Applied Occupational and Environmental Hygiene

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

Use and Analysis of Exposure Monitoring Data in Occupational Epidemiology: An Example of an Epidemiological Study in the Dutch Animal Food Industry

Dick Heederik , Jan S.M. Boleij , Hans Kromhout & Tjabe Smid

To cite this article: Dick Heederik , Jan S.M. Boleij , Hans Kromhout & Tjabe Smid (1991) Use and Analysis of Exposure Monitoring Data in Occupational Epidemiology: An Example of an Epidemiological Study in the Dutch Animal Food Industry, Applied Occupational and Environmental Hygiene, 6:6, 458-464, DOI: 10.1080/1047322X.1991.10387913

To link to this article: https://doi.org/10.1080/1047322X.1991.10387913

Published online: 24 Feb 2011.

Use and Analysis of Exposure Monitoring Data in Occupational Epidemiology: An Example of an Epidemiological Study in the Dutch Animal Food Industry

Dick Heederik,[A] Jan S.M. Boleij,[B] Hans Kromhout,[B] and Tjabe Smid[A,B]
[A]Department of Epidemiology and Public Health, Wageningen University, P.O. Box 238, 6700 AE Wageningen, The Netherlands; [B]Department of Air Pollution, Wageningen University

No general, theoretical framework exists for exposure monitoring strategies for epidemiological purposes nor for the analysis of these data. It is, however, recognized that the variation in exposure over time and between workers is an important element of consideration in the design of measurement strategies and the analysis of exposure data. In this paper, the partitioning of exposure variability is discussed for two types of monitoring strategies. First, a monitoring strategy is considered in which all members of the study population are monitored on various occasions. In this strategy, the ratio of within and between worker variance gives information about the magnitude of underestimation of the exposure–response relationship. The second strategy discussed is one based on "homogeneous" exposure categories. This strategy is illustrated with a study among animal feed workers in The Netherlands. It is shown that the grouping which results from this strategy can be applied for epidemiological purposes successfully if the ratio of the within and between category variance in exposure is relatively small. This implies that it is important to optimize the ratio of these variances in a given study population instead of composing strictly homogeneous exposure categories. In both strategies, the analysis of variance of repeated exposure measurements on the same individuals plays a crucial role. It is recommended that repeated measurements on the same individual should be taken routinely. The analysis of variance should be used as a tool to analyze such data in order to optimize a measurement strategy or a categorization of the population in exposure groups. **Heederik, D.; Boleij, J.S.M.; Kromhout, H.; Smid, T.: Use and Analysis of Exposure Monitoring Data in Occupational Epidemiology: An Example of an Epidemiological Study in the Dutch Animal Food Industry. Appl. Occup. Environ. Hyg. 6:458–464; 1991.**

Introduction

In the past, studies of exposure–response relationships in the work environment were focused on situations in which a substantial exposure and a resulting substantial health risk occurred. Rare monocausal diseases associated with a specific agent were easily detected, e.g., pleural mesothelioma caused by asbestos. Simple approaches of exposure assessments based on accidentally collected information generally are not sufficient anymore. However, when perusing the current literature, one tends to conclude that exposure assessment for prospective epidemiologic studies is still in its infancy. Few examples are found that describe epidemiologic studies with *a priori*-designed exposure assessment strategies.

The possibility of correctly identifying associations between exposure and health effects is partly determined by the health relevance of the exposure measurement. Exposures and health effects are usually linked through the concept of dose.

It is obvious that knowledge of the disease process should influence environmental monitoring strategies. However, in many situations, *a priori* information about disease processes and, thus, the relevant dose is not present. For example, very limited information exists for grain dust about the relevant exposure in relation to the development of lung function changes. Therefore, one often assumes that cumulative exposure is the best exposure index to estimate and temporal patterns in exposure are neglected. The most useful approach, in such cases, is to compare different exposure indices in the analysis.

