CHAPTER 3:
APPLICATION OF MIXED-EFFECTS MODELS FOR EXPOSURE ASSESSMENT

C. Peretz , A. Goren , T. Smid , H. Kromhout

Ann Occup Hyg 2002;46:69-77.

ABSTRACT

The benefits of using linear mixed effects models for occupational exposure assessment were studied by re-analysing three data sets from two published surveys with repeated exposure measurements. The relative contributions of particular characteristics affecting exposure levels were assessed as in a multiple regression model, while controlling for the correlation between repeated measurements. While one-way ANOVA allows one only to estimate unconditioned variance components, a mixed model enables estimation of between- and within-worker variance components of exposure levels while accounting for the fixed effects of work characteristics. Consequently, we can identify the work characteristics affecting each variance component. Mixed models were applied to the data sets with repeated measurements and auxiliary information on work characteristics. The between-worker variance components were reduced by 35%, 66% and 80% respectively in the three data sets when work characteristics were taken into account. The within-worker (day-to-day) variability was reduced only in the pig farmer data set, by 25%, when accounting for work activities. In addition, coefficients of work characteristics from the mixed model were compared with coefficients resulting from originally published multiple linear regression models. In the rubber manufacturing data, the coefficients of the mixed model showed similar relative importance, but were generally smaller than the coefficients from regression models. However, in the pig-farm data, only the coefficients of work activities were somewhat reduced. The mixed model is a helpful tool for estimating factors affecting exposure and suitable variance components. Identifying the factors in the

working environment that affect the between-worker variability facilitates *a posteriori* grouping of workers into more uniformly exposed groups. Identifying the factors that affect the within-worker variance is helpful for hazard control and in designing efficient sampling schemes with reference to time schedule.

INTRODUCTION

Ideally, exposure assessment of air pollutants at the workplace should be based on repeated measurements on randomly selected days of a randomly selected number of workers from *a priori* defined occupational groups.[1-5] Usually measurements done on the same worker are correlated. Since exposure varies both within and between workers in a given exposure group,[1,2,4,6,7] these variance components should be taken into account in exposure assessment and for more effective hazard-control, as well as in compliance testing and evaluation of exposure-response relationships.[8-11]

Understanding the factors in the work environment that affect mean exposure levels enables the estimation of the between- and within-worker variance components conditioned on these factors. The identification of uniformly exposed groups of workers is essential for valid compliance testing and exposure-response evaluation. Identification of the factors in the work environment that are related to the between-worker variance component enables sub-grouping of workers into more uniformly exposed groups. An understanding of the factors affecting within-worker variance assists in the identification of conditions in the work environment that cause varying concentrations from day to day. This is a prerequisite for better protection of the individual worker from hazardous exposures.

So far, in studies with repeated measurements designs, most researchers have used either a) a one-way random-effects model to estimate variance components, ignoring work characteristics[12-19] and/or b) multiple linear regression to model the effect of work characteristics on observed exposure levels, ignoring the correlation between repeated observations from the same worker.[14-15] Mixed-effects models for unbalanced data simultaneously estimate both the effects and the variance components in a more efficient way.[20-22] For exposure groups those models can describe the influence of fixed and random work environment characteristics on the observed exposure levels, and estimate the within- and between-worker variance components controlled for work characteristics and other determinants of exposure. Recently, mixed effects models were used by several researchers for different purposes.[19, 23-25] A time trend can be introduced into these models as a fixed effect. In this paper we present the benefits of using mixed effects models for unbalanced data to estimate variance components while controlling for work characteristics. In addition, we present the coefficients of the work characteristics that affect exposure levels, controlled for the correlation between repeated measurements. Finally, we will show which work characteristics affect between- and within-worker variability in exposure concentrations.

For our analysis, we used three existing data sets with repeated personal exposure measurements as described previously in two published papers.[14,15,18] A common feature of the data sets was that they were from systematic surveys. Auxiliary information on the work environment and activities was collected during the measurements. The data sets stemmed from two industry-wide surveys among workers from the rubber manufacturing industry and one survey among pig farmers in the Netherlands. Detailed information on these studies and the results from the one-way random-effects models and from the multiple linear regression models can be found in the above-mentioned papers.[14,15,18]

METHODS

Study design, data collection and previous statistical analysis

First Example: Industry-wide survey of the rubber manufacturing industry

This study of the rubber manufacturing industry was performed in the Netherlands, to examine relationships between working conditions and chemical exposures. Personal exposures to airborne particulate, rubber fumes and solvents, as well as dermal contaminants, were measured in a representative sample of 10 factories producing an array of different rubber products. For each plant, the measurements and observations took four days (Tuesday-Friday). Auxiliary data on tasks performed, use of personal protection devices, ventilation characteristics and process characteristics were collected through interviews of sampled workers. Workers were selected, stratified by production function and by the job done, and surveyed on randomly chosen days during the course of the four-day measurement period.

