Applied Occupational and Environmental Hygiene

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

Application of Statistical Models for Secondary Data Usage of the U.S. Navy's Occupational Exposure Database (NOED)

J. A. Formisano Jr , K. Still , W. Alexander & M. Lippmann

Applied Occupational and Environmental Hygiene
Volume 16(2): 201–209, 2001

1047-322X/01 $12.00 + .00

Application of Statistical Models for Secondary Data Usage of the U.S. Navy's Occupational Exposure Database (NOED)

J. A. Formisano, Jr.,[1] K. Still,[2] W. Alexander,[2] and M. Lippmann[3]
[1]Commander, Submarine Force, U.S. Pacific Fleet, Pearl Harbor, Hawaii; [2]Navy Health Research Center Detachment (Toxicology), Wright-Paterson Air Force Base, Ohio; [3]New York University School of Medicine, Tuxedo, New York

Many organizations around the world have collected data related to individual worker exposures that are used to determine compliance with workplace standards. These data are often warehoused and thereafter rarely used as an information resource. Using appropriate groupings and analysis of OSHA data, Gómez showed that such stored data can provide additional insight on factors affecting occupational exposures. Using data from the Occupational Exposure Database of the United States Navy, the usefulness of statistical models for defining probabilities of exposure above permissible limits for observed work conditions is examined. Analyses have highlighted worker Similar Exposure Groups (SEGs) with potential for overexposure to asbestos and lead. In terms of grouping data, Rappaport et al. defined the Within-Between Lognormal Model, a scale-independent measure for quantifying between-worker variability within a selected worker group:

$$_{B}R_{.95} = \exp[3.92 s_{sB}],$$

representing the ratio of arithmetic mean exposures received by workers in the 97.5th and 2.5th percentiles. To help search for groups, the Proportional Odds Model, a generalization of the logistic model to ordinal data, can predict probabilities for group exposure above the Occupational Exposure Limit (OEL), or the Action Level (AL), which is one-half of the OEL. Worker SEGs have been identified for asbestos workers removing friable asbestos ($_{B}R_{.95} = 11.0$) and non-friable asbestos ($_{B}R_{.95} = 6.5$); metal cleaning workers sanding specialized equipment ($_{B}R_{.95} = 11.3$), and workers at target shooting ranges cleaning up lead debris ($_{B}R_{.95} = 1.0$). Estimated probabilities for the categories <AL, AL-OEL, and >OEL support current understanding of work processes examined. Differences in probability noted between tasks and levels of ventilation validate this method for evaluating other available workplace exposure determinants, and for predicting probability of membership in categories that may help further define worker exposure groups, and determinants of excessive exposures. Thus, analyses of retrospective exposure data can help identify work site and work practice factors for efficient targeting of remediation resources.

Keywords Retrospective Worker Exposures, Similar Exposure Groups, Within-Between Lognormal Model

Many large organizations have developed complex systems for coding and storing occupational health and industrial hygiene data,[1] where the data are used to determine compliance with workplace standards, are subsequently warehoused and rarely, if ever, used again.[2–6] There is increasing recognition within the occupational health community that these kinds of reliable exposure data are all too often lacking or inaccessible. This research effort seeks to make additional productive uses of this type of stored data, and to understand the full capabilities of the unique construction of a database containing industrial hygiene and other occupational health information.

The Industrial Hygiene Information Management System (IHIMS) was developed for the United States Navy during 1989–1993 by a contractor-based group of industrial hygienists and data analysts, improving the hardware and system architecture of previously developed systems.[7,8] In this system worker information is organized based on job operation code, command, and location, linked by name and social security number (SSN), and is designed to look for trends of increasing exposure levels in the work environment. Industrial hygiene information collected at each Navy facility is submitted either on disk or hard copy to the Navy Environmental Health Center (NEHC) in Norfolk, Virginia, where the data are stored as part of the Navy Occupational Exposure Database (NOED).