Exposure Assessment Strategies

In occupational epidemiological studies, several methods are being used to characterize the exposure of the population under study. Several authors[1,2] provide a hierarchy of exposure characterization methods for use in epidemiological studies. These factoring techniques range from specific exposure measurements to the semiquantitative estimation of specific exposure (ranking) with such surrogates as occupational title, occupational status, de-

 1047-322X/91/0606-458$2.00/7 © 1991 AIH

partment, or sector of industry. Basically, there are two approaches to consider measured exposure information. In the first, all population members are monitored. In the second, only subgroups, sometimes referred to as homogeneous exposure groups, are monitored. The composition of a homogeneous subgroup can be made on the basis of information on a subject's job or by an estimation technique. Semiquantitative exposure estimation is the process of estimating a subject's exposure on the basis of present knowledge on such items as tasks performed and processes applied. This second method is often employed because it is simple, comprehensive, and efficient and because it leads to a reduction in the measurement effort when applied in combination with strategies in which the exposure is actually measured.[3] In this case, the grouping takes place before measurements are taken. An alternative is an *a posteriori* grouping based on randomly sampled exposure measurements. For the purpose of this paper, no distinction will be made between *a priori* and *a posteriori* grouping strategies.

Examples in which detailed, quantitative, exposure information is gathered are given by studies among workers exposed to ionizing radiation in nuclear power plants and other nuclear facilities. Personal monitoring is done permanently during working hours because personal dosimetry is a technically simple method for this particular exposure. The exposures can be expressed as doses from external sources to various parts of the body or the entire body, to the skin of the entire body, and to the extremities.

More often, strategies are applied in which exposure is established for subgroups of workers. Such procedures reduce the measurement effort considerably because only a sample of workers needs to be monitored. This is favorable when sampling techniques are time-consuming and expensive to apply. An early description of such a strategy has been given by Oldham and Roach[4] and Ashford.[5] These authors describe the statistical background of the measurement strategy applied in the Pneumoconiosis Field Research of the British Coal Board. Colliery populations were divided into homogeneous subgroups on the basis of their occupation, place of work, and shift, or the so called "man-shift." This meant that dust exposure measurements were taken for a random sample of all workers employed during a certain shift. Time exposed was measured for all workers based on colliery records. The cumulative exposure was calculated as the product of the exposure measured for a certain occupational group and the time exposed. The sampling effort was allocated to certain man-shifts based on several factors such as the standard deviation in exposure for the particular homogeneous subgroup, the labor turnover of a group, and the number of individuals assigned to the group. An underlying assumption of this sampling strategy is that the exposure of an individual worker is supposed to be equal to the shift average of the whole group. Surveys in different occupationally exposed populations have shown that considerable inhomogeneity may exist between workers.[3,6] The principal reasons for the lack of homogeneity are an inappropriate definition of the subgroup and differences in work practice and environmental conditions. The latter may occur, for instance, when a broad definition of a subject's function or job is used while workers with the same title job or function perform different tasks.

Misclassification of Exposure

In epidemiology, nondifferential and differential misclassification are distinguished. Both are forms of information bias.[7] In the case of nondifferential misclassification, misclassification is independent of the health effect. Differential misclassification of exposure can lead to underestimation as well as overestimation of the effect of exposure on health, depending upon the type of study and the exact form of differential misclassification.[7] Generally, nondifferential misclassification of exposure leads to underestimation of the effect of exposure on the health parameter studied. Large differences in concentrations of an air pollutant in time and space lead to a large inter- and intraindividual (or day-to-day) variation in the worker's exposure. This can be regarded as a specific form of nondifferential misclassification.

First, the influence of the nondifferential exposure estimation error will be illustrated for a situation in which repeated exposure measurements are present for all the members of the population under study with no control group present. Consider an epidemiological study in which the mean personal exposure to some agent has to be related to a health endpoint. A limited number of measurements per worker is, at best, an approximation of that worker's long-term exposure if a considerable within person or day-to-day variation in exposure exists. It has been shown that this within person variation in exposure can result in a considerable bias of the exposure–response relationship.[8,9] The statistical background has been recognized in the statistical literature and is referred to as attenuation.[10,11] The bias that occurs when a dependent variable measured on a continuous scale is regressed on an imperfectly measured exposure variable can be described by the ratio, λ, of the within and between worker variability in exposure:

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

where: b = expected value of the empirical regression coefficient of a dependent variable Y on an independent variable X which is an imperfect measure of the independent variable X
β = the true regression coefficient of Y on X
$\lambda = \sigma_w^2/\sigma_b^2$
σ_w^2 = within subject variance
σ_b^2 = between subject variance

If repeated exposure measurements on an individual are present, it is possible to estimate the within and between worker components of variance by an analysis of variance (ANOVA). The structure of an analysis of variance table, together with the estimated parameters, can be found in the literature[12] and is given in the Appendix to this paper.