Multiple regression models were applied to evaluate the relationships between the collected auxiliary data and exposure levels , for two groups of workers: 234 workers with 620 measurements exposed to inhalable particulate, and a sub-group of 36 workers with 59 measurements exposed to rubber fumes (measured as the cyclohexane-soluble fraction of the inhalable particulate). Details of the study and the modeling can be found elsewhere.[18]

Second Example: Survey on Pig Farmers' Exposure to Inhalable Endotoxin

In a study among 98 pig farmers from the south of the Netherlands, exposure to inhalable dust and endotoxin was monitored by personal sampling. Exposure was measured during one work shift on a randomly chosen day of the week; one day during the summer of 1991 and one day during the winter of 1992. Outdoor temperature was obtained from a monitoring station in the south of the Netherlands. Task activity patterns and farm characteristics were also recorded. Activities, which were represented by time spent in each activity, were based on daily averages during 12-14 days For the purpose of this paper, only the exposure data on endotoxin will be used. Multiple linear regression analysis were applied to evaluate the relationship between farm characteristics, activities and outdoor temperature and log-transformed endotoxin concentrations. One-way random-effects model was applied to estimate variance components. Details of the study and the modeling can be found elsewhere.[14,15]

Sources of exposure variability

We postulated that the variability of the exposure levels in an industrial hygiene group of workers arises from 4 sources:

1. Systematic between worker variation: Systematic differences in factors that define the work conditions of different workers. These factors are mostly spatially related (e.g. local ventilation), varying among workers but constant in time for each worker. Sometimes these factors are both temporal and spatial (e.g. process temperature), meaning that the levels differ among workers (the mean value) and within the same worker along time.

2. Random between worker variation: Differences among workers beyond what can be explained by specific factors. This additional variation may be associated with factors that are not measured due to time/money limitations, inability to measure (e.g. workers' habits) or lack of awareness.
3. Systematic within worker variation: Systematic differences in factors that define the work conditions of the same worker over time. These factors are temporal (e.g. burden, activities) and may be common to all workers in the IH group. Time itself is one possible systematic, within worker, factor (e.g. season, year). Usually these within worker changes are related to the cycle of work, the production, seasonality etc.
4. Random within worker variation: Differences among measurements on the same worker at different time points beyond what can be explained by specified factors. This additional variation may be associated with further within-worker factors (e.g. changes of habits of a worker) that are not measured due to time/money limitations, inability to measure or lack of awareness (e.g. measures taken by different hygienists, measurement errors).

Thus, within the same exposure group along time, the usual partition of the total exposure variance into 2 components: between workers and within workers[1,2,4,6,7] can be refined when work characteristics are taken into account. 1. Systematic variation (sources 1,3) accounts for differences in work characteristics. These differences can be included as explanatory variables in the model whose effects can be estimated. 2. Random variation (sources 2,4), which is partitued into 2 components: a. Between worker variation, "conditioned" on the effects of the observed work characteristics (source 2). This variance component reflects additional variation among workers beyond differences due to work factors. b. Within worker variation (source 4), which reflects additional variation at the within-worker level. The total random variance when work characteristics are taken into account is a conditional variance, so its value is less than the total variance when those factors are not taken into account, and is the sum of the conditioned random between worker variance and the random within worker variance.

Mixed effects models

A mixed effects model is a generalisation of the standard linear model (a multiple regression model) that enables the analysis of data generated from several sources of variation instead of just one.[26] It associates one continuous dependent variable (a response, an outcome,) with several explanatory variables (categorical or continuous). The unique aspect of the mixed effects model is the inclusion of both fixed and random factors. Fixed effects provide estimates of the average responses in the group, like in a common regression model, while random effects (e.g. subjects' effects) account for the natural heterogeneity in the responses of different individuals and allow estimation of responses for each individual in the study. Since measurements done on the same subject are correlated, this correlation must be taken into account in modelling. The dependence among the repeated responses can be of different types leading to specific covariance/correlation structures. The model allows the assumption of several covariance structures and enables estimation of the effects as well as variance parameters. The number of observations per subject can be either the same (a balanced design) or different (an unbalanced design). The time points can be either identical across subjects or not. The time-interval between repeated observations can vary across repetitions.[21]