As of May, 1999, there were 1.5 million people on active duty in the Armed Forces, with 365,922 of them serving with the U.S. Navy. There were at the same time 195,058 civilian workers in direct support of active duty personnel, maintaining 4108 aircraft and 323 ships needed for Navy operations. Personal breathing zone samples of these military and civilian workers make up the data in the NOED, with about 60 percent of the measurements reported for civilian workers.

The NOED uses IHIMS data fields for 392 job operation codes and 608 chemical substances, and contains more than 40,000 records. The personal sampling data used for this research was collected over the period 1987–1997 by Navy industrial hygienists, both civilian and military. If one accepts the estimate that the total cost for collection and analysis of one industrial hygiene air sample is typically from \$500–\$1000, (including collection time, processing, and administrative costs), the 1987–1997 records in the NOED represent an investment of from \$20 million to \$40 million.

This research utilizes data in the NOED to extend discussions in the literature[9–14] that have recommended continued evaluation and comparison of available collections of industrial hygiene data. The NOED appears to be an excellent database resource, linking personal sampling data and job tasks. It contains enough data to test the usefulness of the statistical models, and can be used to enrich the literature with generic estimates of personal exposures associated with commonly encountered industrial processes. The data linking worker exposures to observed tasks is of particular interest, and is not often found in other collections of industrial hygiene measurements. Since this project began in 1997, an additional two years of data have been entered into the NOED, creating additional opportunities for follow-on research and validation projects.

METHODS

Selecting Worker Exposure Groups

It has been noted that the most serious obstacle to effective utilization of an occupational health database lies in the database entries defining the occupational title or job description.[15,16] Such labels are normally designed to satisfy administrative purposes and often do not provide information related to worker exposures. The first step in analyzing the data available in the NOED was to isolate and define worker groups based on job title and location, allowing more detailed investigation. Microsoft Access version 2.0 was used for searching the data, and the resulting file can be exported to various statistical program packages. Minitab was selected as the primary software for analysis because of its usefulness in applying the regression model for cumulative logits.

The available exposure determinants from the database files in the NOED were correlated with the worker 8-hr time-weighted average (TWA). One line of data was recorded for a worker for a selected day of sampling, and includes all available variables. The Within-Between Lognormal Model[17] was used to estimate the between-worker variability for a selected worker group, to see if it met the requirements of a SEG.

Within-Between Lognormal Model

The Within-Between Lognormal Model[17] uses the components of variance determined for the selected worker exposure group, derived from ANOVA on the natural logarithms of the TWA values for the worker exposures.

Rappaport[18] has defined a scale-independent measure for quantifying the between-worker component of variance within a group:

$$ {}_{B}R_{.95} = \exp[3.92 s_{sB}] \qquad [1] $$

where ${}_{B}R_{.95}$ represents the ratio of the individual arithmetic mean exposures received by the 97.5th and 2.5th percentile workers. The estimator for s_{sB} is the square root of the between-worker component of variance determined from the analysis of variance test. This measure for ${}_{B}R_{.95}$ gives a scalar value of variability for 95 percent of the workers within the selected group, and allows different groups of workers to be compared to one another based on the magnitude of separation between worker mean values. Worker mean values relatively close to one another allow an analysis of the assignment of workers to a similar exposure group (SEG), with the expectation that all workers in the SEG would experience similar risks of health effects.[19]

What values for ${}_{B}R_{.95}$ are too large, and what values are low enough to consider a worker group "uniformly exposed" to an average exposure level is a matter of ongoing discussion. Rappaport has suggested 2 as a reference value, but stated that higher values may be sufficient for discussing groups depending on the needs of the analyst.[20] The Exposure Assessment Strategies Committee (EASC) of the American Industrial Hygiene Association (AIHA) stated that whether a factor of two is appropriate depends on the nature of the exposures, the goals of the exposure assessment strategy, and the resources available.[17] These concepts served as the basis for the decision to consider values of ${}_{B}R_{.95}$ near 10 to be sufficient to discuss the selected group as an SEG. In this case, the primary need of the analyst is to identify worker groups that are sufficiently similar to be classified as one group. The nature of the exposures within the NOED is that these exposures are not due to repetitive, similar work actions (as on a manufacturing assembly line), but combinations of work actions performed as maintenance, which can vary somewhat each time the work is performed. Values of ${}_{B}R_{.95}$ may range from 2–6 for workers performing the same work process under similar conditions almost every day.[19] The rationale here is that a tenfold difference in exposure levels for maintenance work would not be an unreasonable upper limit for discussing the individuals in the work group as having similar exposure levels over time.