If, for instance, $\lambda = 1$, which means that the within and between subject variances in exposure are equal, the empirical regression coefficient underestimates the true coefficient by 50 percent. The equation clearly shows that the underestimation is a function of both the within and the between worker variance in exposure. In a situation that has a high within worker variability in exposure, almost unbiased regression coefficients are obtained as long as the between worker variability in exposure (e.g., in the population) is a multitude of the within worker variability in exposure. This implies that a strict rule cannot be formulated for an acceptable or required within or between worker variance. The ratio between the two variance components determines the magnitude of the underestimation rather than the absolute values. Before starting a large epidemiological study, information about the partitioning of the exposure variability is necessary for further planning. Suppose a population with a large λ or a relatively large within worker variance in exposure is the object of a study. The exposure measurement strategy can then be modified in three ways to avoid bias of the exposure–response relationship. First, one can increase the measurement effort to attain a reliable estimate of the personal exposure with a minimal bias due to within person variance in exposure. This effect can be calculated according to the mathematics described by Liu *et al.*:[13]

$$k = \left(\frac{p}{1-p}\right)\lambda \qquad (2)$$

where: k = the number of repeated measurements per worker
p = the bias in a regression coefficient

For instance, when $\lambda = 6$, one needs at least $k = 54$ measurements per worker in order to reduce the bias in a regression coefficient to maximally 10 percent ($p = 0.90$). If this first approach leads to a measurement effort that is too large, repeated measurements in a subsample of workers can be performed to obtain a precise estimate of within and between person variability, after which a correction for attenuation can be applied.[8,14] If the first two approaches prove inadequate, a completely different study design can be chosen, e.g., by maximizing the contrast in exposure by means of an external control group, which basically means increasing the between worker variance in exposure.

In a grouping strategy, one assumes that subjects in an exposure group have a comparable exposure profile and that the mean exposure can be used to describe the exposure of all group members. However, occupational hygiene surveys have shown that considerable differences in mean exposure may exist between workers. This indicates that the actual exposures for some members of the homogeneous exposure group have been under- or overestimated. In analogy with the previous section, it seems clear that the underestimation of an exposure–response relationship is a function of the within and between group variation in exposure. A within group variance which is larger than the between group variance means that the exposure distributions of the groups are strongly overlapping and that the grouping did not lead to any distinction in exposure between groups compared with the whole population. If the between group variance is larger than the within group variance, distributions do not or only partially overlap. An example of the magnitude of underestimation of the exposure–response relationship has been reported by Heederik and Miller[14] in an epidemiological study among British coal miners. They estimated that up to 30 percent of the total variance of the cumulative dust exposure might be due to inhomogeneity of exposure groups. Correction for this inhomogeneity showed that the exposure–response relationship between cumulative coal dust exposure and eight-year lung function changes was underestimated by approximately 200 percent.

This suggests that, in general, underestimation of an exposure–response relationship is a function of two variances, the within and between person or group variance, and that the underestimation cannot be minimized by an absolute criterion for one of the two. The two basic monitoring strategies require a specific procedure in the analysis of the data. In the first strategy, in which all population members are being monitored, primary emphasis must be placed on the calculation of within and between worker variances and the ratio of these two to establish the potential bias in the relationship between exposure and response due to attenuation. The second approach, based on monitoring of homogeneous subsamples of workers, also requires the calculation of within and between worker variances; however, in this case, the between worker variance is the primary aim of the analysis of variance because it gives an impression of inhomogeneity of an exposure group and can give information about the overlap between exposure groups. A comparison of the variance ratio, λ, of the whole group with the variance ratio of subsets of workers gives information about the homogeneity of subgroups relative to the entire population. The calculation is similar to the calculation of the within and between worker variance (see Appendix). A similar comparison can be based on the 95 percent confidence limits of the distribution of worker mean exposures, $R_{0.95}$, as a measure of homogeneity of the group.[15] Partitioning of exposure variability into between and within exposure group variance can be derived by a nested type of analysis of variance.