Current application of the mixed effects model

The application of this model for identifying the determinants of exposure and assessing variance components is presented in this paper using two examples: one from the rubber manufacturing industry, and another from pig farming. In both examples, the dependent variable was the exposure level of a pollutant and the explanatory variables were workplace characteristics. These fixed effects were either time-dependent (e.g. outdoor temperature) or fixed along time (e.g. stable flooring). The individual worker's effect was taken as a random effect. We looked at two sources of random variance: the random variance between workers and the random variance among repeated measurements within workers. In the first example we had three repetitions, and in the second example we had only two repetitions per worker. We assumed that any two repeated measurements of the same worker have equal correlation irrespective of the time interval between them; a compound symmetry covariance structure. This is the covariance structure assumed in classical repeated measurements ANOVA. Furthermore, we assumed that the variance between workers is equal across the fixed factors of work characteristics (defined as homogeneity) and that the effects as well as the variance within workers are equal across all workers in the same group. The residuals were assumed to be independent of each other. Exposure concentrations were assumed to be log-normally distributed.[1,2,4,7] Mixed models were applied with and without the fixed effects (without- is equivalent to the random effects model), to determine the impact of the fixed effects upon the variance component associated with the random effects. The estimated variance components of both models were compared.

PROC MIXED from SAS System Software Version 6.1[26] was used for the analysis. The procedure enables simultaneous estimation of the effects, their standard error and significance tests, as well as variance components and their confidence limits. Variance components were estimated using the Restricted Maximum Likelihood (REML) method. Nested models were compared by likelihood ratio test.

The mixed-effects model for unbalanced data is specified by the following expressions:

$$y_{ij} = \beta_0 + \beta_1 X_{ij1} + \ldots + \beta_p X_{ijp} + b_1 z_1 + \ldots b_k z_k + \varepsilon_{ij}$$

For i=1,...,k (workers) and j = 1,...,n_i (repetitions of the i'th worker)

Where:

Y_{ij} = log-transformed exposure level

β_0 = an overall intercept for the group that corresponds to mean background exposure (log-transformed) when all factors equal zero

$\beta_1, \ldots, \beta_p$ = fixed effects

$x_{ij1}, \ldots, x_{ijp}$ = values of the variables for the i'th worker on the j'th day

$b_1, \ldots, b_k$ = workers' random effects

b_i = i'th worker random effect, which corresponds to the discrepancy between his intercept and the group intercept β_0

$Z_1, \ldots, Z_k$ = workers' indicators (0/1)

It is furthermore assumed that:

$b_i \sim$ N(0, σ_b^2), b_i 's are all independent ; $\varepsilon_{ij} \sim$ N$(0, \sigma_w^2)$, ε_{ij} 's are all independent,

b_i 's and ε_{ij} 's are all independent of each other

$\sigma^2 = \sigma_b^2 + \sigma_w^2$; $\sigma_b^2 =$ variance between-workers; $\sigma_w^2 =$ variance within-workers.

$\rho = \sigma_b^2 / \sigma^2$; $\rho =$ correlation between any two repeated measures of the same worker

$$corr(y_{ij}, y_{il}) = \begin{cases} 1 & j = l \\ \rho & j \neq l \end{cases} \qquad cov(y_{ij}, y_{il}) = \begin{cases} \sigma^2 & j = l \\ \rho\sigma^2 & j \neq l \end{cases}$$ compound symmetry structure

For every i , i= 1,...,k and for every j,l , j,l = 1,...,n_i

(All y_{ij} of the same worker are correlated, and those of different workers are uncorrelated)

To model the influence of work characteristics on the exposure levels they were considered as fixed effects in the above model, for example:

β_1 = process pressure effect

β_2 = process temperature effect

β_3 = local exhaust ventilation effect

x_{ij1} = process pressure value for the *i*-th worker on the *j*-th day

x_{ij2} = process temperature value for the *i*-th worker on the *j*-th day

x_{ij3} = local exhaust ventilation indicator (0-non present/1-present) for the *i*-th worker on the *j*-th day

To identify time trends in the above model a time term was added to the model:

β_4 = period effect

x_{ij4} = period indicator (0-first period /1-second period), for the *i*-th worker on the *j*-th day

RESULTS

First Example: Industry-wide survey of the rubber manufacturing industry

Results of the application of both the random-effects model (the model without the fixed effects) and the mixed-effects model (the model with the fixed effects) were compared in the 2 data sets with exposure to inhalable particulate and rubber fumes (see Table 3.1, 3.2). The effect of the day of the week on the exposure level was tested separately. Table 3.1 shows that for workers exposed to inhalable particulate, the factors affected the between-worker component of variance (s^2_{bw}) considerably (35% reduction from 1.30 to 0.84), but did not alter the within-worker component of variance (s^2_{ww}). In the two models s^2_{bw} is different from zero ($p<.05$) and the models were significantly different ($p=.0023$).

