Applied Occupational and Environmental Hygiene

ISSN: 1047-322X (Print) 1521-0898 (Online) Journal homepage: https://www.tandfonline.com/loi/uaoh20

Effect of Different Grouping Strategies in Developing Estimates of Personal Exposures: Specificity Versus Precision

Mark A. Werner & Michael D. Attfield

To cite this article: Mark A. Werner & Michael D. Attfield (2000) Effect of Different Grouping Strategies in Developing Estimates of Personal Exposures: Specificity Versus Precision, Applied Occupational and Environmental Hygiene, 15:1, 21-25, DOI: 10.1080/104732200301818

To link to this article: https://doi.org/10.1080/104732200301818

Published online: 30 Nov 2010.

Applied Occupational and Environmental Hygiene
Volume 15(1): 21–25, 2000
1047-322X/00 $12.00 + .00

Effect of Different Grouping Strategies in Developing Estimates of Personal Exposures: Specificity Versus Precision

Mark A. Werner[1] and Michael D. Attfield[2]
[1] *Wisconsin Division of Public Health, Madison, Wisconsin;* [2] *Field Studies Branch, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, Morgantown, West Virginia*

There has been increasing interest in recent years in developing strategies in epidemiology for the summarization of occupational exposures, strategies that serve to clarify observed relationships between occupational exposure and health outcomes. Where source occupational exposure data are scarce, it is common to assemble exposure groups with the goal of increasing the extent to which data-based exposure estimates are available for an entire cohort. There has been little guidance, however, on the effect of different grouping strategies on the observed fit between exposure and health outcome. This investigation examined the effect of the use of different exposure summarization strategies on observed relationships between dust exposure and lung function decline among coal mine workers. The dust exposure and spirometry data employed were gathered in the National Study of Coalworkers' Pneumoconiosis. An analysis of variance procedure was carried out to characterize the variability of the dust exposure data, employing single variables relating to mine identity, occupation, and year, as well as two- and three-way multivariate combinations of these variables. The resulting combinations were ranked according to the standard deviation of the observed exposure range to reflect the relative specificity of the various approaches. Sequential arrangements of single- and multiple-variable combinations were constructed, alternately employing highly specific codes or broad categories for mine, occupation, and year. Annual exposure estimates were constructed on the basis of these sequences and used in tandem with longitudinal change in forced expiratory volume (FEV_1) in linear multiple regression procedures. Height, age, smoking status, and dust exposure were employed as predictor variables. The results show that the use of broad categorization approaches had a substantial impact on observed regression coefficients. The largest change was observed for categorization according to occupation, which resulted in two- to three-fold increases in the magnitude of observed regression coefficients. These results suggest that the use of highly specific exposure summarization approaches may result in regression outcomes which are marked by a high degree of attenuation, and that consideration of the precision of summarized exposure estimates is an important component of an effective exposure assessment strategy.

Keywords Occupational Exposure Assessment, Occupational Epidemiology, Coal Mining

There has been increasing interest in recent years in how occupational exposure data are used in developing quantitative exposure-response relationships for the risk of adverse health outcomes.$^{(1-3)}$ In many cases, the ability to obtain reliable quantitative estimates of exposure for exposed individuals is hampered by the scarcity of exposure data. When data are limited, it is common to pool exposure data across individuals to create grouped estimates. This use of categorization allows for the creation of exposure estimates based upon a larger number of samples, reducing the amount of random variability associated with the estimate. These estimates can also be applied to individuals likely to have similar exposures but for whom no measurements exist. One disadvantage to this general approach is that the grouped estimate of exposure may not be truly applicable to all individuals in the group, resulting in misclassification of exposure for those people.

The use of exposure grouping strategies is carried out with two competing interests in mind. First, the ideal grouping strategy would result in the creation of exposure groups that are as large as possible, thereby leading to exposure estimates that are substantially more precise than the individual worker mean exposures they replace. Second, the ideal strategy would strive for the greatest homogeneity within the groups, at the same time maximizing the differences between groups. The ability of a grouping strategy to retain true individual differences in exposure among individuals may be termed *specificity*, with outcomes characterized by within-group, between-worker variability that is low compared to between-group variability. In practice, however, efforts to gain precision generally result in a loss of specificity, and attempts to retain specificity tend to create groups with imprecise estimates of exposure.