The Animal Food Industry Example

In The Netherlands, approximately 6000 people are employed in animal feed production. Several raw materials are used such as grain, pulses, and waste products of the vegetable oil and starch industry. A large-scale study was initiated in 1985 and is still in progress. The primary aim of this project was to study the influence of occupational dust exposure in animal feed mills on respiratory symptoms and lung function. Dust exposure of workers in eight production and storage facilities was studied for 1 to 24 days, in most cases, in two seasonal periods (spring and

TABLE I. Characteristics of Personal Dust Concentrations by Job Category

Job Category	n	j	AM (mg/m³)	GM (mg/m³)	GSD	Range (mg/m³)
Unloaders	69	29	30.0	9.8	4.6	0.2–450
Crane drivers	23	12	6.1	2.5	4.7	0.2–27
Facility operators	63	20	1.4	0.8	2.7	0.2–9.4
Press operators	54	17	4.2	1.3	3.2	0.2–100
Production managers	42	16	3.8	1.6	3.5	0.2–49
Expedition workers	61	18	2.1	1.2	2.8	0.2–14
Sackers	20	9	6.6	3.4	2.9	1.1–45
Other	198	121	11.3	3.2	4.9	0.2–250
Total	530	242	9.8	2.4	4.7	0.2–450

AM = arithmetic mean.
GM = geometric mean.
GSD = geometric standard deviation.
n = number of exposure measurements.
j = number of workers measured at least on one occasion.

autumn). For five facilities, repeated measurements on the same worker were available. The goals of the exposure measurements were to study:

- If and how the population could be divided into homogeneous exposure groups.
- If and how these data could be used to calculate cumulative exposure estimates.
- If and how the measurements in 8 facilities could be used to estimate the exposure of workers in a total of 14 facilities.

The facilities differed in a number of characteristics such as age, number of employees, and technology applied. All workers were *a priori* grouped into eight occupational categories according to their job title: unloaders, crane drivers, operators, press operators, production managers, expedition workers, sackers, and other production workers. The exposure categories consisted of one or more job titles with similar facility tasks and working environments. Personal, breathing zone, inspirable dust sampling was performed on workers from all occupational categories during full-shift periods of approximately 8 hours. The endotoxin concentration was determined in all samples according to methods described in the literature.(16) Lung function of 315 workers from 14 animal feed production mills was measured according to the European Committee of Coal and Steel protocol.(16)

The data were analyzed with SAS statistical software. Analyses of variance were performed using Proc GLM. The within and between worker mean squares were recalculated to within and between worker geometric standard deviations according to the equations in the Appendix. On the basis of the between worker variance, a $R_{0.95}$, as a measure of homogeneity of the group, was calculated.(15,17) The within and between occupational category variances were calculated with a nested analysis of variance (Proc NESTED) in which the occupational group was the nesting variable. The relationship between the current dust exposure, the cumulative dust exposure, and endotoxin exposure and lung function was studied in a multiple regression analysis after a correction for age, standing height, and pack-years smoked (Proc REG).

Results

The results for the dust and endotoxin concentration are presented in Tables I and II. Tabular analysis, as well as the analyses of variance with the dust and endotoxin concentration as dependent variables and occupational category and feed mill as explanatory variables, showed that

TABLE II. Characteristics of Personal Endotoxin Measurements by Job Category

Job Category	n	j	AM (mg/m³)	GM (mg/m³)	GSD	Range (mg/m³)
Unloaders	69	29	102	28.5	5.4	0.2–1150
Crane drivers	23	12	49.6	10.6	8.7	0.2–420
Facility operators	63	20	3.9	1.2	4.3	0.2–50
Press operators	54	17	9.4	2.3	4.2	0.2–160
Production managers	42	16	15.2	2.5	5.0	0.3–160
Expedition workers	61	18	29.7	2.4	5.2	0.2–880
Sackers	20	9	9.3	3.7	3.8	0.5–75
Other	198	121	40.9	7.0	6.0	0.2–1870
Total	530	242	37.1	5.1	6.7	0.2–1870

AM = arithmetic mean.
GM = geometric mean.
GSD = geometric standard deviation.
n = number of exposure measurements.
j = number of workers measured at least on one occassion.