Table 3.1 Variance components estimates for the one-way random-effects and mixed-effects models for exposure to inhalable particulate among workers in the rubber manufacturing industry (n=620, k=234, l=10)

Variance	Random-effects model est (CI)	Mixed-effects model est (CI)
s^2_{bw}	1.30 (1.07-1.59)	0.84 (.68-1.06)
s^2_{ww}	0.29 (.25-.34)	0.30 (.26-.34)

NOTES: n = no. of measurements; k = no. of workers; l = no. of factories;
s^2_{bw} = between-worker variance component; s^2_{ww} = within-worker variance component;
est= estimator; CI= 95% Confidence Interval

Table 3.2 Variance components estimates for the one-way random-effects and mixed-effects models for exposure to rubber fumes among workers in the rubber manufacturing industry (n=59, k=36, l=7)

Variance	Random-effects model est (CI)	Mixed-effects model Full model est (CI)
s^2_{bw}	0.53 (.31-1.17)	0.18 (.08-.70)
s^2_{ww}	0.32 (.20-.60)	0.30 (.19-.55)
	Mixed-effects model including local ventilation	Mixed-effects model including process' temperature and pressure
s^2_{bw}	0.49 (.28-1.08)	0.26 (.13-.83)
s^2_{ww}	0.31 (.20-.58)	0.32 (.20-.59)

NOTES: n = no. of measurements; k = no. of workers; l = no. of factories;
s^2_{bw} = between-worker variance component; s^2_{ww} = within-worker variance component;
est= estimator; CI= 95% Confidence Interval

Table 3.3 Coefficients for binary (0/1)[a] factors affecting exposure to inhalable particulate among workers in the rubber manufacturing industry in the mixed effects model and multiple linear regression model (n=620, k=234, l=10)

	Mixed-effects model		Multiple linear regression	
	β	P	β	p
Intercept	-0.05		-0.17	
Punching powdered products	3.41	<.01	3.79	<.01
Tube inspection	3.47	<.01	3.34	<.01
Packing powdered products	2.67	<.01	2.82	<.01
Jointing	2.41	<.01	2.71	<.01
utoclave	0.63	.01	1.66	<.01
Weighing	1.26	<.01	1.43	<.01
Heating mill	0.26	ns	1.10	<.01
Repair buffing	0.71	.02	0.82	<.01
Bench fitting	0.51	.01	0.78	<.01
Open mill	0.20	ns	0.61	.03
Internal mill	0.31	ns	0.52	.03
Packing	0.51	.01	0.48	.02
Cleaning	0.15	ns	0.37	<.01
Transport	0.15	ns	0.34	.02
Rubber cutting	-0.18	ns	-0.24	bs
Inspection	-0.29	ns	-0.29	ns
Breakdown work	-0.51	<.01	-0.38	.05
Punching	-0.20	ns	-0.46	ns
Unrolling	-0.49	bs	-0.60	.05
Calendering	-0.13	ns	-0.64	<.01
Assembling machine	-0.69	ns	-0.68	bs
Manual assembling	-0.41	.04	-0.73	<.01
Loading-unloading	-0.43	ns	-0.80	.05
Weighing products	-0.71	.04	-0.93	.02
Extruding-slicing	0.05	ns	-0.97	.03
Lead extrusion	-0.24	ns	-1.00	.05
UHF curing	-0.48	ns	-1.05	<.01
Braiding machine	-1.22	bs	-1.21	.02
Laboratory work	-0.87	.03	-1.28	<.01
Autoclave LEV	-0.69	.03	-1.39	<.01
General trimming	-1.56	.03	-1.94	<.01

NOTES: n = no. of measurements; k = no. of workers; l = no. of factories; bs .05<p≤.10; ns p>.10
[a] binary variable: 0= non-present, 1=present

The difference between the models suggests no systematic changes in work tasks and production characteristics for individual workers during a week. The main difference between the mixed-effects model and the original multiple regression model[18] assuming independence between repeated measurements, was that fewer exposure-affecting factors were statistically significant or borderline statistically significant ($p<0.10$) (17 versus 28 factors, see Table 3.3). However, the coefficients had the same sign for all factors except one task ("Extruding-Slicing"), whose coefficient was nearly 0 in the mixed-effects model. The tasks "Punching Powdered Products", "Packing Powdered Products", "Tube Inspection" and "Jointing" affected exposure to inhalable particulate most dramatically, according to both the mixed-effects model and the original multiple regression model.