The work described here represents an attempt to examine the trade-off between specificity and precision in constructing estimates of workplace dust exposures. This exercise was undertaken to provide insight into the utility of grouping strategies in improving observed exposure-response relationships, and to aid exposure assessors and epidemiologists in gauging the relative degree to which imprecision and inaccuracy in exposure estimation may be overcome in their work.

METHODS

Sources of Data

This exercise was carried out with data from two primary sources: (1) medical and work history data from the National Study of Coalworkers' Pneumoconiosis (NSCWP)—a large epidemiological study of lung disease in coal miners; and, (2) personal respirable dust exposure data obtained from mine operators as required by the 1969 Federal Coal Mine Health and Safety Act (FCMHSA).[4] This law requires that mine operators carry out environmental monitoring in coal mines and submit the results to the U.S. Department of Labor.

Medical Data

Information on the health status of coal mine workers was obtained from the first and third rounds of medical examinations of the NSCWP.[5] The first round of medical examinations was administered between 1969 and 1971, with subsequent rounds of examinations performed in 1972–1975, 1978–1981, and 1985–1988. The study involved the collection of radiological data, as well as data on ventilatory function, respiratory symptoms, height, age, smoking status, and work history. Measures of forced expiratory volume (FEV_1) were obtained by use of a rolling seal volume spirometer, with the highest of five forced expiratory maneuvers utilized (methods are fully described by Morgan and colleagues).[6] Longitudinal change in lung function was measured as the difference between FEV_1 measurements taken in Round 1 (1969–1972) and Round 3 (1978–1981) of the NSCWP divided by the time interval between the two surveys. These particular rounds were selected because of the availability of large amounts of exposure data for the interval between these rounds. In addition, the number of subjects examined in rounds 1 and 3 was considered sufficiently large to permit the detection of an exposure-response relationship. Work history information was taken from that supplied by individuals at the third round, and consisted of records of each job worked, with starting and ending dates. The work history was coded according to the scheme used by the Mine Health and Safety Administration (MSHA).[7]

Environmental Data

Construction of dust exposure estimates was carried out on the basis of data submitted to MSHA by mine operators. These data have been described previously.[8–10] Data employed for analysis included only those samples collected in the time interval between the administration of ventilatory function testing in rounds 1 and 3 of the NSCWP for each particular mine. When limited to data collected between 1970 and 1979, this database yielded over 36,000 data points for initial consideration in this study.

Construction of Exposure Estimates

The primary goal of this step was to derive and compare quantitative dust exposure estimates based on strategies involving various degrees of specificity and precision. This was done by first identifying three major dimensions of variability in exposure level corresponding to *mine*, *job*, and *year*. Each of these dimensions was then summarized into many groups corresponding to a highly specific breakdown of the data, and also into a few groups representing low specificity but high precision. The groups so derived were then linked with the work history records to obtain numerical exposure estimates for each miner for each strategy. The exposures were then tabulated and analyzed, and also correlated with the medical data, to explore how different levels of specificity and precision affected the exposure-response relationship.

Each exposure dimension was summarized as follows:

Mine Category

Initial classification was made according to mine identification numbers assigned at the outset of the NSCWP. In this way, each mine was represented separately to give the most specific derivation of exposures. A much coarser, less specific grouping was constructed on the basis of four groupings according to the carbon content of the coal and mine location (as previously described by Attfield and Morring).[11]

- Medium/low bituminous coal (>88% carbon; central Pennsylvania and southeastern West Virginia)
- High volatile "A" bituminous coal (80–87% carbon; Alabama, eastern Kentucky, eastern Ohio, western Pennsylvania, southwestern Virginia, and northern and southwestern West Virginia)
- High volatile Midwest (Illinois and western Kentucky)
- High volatile West (Colorado and Utah).