TABLE III. Analysis of Variance of Log-Transformed Personal Dust and Endotoxin Concentration (n = 530)

		Dust		Endotoxin	
Factor	df	MS	F	MS	F
Job category	7	28.2	15.8*	37.2	14.3*
Facility	7	11.6	6.5*	18.8	7.2*
Residual	515	1.8	2.6		

df = Degrees of freedom
MS = Mean squares
F = F-test
* = $p < 0.001$

the occupational category explained most of the variance in dust and endotoxin concentrations (Table III). In the eight mills monitored, this meant that 1) more or less the same pattern of exposure existed for the various occupational categories, 2) differences in mean exposure between mills were small, despite differences in age and technology level, and 3) sampling would not necessarily have to be expanded to all 14 mills. This last point is clearly illustrated by a breakdown of the exposure to occupational groups and to facility. The differences in geometric mean dust and endotoxin exposures between the highest and lowest exposed occupational group were, respectively, a factor 12 and 24, while for the facilities, this was maximally a factor 2 to 4. Because the facilities had a different age and technology level, the results also indicated that the current exposure could probably serve as an acceptable estimate of a previous exposure.

An analysis of variance showed that, for the total population, the between worker variation in exposure, expressed as a geometric standard deviation, was somewhat larger than the within worker variation both for dust and endotoxin ($\lambda < 1$) (Table IV). A similar analysis of variance for each occupational group revealed that the within person variation in exposure (or day-to-day variability) was equal to or considerably greater than the interpersonal variability in exposure for most occupational groups. Crane drivers formed a clear exception for both endotoxin and dust. A more detailed analysis of the exposure of crane drivers showed that some of them had specific tasks of short duration which led to a relatively high exposure and, thus, inhomogeneity within the group of crane drivers. The group was also very small and exclusion of one outlier increased the λ and reduced the $R_{0.95}$ drastically. The occupational category "Other production workers" showed large differences in dust exposure within the group, probably because they were too heterogeneous in personal mean exposures due to shifting of tasks. However, in most cases, the $R_{0.95}$ for each occupational group appeared to be smaller than the $R_{0.95}$ for the total group. This indicated that the differences between workers within an occupational group were smaller than the differences between workers in the whole population. This implies that the occupational group mean exposure was probably an acceptable exposure estimate for each member of the group, except for "crane drivers" (dust and endotoxin) and "other production workers" (dust only). Reduction of the $R_{0.95}$ for most occupational groups compared with the overall $R_{0.95}$ suggests that calculation of a cumulative exposure based on occupational categories could still be useful. This was also confirmed by the nested analysis of variance. This analysis resulted in a ratio of the within and between exposure group variance which was 1.24 for dust and 0.97 for endotoxin. This means that the within exposure group variance was larger than the between exposure group variance for dust exposure. For endotoxin exposure, the within exposure group variance was smaller than the between group variance suggesting that a breakdown to occupational categories was more informative than in case of dust. Therefore, the calculation of the cumulative dust and endotoxin exposures was based on the different occupational categories performed, the arithmetic average dust exposure in that particular occupational category, and the time spent in that occupational category. Differences in mean exposure for workers with the same occupational category but working in different mills were neglected.

Statistically significant relationships were found between

TABLE IV. Within and Between Worker Variation, Expressed as a GSD, λ, and $R_{0.95}$, for Log-Transformed Personal Dust and Endotoxin Exposure per Occupational Category. Only Workers with $k > 1$ Repeated Measurements Have Been Taken into Account

		Dust Exposure				Endotoxin Exposure			
	j^1	GSD_w	GSD_b	λ	$R_{0.95}$	GSD_w	GSD_b	λ	$R_{0.95}$
All workers	120	2.88	3.08	0.89	82	3.88	4.02	0.95	234
Unloaders	15	4.05	1.55	10	6	4.30	2.46	2.6	34
Crane drivers	5	3.07	3.37	0.8	117	4.20	4.75	0.8	450
Facility operators	13	2.39	1.73	2.5	9	3.26	2.11	2.5	19
Press operators	12	2.50	2.12	1.5	20	2.80	2.67	1.1	48
Production managers	10	2.29	2.66	0.7	47	3.22	3.40	0.9	125
Expedition workers	14	2.51	1.30	12.2	3	4.69	1.83	6.6	11
Packers	3	3.16	1.40	11.4	4	4.86	1.01	>100	1.04
Other production	48	3.04	3.13	0.95	88	4.05	3.29	1.4	107

j^1 = number of persons measured more than one occasion ($k > 1$).
GSD_w = within worker geometric standard deviation.
GSD_b = between worker geometric standard deviation.
λ = variance ratio of within and between worker variance.
$R_{0.95}$ = ratio of 5 and 95% confidence interval of between worker variance.