For workers exposed to rubber fumes (Table 3.2), the same phenomenon for the components of variance was observed. The three factors, one pure spatial between workers factor (local ventilation) and the rest (process- temperature and pressure) both spatial and temporal factors, affected the between-worker component of variance (66% reduction from 0.53 to 0.18). The within-worker component of variance (s^2_{ww}) was not affected. In the 2 models s^2_{bw} is different from zero ($p<.05$) and the models were not significantly different. The coefficients of the factors were almost identical when compared to the original multiple linear regression (Table 3.4).

Table 3.4 Coefficients for factors affecting exposure to rubber fumes among workers in the rubber manufacturing industry in the mixed-effects model and multiple linear regression model ($n=59$, $k=36$, $l=7$)

	Mixed-effects model		Multiple linear regression	
	β	p	β	p
Intercept	4.94		4.97	
Process-Temperature[a]	0.0056	.04	0.0049	.05
Process-Pressure[b]	0.0038	<.01	0.0042	<.01
Local Exhaust Ventilation[c]	-0.69	<.01	-0.68	<.01

NOTES: n = no. of measurements; k = no. of workers; l = no. of factories

[a] per 1 °C;

[b] per 1 bar;

[c] binary variable: 0=non-present,1=present

The effect of 'day of the week' (Tuesday-Friday) on the mean exposure and on the components of variance was non-significant for both exposure modelling (to inhalable particulate and rubber fumes), suggesting no systematic differences in exposure over the course of a week.

Second Example: Survey on Pig Farmers' Exposure to Inhalable Endotoxin

Table 3.5 presents the results of both the original one-way random-effects model and a mixed-effects model with 21 fixed effects for both farm characteristics and activities. Furthermore, results from three additional mixed models with only farm characteristics, only activities and outdoor temperature, and only a season term, respectively were elaborated and are presented in the same table.

From the random effects model (Table 3.5) it was clear that the within-worker variance component had a greater weight than the between-workers variance component (85% versus 15% of the total exposure variance). The between-worker variance component was low 0.11. Table 3.5 shows the extent of reduction in both the within- and between-worker variance components by including all the statistically significant factors from the original multiple regression model. The within-worker variance component was reduced by 0.16 (25%) and the 2 models are significantly different ($p<.0001$). The between-worker variance component was reduced by 0.09 (80%) and while in the random effects model s^2_{bw} is different from zero ($p<.05$) in the mixed model with the farm characteristics $s^2_{bw} \sim 0$.

In Table 3.5 we see that farm characteristics appeared to be solely responsible for the reduction in the between-worker variance component ($s^2_{bw}=0.01$). The model with outdoor temperature and farmers' activities pure within worker factors had no effect on the between-worker variance component ($s^2_{bw}=0.12$), while it reduced the within-worker variance component by 25%.

Table 3.5 Variance components estimates for the one-way random-effects and mixed-effects models (full model and three sub-hierarchical models) for exposure to endotoxins among pig farmers (n=348 , k=198)

Variance	Random-effects model est (CI)	Mixed-effects model Full model est (CI)
s^2_{bw}	0.11 (.05-.45)	0.02 ([a])
s^2_{ww}	0.64 (.52-.80)	0.48 (.04-.60)

Variance	Mixed-effects model including farm characteristics est (CI)	Mixed-effects model including activities and temperature est (CI)	Mixed-effects model including season effect est (CI)
s^2_{bw}	0.01 ([a])	0.12 (.06-.33)	0.14 (.07-.39)
s^2_{ww}	0.64 (.52-.80)	0.48 (.39-.61)	0.59 (.48-.74)

NOTES: n = no. of measurements; k = no. of workers; l = no. of factories
s^2_{bw} = between-worker variance component; s^2_{ww} = within-worker variance component;
est= estimator; CI= 95% Confidence Interval ;
[a] cannot be computed

This clearly shows that different work environment characteristics contributed independently to the variance components. Farm characteristics, which are almost constant over the time period studied (one year), were responsible for differences in average endotoxins concentration between farmers, while changes from day to day in temperature and activities performed, led to temporal variability in exposure concentrations. The season factor was found to have a minor influence on the within- worker variance component.

The two models (the random effects model and the mixed effects model with season effect) were found to be significantly different ($p<.0001$). When compared to the coefficients of the original multiple linear regression model[14] the estimated coefficients from the Unbalanced Mixed Effects Model were almost exactly the same for the farm characteristics, but somewhat smaller for the activities (Table 3.6). Neither the relative order nor the p-values changed.