Job

Jobs were initially classified according to each specific occupation using the standard MSHA job classification scheme for coal mines to give the most specific breakdown.[7] For comparison, a four-part grouping was used in which categories were constructed according to the primary work location as follows:

- all underground face jobs
- all underground non-face jobs including transportation
- all surface jobs
- all supervisory jobs (regardless of location)

Year

Temporal categories were initially created according to each of the 10 distinct years from 1970 to 1979 to give the most specific breakdown. For comparison, four categories were created to represent time periods of uniform exposure. A previous examination of the dust exposure data had revealed a decreasing trend in the dust exposure of coal mine workers in the first half of the 1970s. The sharpest decline was observed from 1970–1972, and no clear trend observed between 1973 and 1979.[12] With this in mind, the following four categories were constructed:

- 1970
- 1971
- 1972
- 1973–1979

In pursuing this approach, it was found that for the highly specific approaches utilizing all three exposure dimensions, exposure data were not available for all mine/occupation/year cells represented in the collective work history of the cohort. For this reason, it was necessary to adopt a procedure by which numerical exposure estimates could be derived for those cells unfilled by the mine/occupation/year combination. For these unfilled cells, successively lower-order combinations (e.g., mine/year, mine/occupation, mine, etc.) were employed. To specify the sequence in which lower-order combinations should be applied, the dust exposure data were subjected to an analysis of variance procedure in which each combination of the three dimensions was examined. The various combinations were then ranked according to the standard deviation of the observed range of cell-specific mean exposures. This statistic relates to the degree to which the observed range of mean exposures was preserved by the summarization, and is used here as an index of specificity. In this way, work histories were matched with exposure estimates based on successively less specific strategies until all cells of the exposure matrix were filled.

Using the fine and coarse groupings of each of the three exposure dimensions gave rise to eight overall estimation strategies. These are summarized in Table I (M = mine [high specificity], MC = mine categorized [low specificity]; O = occupation [high specificity], OC = occupation categorized [low specificity]; Y = year [high specificity], YC = year categorized [low specificity]).

The sets of exposure estimates resulting from these eight strategies were summed by year for the time interval between medical examinations. The resulting cumulative exposure estimates were used in tandem with information from medical examinations in linear multiple regression procedures. Age, height, smoking status, and estimated dust exposure were employed as independent variables, and change in FEV_1 from Round 1 to Round 3 of the NSCWP was used as the dependent variable.

RESULTS

Table II displays two key statistical parameters for summarization strategies based on seven exposure-related variables and multivariate combinations. The parameters shown are the standard deviation of the exposure range and the standard error of the group means, which serve as indices of specificity and precision, respectively, with a large standard deviation suggestive of a strategy with high specificity and a small standard error of group means suggestive of a strategy with high precision. Based on these results, the sequence of variables used in each strategy was created according to decreasing specificity (as shown in Table I).

Data from the eight strategies employed, including values for indices of precision and specificity, are shown in Table III. The information presented there shows the varying specificity across the different exposure dimensions, as well as the decreasing specificity that occurs with use of fewer exposure dimensions.

TABLE I
Description of the eight grouping strategies employed to investigate specificity and precision in this study[A]

	Iteration[B]						
Strategy	1	2	3	4	5	6	7
1	M*O*Y	M*O	O*Y	O	M*Y	M	Y
2	MC*O*Y	MC*O	O*Y	O	MC*Y	MC	Y
3	M*OC*Y	M*OC	OC*Y	OC	M*Y	M	Y
4	MC*OC*Y	MC*OC	OC*Y	OC	MC*Y	MC	Y
5	M*O*YC	M*O	O*YC	O	M*YC	M	YC
6	MC*O*YC	MC*O	O*YC	O	MC*YC	MC	YC
7	M*OC*YC	M*YC	OC*YC	OC	M*YC	M	YC
8	MC*OC*YC	MC*YC	OC*YC	OC	MC*YC	MC	YC

[A]The sequences in which the seven variable combinations are arranged are also shown for each of the eight strategies.
[B]M—Mine code (MSHA); MC—Mine category; O—Occupation code (MSHA); OC—Occupation category; Y—Calender year; YC—Year category.

