 1952;9:112–19.