Table 3.6 Coefficients for factors affecting exposure to endotoxines among pig farmers, in the mixed-effects model and multiple linear regression model (n=348, k=198)

	Mixed-effects model		Multiple linear regression	
	β	p	β	p
Intercept	4.44		4.44	
Temperature				
Outdoor temperature (per 10%)[a]	-0.35	<.01	-0.35	<.01
Farm characteristics				
Feeding				
manual dosage dry feeding (1/0)[b]	-0.38	<.01	-0.37	<.01
pig starter (1/0)	0.35	.03	0.35	.03
automated dry feeding (per 10%)	-0.06	.02	-0.06	.02
Flooring				
Convex floor (1/0)	-0.22	.02	-0.22	.01
Fully slatted floor (per 10%)	0.08	<.01	0.08	<.01
Fully slatted floor with piglet mat (per 10%)	-0.08	<.01	-0.08	<.01
Synthetic grid (per 10%)	-0.13	.04	-0.13	.03
Concrete and metal grid (per 10%)	-0.15	<.01	-0.15	<.01
Floor heating (per 10%)	0.07	<.01	0.07	<.01
Floor heating with delta heating tubes (per 10%)	0.10	.04	0.10	.03
Other				
Overall very dusty	0.12	.02	0.12	.01
Air exhaust via pit (1/0)	-0.33	<.01	-0.33	<.01
Activities (per 10 minutes)[c]				
Feeding	0.04	<.01	0.03	<.01
Controlling	0.03	.03	0.02	.02
Re-penning	0.04	.05	0.02	.05
Floor sweeping	0.08	<.01	0.06	.01
Iron injection	0.09	.03	0.09	.03
Castrating	0.07	.03	0.05	.03
Teeth cutting	0.25	<.01	0.17	.01
Ear tagging	0.22	<.01	0.14	<.01

NOTES: n = no. of measurements; k = no. of workers

[a] per 10% of total time spent in pig farming;

[b] binary variable: 0= non-present, 1=present

[c] per 10 minutes spent on a task

DISCUSSION

The above examples illustrate the major contribution of the mixed-effects model in unbalanced designs to the investigation of exposure variance components and exposure affecting factors. In contrast with the one-way random-effects model, the mixed effects model deals with both fixed and random effects. It estimates the between- and within-worker variance while adjusting for fixed effects. Simultaneously, it assesses the linear relationships between the determinants of exposure (usually fixed factors) and exposure levels. Common multiple linear regression can be correctly applied when each worker has only a single measurement.

With repeated measurements of each worker, some dependence amongst repeated measurements will exist and the correlation between values for a given person must be taken into account when estimating the relationship between the determinants and exposure levels.[21,22] The mixed effects model is capable of taking this dependence into account in the modeling process.

In this study we used mixed effects models to understand the relationship between specific work environment characteristics and between- and within-worker exposure variance components. Identification of specific work characteristics, which affect the between-worker variance component, will enable development of criteria for defining uniformly exposed groups of workers.[1,5,10,19] Grouping workers into sub-groups is an inherent part of the work of an industrial hygienist in exposure surveys, as well as in compliance tests and epidemiological studies.[1,5,6] Despite the widespread use of grouping strategies, there is so far only limited experience with optimisation of these strategies.[5,27,28]

The analyses with the mixed effects models form the basis for the creation of uniformly exposed groups of workers. Reliance on observational factors such as a job title, which may lead to non-uniformly exposed groups, seems no longer necessary.[5] For instance, classifying curing workers in rubber manufacturing based on the determinant process-temperature and -pressure will lead to more uniform and distinctly different rubber fume exposure groups. It is further recommended that if the IH group cannot be split into sub-homogeneous groups due to a small number of workers, the testing of overexposure in IH groups, as suggested by Lyles et al,[11] which relate to both within and between variance components, should be refined to account for the effects of work characteristics in these groups. As we have shown here, these effects can be estimated with mixed effects models. Also, in this study, specific factors were identified which mainly influenced variability in exposure levels from day to day (within-worker). Hazard control should focus on these factors.

The two examples presented in this paper describe different work situations and measurement strategies. In the first example, one to three randomly chosen measurements were performed within a week on two groups of rubber workers in The Netherlands.[18] In the second example, one to two measurements were collected in two different seasons of a particular year among Dutch pig farmers.[14,15] Different work characteristics were documented in each examples. In the first example, for exposure to particulate, a reduction of 25% of the between-worker variance component was the effect of 17 factors affecting exposure. Classification of rubber workers into uniformly exposed groups will have to rely on those identified factors. In the second example, exposure to endotoxin among pig-farmers, the reduction of 82% in between-worker variance was mainly an effect of the inclusion of 12 farm characteristics. Consequently, classification of pig farmers into uniformly exposed groups will have to rely on those farm characteristics.