TABLE II
Values for statistical parameters from summarization strategies for coal dust exposure data based on seven exposure-related variables and multivariate combinations[A]

Exposure dimension combination	Standard deviation of exposure range	Standard error of group means
Mine + Occupation + Year	8.1	5.4
Mine + Occupation	6.7	2.8
Occupation + Year	6.4	2.2
Occupation	6.1	0.9
Mine + Year	5.7	1.8
Mine	3.9	0.9
Year	3.0	0.3

[A]the parameters shown are the standard deviation of the exposure range and the standard error of the group means, which serve as indices of specificity and precision, respectively. Note that an optimal strategy would seek to minimize the precision index and maximize the specificity index.

The results of the regression analyses derived from use of the exposure estimates generated by each strategy are presented in Table IV. The observed regression coefficients for dust exposure vary by greater than a factor of three (-2.02×10^{-4} to -6.79×10^{-4}). Regression coefficients for dust exposure differed significantly from zero ($\alpha = 0.05$) for two of the eight sets of exposure estimates. Both of these sets employed specific codes for mine designation and broad categorization for occupation.

DISCUSSION

This investigation was undertaken to explore the effect of different exposure grouping strategies, reflecting various levels of specificity and precision of grouped cell means, on observed exposure-response relationships in a cohort for which a relationship between exposure and response had been previously demonstrated. These findings suggest that such differences may lead to the derivation of quantitative exposure-response relationships that vary by factors of two or more. Inspection of Table IV reveals the following trends:

- The replacement of *mine*-specific codes with broad mine categories results in a *moderate decrease* in the fit between exposure and response;
- The replacement of *occupation*-specific codes with broad occupational categories results in a *substantial increase* in the fit between exposure and response; and
- The replacement of individual calendar *year* with broad year categories results in a *slight decrease* in the fit between exposure and response.

The observation that the use of mine-specific codes and calendar year is superior to broad mine and year categories in fitting exposure-response relationships suggests that for these variables as defined, differences in exposure across mines and across time are best accounted for by applying the more specific designator. In addition, the finding that the use of occupational category improves the observed exposure-response relationship suggests that any potential gain in fit which may be attained by using the more specific code is offset by the loss of precision which results from the use of estimates based on smaller number of samples. Of the three coded variables, it would be expected that the use of broad categories for occupation would be most beneficial for two reasons. First, the four categories selected are likely to represent groups of workers for whom exposures are similar in magnitude and in the physical characteristics of airborne dust (e.g., particle size and moisture content). In addition, it is in the use of categorization for occupation that the greatest decrease is seen in the number of exposure estimates, and the largest increase in the number of dust samples per cell. It is likely that both of these factors contribute to the increase in fit between exposure and response observed with the use of occupational categories.

Care should be taken in applying the results of this exercise in designing exposure summarization strategies for use in other industries.[13] This work seeks to characterize the effect of exposure grouping strategies on observed exposure-reponse relationships for workers in the U.S. coal mining industry. This industry is characterized by a large number of workplaces which, when compared with other industries, are quite similar with respect to the magnitude and nature of exposure and workplace organization. The exceptionally large number of samples in the data set and the use of dust exposure collected by mine operators and submitted to MSHA for regulatory purposes also represent unique features of this investigation.

TABLE III
Indices of precision (within-group standard error = *S.E.*) and specificity (between-worker standard deviation = *S.D.*) for exposure estimates obtained from applying sequential grouping strategies employing different combinations of variables[A]

	Year			Year category	
	S.E.	S.D.		S.E.	S.D.
			Mine		
Occupation	0.41	0.61	Occupation	0.46	0.61
Occup. Cat.	0.23	0.50	Occup. Cat.	0.23	0.51
			Mine category		
Occupation	0.40	0.52	Occupation	0.41	0.51
Occup. Cat.	0.13	0.42	Occup. Cat.	0.13	0.43

[A]Note that an optimal strategy would seek to minimize the precision index and maximize the specificity index.