TABLE V. Change in Lung Function Per Unit Cumulative Dust and Endotoxin Exposure Among Animal Feed Production Workers Adjusted for Age, Standing Height, and Pack-Years Smoked with Linear Regression (n = 263)

	Cumulative Dust Exposure		Cumulative Endoxin Exposure	
	β_{dust}	SE	$\beta_{endotoxin}$	SE
FVC (L)	−0.0011	0.0014	−0.0003	0.0006
FEV_1 (L)	−0.0021	0.0013	−0.0009	0.0006
FEV_3 (L)	−0.0012	0.0014	−0.0004	0.0006
PEF (L/s)	−0.0206[A]	0.0049	−0.0084[A]	0.0021
MEF_{75} (L/s)	−0.0141[A]	0.0045	−0.0067[A]	0.0019
MEF_{50} (L/s)	−0.0059	0.0032	−0.0032[B]	0.0014
MEF_{25} (L/s)	0.0005	0.0015	−0.0003	0.0006
MMEF (L/s)	−0.0028	0.0026	−0.0019	0.0011

[A]$p < 0.01$.
[B]$p < 0.05$.
β = regression coefficient.
SE = standard error.

most of the lung function variables and the cumulative dust, as well as endotoxin exposure, after correction for the confounding variables. Such relationships were also found for the current exposure to dust and endotoxin. The results of the analysis with cumulative exposure are given in Table V. The relationship between the endotoxin exposure and lung function appeared to be a statistically stronger association than that between dust exposure and lung function.

Discussion and Conclusions

Large variations in exposure in time and between workers are an important source of difficulties in the design and analysis of exposure–response relationships because these variations can be responsible for considerable misclassification. Theory showed that misclassification of exposure can result in underestimation of the relationship between an exposure and some health endpoint. The magnitude of this bias is determined by the ratio of the within and between worker variance in exposure in a study in which all workers have been monitored. The magnitude of bias can be calculated on the basis of the variance ratio of within and between worker variance in exposure. Several strategies are recommended to reduce this bias.

For grouping strategies, no strict criteria can be given for a required level of homogeneity. The inhomogeneity which is acceptable in a specific situation is an analogy determined by the magnitude of the variation in exposure in the whole population and the variation within the exposure groups. It is, therefore, not necessary to define absolute criteria for homogeneity. Criteria need to be developed for the acceptable ratio of the within and between group variances.

Some elements of the theory were illustrated with a study in the Dutch animal food industry. Clear differences in exposure existed between some of the *a priori*-defined occupational categories. These differences were such that differences within an occupational category due to different mills explained only a limited part of the total variance in the exposure. However, the analyses of the exposure data were performed before the lung function results were analyzed. The initial analysis of exposure data was meant to decide on further steps to be taken. It was planned that additional measurements would be taken if considerable differences between mills would exist or differences within occupational groups would exist. Cumulative exposure could have been calculated in a different way, based on the occupational categories performed in a certain facility where the tasks were performed. For the subsample of eight facilities, cumulative exposure was calculated in this alternative way. Similar exposure–response relationships were found suggesting that the assumptions made were acceptable.

The between worker variance for the subgroups was reduced compared to the between worker variance for the whole group. This showed that the grouping lead to a reduction of exposure variability. The fact that clear exposure–response relationships were found of the dust and endotoxin exposure with lung function suggests that an acceptable ratio between these two variance components was present.

It is recommended to establish the within and between group variance in exposure with an analysis of variance using repeated measurements on the same individual. The success of different groupings can be judged by comparing ratios of within and between group variances of the alternative groupings. The optimal grouping of the population can then be selected for the further epidemiologic analysis of the data.

Acknowledgments

The animal feed production industry study was granted by the Dutch Directorate General of Labor of the Ministry of Social Affairs and Employment. This analysis was conducted in the framework of the EC Concerted Action "Retrospective Evaluation of Occupational Exposures in Cancer Epidemiology."