In the second example (pig farmers), there was a clear distinction between characteristics influencing each of the variance components. Eight time-dependent activities, as well as outdoor temperature, all pure within worker factors, reduced the within person (day-to-day) variability by 25%. From diary information collected among the pig farmers, it appeared that some of the activities followed a distinct temporal pattern, with some activities taking place only on particular days of the week.14 The farm characteristics, pure between worker factors, were responsible for reducing the between-worker variance component to zero.

In the mixed model, when the between-worker variance component is close to zero, the coefficients for the fixed effects would be very similar to those from a multiple regression model, assuming independence between repeated observations. In the model for the pig farmers, this was indeed the case (between-worker component 0.02). For the rubber workers' exposure to inhalable particulate the opposite was true a very large between-worker variance component, led to changes in the coefficients. In addition, fewer statistically significant exposure-affecting factors were found in the mixed model, since the significance tests of the coefficients in the mixed model are different.

In all the examples, we assumed that any two repeated measurements of the same worker had equal correlation, irrespective of the time interval between them; a compound symmetry covariance structure. This, since we had only two repetitions per worker in the second example and three in the first example. We also assumed independence across subjects. In general, the residuals can be analysed to check for departures from the independence assumption. However, such analyses will be relevant only when there are a reasonable number of observations within workers, say at least 8-10. Other models with different dependence structure may be applied when the time course of the exposure level of each person is of primary interest, that is, when the correlation itself has scientific relevance. The unbalanced mixed-effects model is a specific case of the generalised linear models. These models were developed over the last decade and were recently introduced in common statistical packages such as SAS (Proc Mixed), BMDP (5V) and S-Plus.[26,29,30] The computerised procedures enable simultaneous estimation of parameters, approximate standard errors and significance tests that have not been available before. One should take into account the fact that in models for unbalanced data, the estimators are proxies since the Maximum Likelihood Estimates have approximately normal distribution. Balanced data (the same number of repeated measurements for each worker) is preferable whenever possible, in order to obtain more accurate estimators.

However, the mixed effects model enables estimation of variance components of exposure levels that have been adjusted for workplace factors in order to improve the assessment of exposure. This statistical method can be used to improve future sampling strategies through the grouping of workers into more uniformly exposed groups, and the identification of specific workplace conditions that should be controlled.

ACKNOWLEDGMENTS

The authors are indebted to Dick Heederik for suggesting the idea for this paper. We also would like to acknowledge Liesbeth Preller for providing the pig farmers' exposure data sets, Prof. D. Steinberg for his statistical review, P. Goldberg and Prof. S.M. Rappaport for their helpful comments and to the reviewers for improving the paper.

REFERENCES

1. Rappaport SM. Assessment of Long-Term Exposure To Toxic Substances In Air. Ann Occup Hyg 1991; 35, 61-121.
2. Rappaport SM, Smith TH. Exposure Assessment for Epidemiology and Hazard Control. ACGIH, Michigan, Lewis Publishers Inc..1991.
3. Burdorf, A. Shifting Concepts in Assessment of Occupational Exposures. Ann Occup Hyg 1993; 37, 447-450.
4. Kromhout H, Symanski E, Rappaport SM. A Comprehensive Evaluation of Within- and Between-Worker Components of Occupational Exposure to Chemical Agents. Ann Occup Hyg 1993;37, 253-270.
5. Boleij J, Buringh E, Heederik D, Kromhout H. Occupational Hygiene of Chemical and Biological Agents, Amsterdam, Elsevier 1995.
6. Rappaport SM, KromhoutH, Symanski E. Variation of Exposure Between workers in Homogenous Exposure Groups. American Industrial Hygiene Association Journal .1993; 54, 654-662.
7. Peretz C, Goldberg P. Kahan E, Grady S, Goren A. The Variability of Exposure over Time: A Prospective Longitudinal Study. Ann Occup Hyg .1997;41, 485-500.
8. Buringh E, Lanting R. Exposure Variability in the Workplace Its Implication for the Assessment of Compliance. American Industrial Hygiene Association Journal 1991; 52, 6-13.
9. Heederik D, Boleij JSM, Kromhout H. Use and Analysis of Exposure Monitoring Data in Occupational Epidemiology. An Example of an Epidemiological Study in the Dutch Animal Food Industry. Appl. Occupational Environ. Hygiene 1991;5, 458-464.
10. Rappaport SM., Lyles RH, Kupper LL. An Exposure Assessment Strategy Accounting for Within- and Between Worker Sources of Variability. Ann Occup Hyg 1995;39, 469-495.
11. Lyles RH, Kupper LL, Rappaport SM. A Lognormal Distribution-Based Exposure Assessment Method for Unbalanced Data. Ann Occup Hyg 1997;41, 63-76.
12. Burdorf A, Lillienberg L, Brisman J. Characterization of Exposure to Inhalable Flour Dust in Swedish Bakeries. Ann Occup Hyg 1994;38, 67-78.
13. Woskie SA., Shen P, Eisen EA, Finkel MH, Smith TJ, Smith R, Wegman DH. The Real Time Dust Exposures of Sodium Borate Workers: Examination of Exposure Variability. American Industrial Hygiene Association Journal 1994; 55, 207-217.
14. Preller L, Kromhout H, Heederik D, Tielen M. Modelling Long-Term Average Exposure in Occupational Exposure-Response Analysis. Scandinavian Journal of Work, Environment and Health 1995; 21, 504-512.