TABLE IV
Results of regression analyses relating exposure estimates constructed using different strategies versus longitudinal change in FEV_1[A]

	Year				Year category		
	Pr > F	Coeff. ($\times 10^{-4}$)	Std err		Pr > F	Coeff. ($\times 10^{-4}$)	Std err
			Mine				
Occupation	0.1235	−3.25	2.11	Occupation	0.1936	−2.81	2.08
Occup. Cat.	**0.0072**	**−6.79**	2.54	Occup. Cat.	**0.0110**	**−6.37**	2.50
			Mine category				
Occupation	0.3548	−2.35	2.54	Occupation	0.4338	−2.02	2.58
Occup. Cat.	0.0521	−6.18	3.18	Occup. Cat.	0.0629	−5.85	3.14

[A]Regression coefficients significantly different from zero and their accompanying p-values are displayed in bold.

CONCLUSIONS

This investigation was carried out to examine the effect of different exposure grouping strategies on observed relationships between dust exposure and lung function decline among coal mine workers. The results suggest that both precision and specificity must be considered in designing exposure metrics for use in occupational investigations, and that the use of grouping strategies which seek precise exposure estimates at the expense of specificity may be superior to more specific estimates to achieve this. Consideration should be given to the a priori development of exposure assessment strategies for occupational epidemiology that are cognizant of the role of precision and specificity in seeking to optimize observed exposure-response relationships.

REFERENCES

1. Tielemans, E.; Kupper, L.L.; Kromhout, H.; et al.: Individual-Based and Group-Based Occupational Exposure Assessment: Some Equations to Evaluate Different Strategies. Ann Occup Hyg 42:115–119 (1998).
2. Rappaport, S.M.; Kromhout, H.; Symanski, E.: Variation of Exposure Between Workers in Homogeneous Exposure Groups. Am Ind Hyg Assoc J 54:654–662 (1993).
3. Kromhout, H.; Tielemans, E.; Preller, L.; et al.: Estimates of Individual Dose from Current Exposure Measurements. Occup Hyg 3:23–29 (1996).
4. Coal Mine Health and Safety Act, Public Law 91-173 (S. 2917), December 30, 1969 (1969).
5. Attfield, M.A.; Castellan, R.M.: Epidemiological Data on U.S. Coalminers' Pneumoconiosis, 1960 to 1988. Am J Publ Health 82:964–970 (1992).
6. Morgan, W.K.C.; Handelsman, L.; Kibelstis, J.; et al.: Ventilatory Capacity and Lung Volumes of U.S. Coal Miners. Arch Environ Health 28:182–189 (1974).
7. U.S. Department of Labor, Mine Safety and Health Administration: Coal Mining Occupations. MSHA Form 2000-169, MSHA, Washington, D.C. (1984).
8. Seixas, N.: Estimation of Exposures to Respirable Coal Mine Dust and an Evaluation of Its Effect on Respiratory Health. Ph.D. dissertation, Department of Environmental and Industrial Health, University of Michigan, Ann Arbor (1990).
9. Seixas, N.; Robins, T.G.; Rice, C.H.; et al.: Assessment of Biases in the Application of MSHA Compliance Data to an Epidemiologic Study. Am Ind Hyg Assoc J 51:534–540 (1990).
10. Seixas, N.S.; Moulton, L.H.; Robins, T.G.; et al.: Estimation of Cumulative Exposures for the National Study of Coalworkers' Pneumoconiosis. Appl Occup Environ Hyg 6:1032–1041 (1991).
11. Attfield, M.D.; Morring, K.: Investigation into the Relationship Between Coalworkers' Pneumoconiosis and Dust Exposure in U.S. Coal Miners. Am Ind Hyg Assoc J 53:486–492 (1992).
12. Attfield, M.D.; Morring, K.: Derivation of Estimated Dust Exposures for U.S. Coal Miners Working Before 1970. Am Ind Hyg Assoc J 53:248–255 (1992).
13. Heederik, D.; Attfield, M.D.: Characterization of Cumulative Exposure for the Study of Chronic Occupational Disease (1999).