Values for the number of workers (k) and total TWA measurements collected (N) can be found in the NOED that are in range for results as expressed by other researchers[12,19,21] when applying this model to various data sets.

To determine worker group variance components, a one-way ANOVA was selected using lnTWA as the target variable and worker ID as the factor. Let X_{ij} represent the exposure received by the i-th worker ($i = 1, 2, \ldots, k$) on the j-th day ($j = 1, 2, \ldots, n$), in an observational group. In keeping with current understanding of occupational exposure, the X_{ij} are assumed to have a lognormal density function.[22] Thus, the one-way random effects model applies to the logged exposures (Y_{ij}):

$$Y_{ij} = \ln(X_{ij}) = m_{my} + b_{bi} + e_{eij} \quad [2]$$

where b_{bi} represents the random deviation of the i-th worker from m_{my} and E_{eij} represents worker i's random deviation from his or her mean value (m_{myi}) on the j-th day.[23] Under the model it is assumed that b_{bi} and e_{eij} are normally distributed with zero means; that is, the normal distribution for $b_{bi} \sim N(0, s_{s^2B})$, and the normal distibution for $e_{eij} \sim N(0, s_{s^2N})$. The sets of $\{b_{bi}\}$ and $\{e_{eij}\}$ are mutually independent. The parameters s_{s^2B} and s_{s^2W} represent the between and within-worker variance, and comprise the total variance, s_{s^2y}. As such, $s_{s^2y} = s_{s^2B} + s_{s^2W}$ and $Y_{ij} \sim N(m_{my}, s_{s^2y})$.

With a balanced ANOVA model, the difference between the mean squares and error squares is divided by n_i, the number of measurements taken for each worker. The NOED data present an unbalanced model, and in general a worker can have from two to twelve measurements for a particular job. To estimate the between-worker component, the difference in mean squares must be divided by n_0, determined by:

$$n_0 = \left[N - \left(N/\left(sn_i^2\right)\right]/(k-1) \quad [3]$$

where N = the total measurements for the group, (sn_i^2) = the total sum of squares, and k = the number of workers in the group. The equation gives an estimator for n_0.[23] This value is usually between 2 and 3; approximately 30 percent of the values for n_0 calculated for selected groups in the NOED are greater than 3. Finding the point estimator for the between worker variability component (s_B^2) allows the determination of ${}_BR_{.95}$ for the selected worker group.

Proportional Odds Model

The variables listed in the NOED describe situations that are categorical, not continuous. For example, within the data file ventilation is designated as being adequate or not adequate, work is performed in one of six or seven locations, there may be two or more separate tasks, and so on. To take maximum advantage of these kinds of categorical variables, the cumulative logit model is most appropriate, and is the second major statistical model selected for this research. It was used by first transforming the target variable (TWA measurements of exposure) into three ordinal categories, Below one-half of the occupational exposure limit (OEL), that is, the action level (<AL); from the action level to the occupational exposure limit (AL-OEL); and above the occupational exposure limit (>OEL). A generalization of the logistic model to this type of ordinal data, based on cumulative logits, is called the proportional odds model, which is another way of describing the technique referred to as the polytomous or multinomial logistic regression model.[24] Traditionally, polytomous dependent variables have been handled with discriminant analysis.[24] Polytomous logistic regression may be preferable because it is a natural extension of logistic regression for a binary response.[24]

When the dependent variable is both polytomous and ordinal, it makes sense to form logits that take advantage of the ordered nature of the categories.[25] Cumulative logits are particularly appropriate when the construct underlying the ordinal measure is actually continuous, as is the case with using the worker TWA in the NOED data. Using the information available in the NOED this model can be used to predict probabilities for worker measurements that fall into the categories mentioned above, below the action level (<AL), between the AL and the OEL (AL-OEL), or greater than the OEL (>OEL).