15. Preller, L., Heederik, D. , Kromhout, H. and Tielen, M. .1995; Determinants of Dust and Endotoxin Exposure of Pig Farmers: Development of a Control Strategy Using Empirical Modelling. Ann Occup Hyg 39, 545-557.
16. Nieuwenhuijsen MJ, Lowson D, Venable KM, Newman Taylor AJ. Flour Dust Exposure Variability in Flour Mills And Bakeries. Ann Occup Hyg 1995; 39, 299-305.
17. Kumagai S, Kusaka Y, Goto S. Cobalt Exposure Level and Variability in The Hard Metal Industry of Japan. American Industrial Hygiene Association Journal 1996;57, 365-369.
18. Kromhout, H, Swuste, P. and. Boleij, J. S. M. .1994; Empirical Modelling of Chemical Exposure in The Rubber Manufacturing Industry. Ann Occup Hyg 38, 3-22.
19. Symanski E, Kupper, L. L., Kromhout, H. and Rappaport, S. M. .1996;. An Investigation Of Systematic Changes In Occupational Exposure. American Industrial Hygiene Association Journal 57, 724-735.
20. Searle, S. R Mixed Models and Unbalanced Data: Wherefrom, Whereat And Whereto Commun. Statistics. A-Theory And Methods .1988; 17, 935-968.
21. Lindskey, J. K. Models for Repeated Measurements. Oxford, Clarendon Press,1993.
22. Burton, P., Gurrin, L. and Sly, P. Extending the Simple Linear Regression Model to Account for Correlated Responses: An Introduction to Generalized Estimating Equation and Multi-Level Mixed Modelling. Statistics in Medicine .1998; 17, 1261-91.
23. Nylander-French, L.A., Kupper, L. L. and Rappaport, S. M. An investigation of Factors Contributing to Styrene and Styrene-7,8-oxide Exposures in the Reinforced-Plastics Industry. Ann Occup Hyg .1999; 43, 99-109.
24. Rappaport, S. M., Weaver, M., Taylor, D., Kupper, L. and Susi, P.; Application of Mixed Models to Assess Exposures Monitored by Construction Workers during Hot Processesr. Ann Occup Hyg .1999;43, 457-69.
25. Burstyn, I., Kromhout, H., Kauppinen, T., Heikkila, P., and Bofefetta, P. Statistical Modelling of the Determinants of Historical Exposure to Bitumen and Polycyclic Aromatic Hydrocarbons Among Paving Workers Annals of Occupational Hygiene.2000; 44, 43-56.
26. SAS, SAS/ STAT Software Changes and Enhancements ,Carl, North Carolina, SAS Institute.1996; .
27. Kromhout, H., Loomis, D. P., Kleckner, R. C. and Savitz, D. A.; Sensitivity of The Relation between Cumulative Magnetic Field Exposure and Brain Cancer Mortality to Choice of Monitoring Data Grouping Scheme. Epidemiology .1997;8, 442-445.
28. Seixas NS, Sheppard L. Maximizing Accuracy and Precision Using Individual and Grouped Exposure Assessments. Scandinavian Journal of Work, Environment and Health. 1996; 22, 94-101.
29. BMDP, Statistical Software, Vol 2, Berkeley, University Of California 1990.
30. S-Plus, Data Analysis Products Division, Seattle, WA: Math-Soft. 1997.