References

1. Vihma T.: Health Hazards and Stress Factors in Small Industry. Scand. J. Work Environ. Health 7(suppl. 3) (1981).
2. Checkoway, H.; Dement, J.M.; Fowler, D.P.; et al.: Industrial Hygiene Involvement in Occupational Epidemiology. Am. Ind. Hyg. Assoc. J. 48:515–523 (1987).
3. Kromhout, H.; Oostendorp, Y.; Heederik, D.; Boleij, J.S.M.: Agreement Between Qualitative Exposure Estimates and Quantitative Exposure Measurements. Am. J. Ind. Med. 12:551–562 (1987).
4. Oldham, P.D.; Roach, S.A.: A Sampling Procedure for Measuring Industrial Dust Exposure. Br. J. Ind. Med. 9:112–119 (1952).
5. Ashford, J.R.: The Design of a Long-term Sampling Program to Measure the Hazard Associated with an Industrial Environment. R. Stat. Soc. Series A, 121:333–347 (1958).
6. Rappaport, S.M.: Smoothing of Exposure Variability at the Receptor: Implications for Health Standards. Ann. Occup. Hyg. 29:201–214 (1985).
7. Rothman, K.J.: Modern Epidemiology. Little, Brown and Company, Boston (1986).
8. Armstrong, B.G.; Oakes, D.: Effects of Approximation in Exposure Assessment on Estimates of Exposure–Response Relationships. Scand. J. Work Environ. Health 1:20–23 (1982).

9. Ulvarsson, U.: Limitations to the Use of Employee Exposure Data on Air Contaminants in Epidemiologic Studies. Int. Arch. Environ. Health 52:285–300 (1983).
10. Cochran, W.G.: Errors of Measurement in Statistics. Technometrics 10:637–666 (1968).
11. Snedecor, G.W.; Cochran, W.G.: Statistical Methods, 7th ed. State University Press, Iowa (1980).
12. Searle, S.R.: Linear Models. John Wiley & Sons, New York (1971).
13. Liu, K.; Stamler, J.A.; 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. Chron. Dis. 31:399–418 (1978).
14. Heederik, D.J.J.; Miller, B.G.: Weak Associations in Epidemiology: Adjustment for Exposure Estimation Error. Int. J. Epidemiol. 17:970–974 (1988).
15. Rappaport, S.M.: Exposure Assessment Strategies. In: Exposure Assessment for Epidemiology and Hazard Control, pp. 219–250. Lewis Publications, Chelsea, MI (1991).
16. Kateman, E.; Heederik, D.; Pal, T.M.; et al.: Relationship of Airborne Microorganisms with the Lung Function and Leukocyte Levels of Workers with a History of Humidifier Fever. Scand. J. Work Environ. Health 16:428–433 (1990).
17. Rappaport, S.M.: Assessment of Long-term Exposures to Toxic Substances in Air. Ann. Occup. Hyg. (accepted for publication, 1991).

APPENDIX

An analysis of variance of between and within worker variance components using repeated measurements on the same worker can be performed with various procedures in SAS such as Proc GLM, Proc NESTED, and Proc VARCOMP. The results are presented schematically in Table A. The between and within group variances can be calculated in a similar way with group and worker as an explanatory variable in the model depending on the procedure used.

Several variables can be calculated from the outcomes of an analysis of variance:

$$GSD_w = \exp(\sigma_w^2)^{1/2} = \exp(MS_w)^{1/2}$$

$$GSD_b = \exp(\sigma_b^2)^{1/2} = \exp[(MS_b - MS_w)/k]^{1/2}$$

$$R_{0.95} = \exp(\sigma_b \times 3.92)$$

if the number of repeated measurements differs from worker to worker k is approximated by:

$$k' = \frac{t-(k_1^2 + k_2^2 + \ldots + k_n^2)/t}{n-1}$$

where: t = total number of observations
k_1 = number of observations on day 1
n = maximum number of monitoring days

or can be approximated by:

$$k'' = \left(\frac{df_{\text{within persons}}}{df_{\text{between persons}} + 1}\right) + 1$$

TABLE A. Analysis of Variance (ANOVA) of Repeated Exposure Measurements on the Same Worker[12]

Variation Source	SS	df	MS	Estimated Parameter
Between worker	SS_b	$(n-1)$	$SS_b/(n-1)$	$\sigma_w^2 + k\sigma_b^2$
Within worker	SS_w	$n(k-1)$	$SS_w/n(k-1)$	σ_w^2

SS = sum of squares.
MS = mean of squares.
df = degrees of freedom.
σ_w^2 = within subject variance.
σ_b^2 = between subject variance.
k = number of repeated measurements per worker.