To define the proportional odds model, consider a model with one predictor variable x and target group identifier Y.
Let:

$$L_j(x) = \text{logit}[F_j(x)] = \log\frac{[F_j(x)]}{[1 - F_j(x)]}, \quad j = 1 \ldots, J-1 \quad [4]$$

where $F_j(x) = P(Y \leq j \mid x)$ is the cumulative probability for response category j, and J is the total number of response categories (so Y takes on values $\{1 \ldots, J\}$). The proportional odds model states:

$$L_j(x) = a_j + bx, \quad j = 1 \ldots, J-1 \quad [5]$$

Thus a positive b is associated with increasing odds of being less than a given value j; a positive coefficient implies increasing probability of being in lower-numbered categories with increasing x. When applied to the NOED task data, the higher the odds ratio, the greater the probability of workers performing that task being in the group representing exposures below the OEL for the stressor in question. The model can indicate which combination of categorical variables will give the highest probability of exposures below the OEL based on the collected data.

Because most of the worker measurements are taken with a month or more between samples, the measurements collected for each worker are assumed to be independent. Some workers have measurements taken in a daily sequence of three or more days in a row. Time series analyses of personal exposure data by other researchers have provided little evidence of large day-to-day correlation.[26] The assumption is that personal exposure measurements of the same worker performing the same job two or three days in a row are not likely to lead to biased estimates of the mean and variance of the exposure distribution for that worker.

RESULTS

Analyses were performed for workers exposed to asbestos during asbestos removal operations, for workers exposed to lead while performing metal cleaning operations such as sanding and grinding, and for workers exposed to lead while working at military target shooting ranges. The findings were described in a doctoral dissertation.[27] This article discusses the results obtained for asbestos workers to illustrate the approach taken and the interpretation of the results.

The job operation codes for asbestos within the NOED presented good cases for testing the two models. As a regulated carcinogen with stringent requirements for sampling, extensive measurements for various types of asbestos work appear in the NOED, with a total of 1595 samples for asbestos in this data set. Table I lists stressors and numbers of measurements for the 20 chemical stressors having more than 500 measurements in the NOED, ranging from 537 to 6169. Table II shows results from the within-between lognormal model for five jobs involving asbestos. The selected stressor used in the analysis of each of these jobs is asbestos, non-specified, CAS No. 12001-29-5B.

For two jobs in Table II, removal using a glove bag (Line 1) and removal using negative pressure containment (Line 5), $_BR_{.95}$ values were evaluated using all measurements in the database, from all locations. The resulting $_BR_{.95}$ values of 15.0 and 22.9 are too large to allow selection of a mean value for exposure that could be applied to any one worker in these groups. For the other three jobs, gasket work (Line 2), insulation removal not elsewhere classified (Line 3) and brake work (Line 4) the scalar values of $_BR_{.95}$ are 4.5, 10.9, and 9.2. These are near or below the proposed criteria of 10 for selection of SEGs for this data. The implication would be that perhaps workers performing these job functions could be considered to belong to a SEG.

In terms of specific locations where asbestos removal is performed, it can be seen from Table II that the value of $_BR_{.95}$ for Location A is high (24.9), while Location B is low enough (2.1) to consider the workers at Location B to be an SEG. Clearly there is some difference between how this job is performed at Location A as compared to the same job at Location B, made apparent by using this model. Further investigation of all available exposure determinants might explain such an observed difference.

This being the case, the next step in the analysis is to follow the within-between lognormal model with the proportional odds model, to see if any other available variables can contribute information that would help form additional groups from measured worker exposure values, and create SEGs with $_BR_{.95}$ less than 10. The $_BR_{.95}$ value of 24.9 for Location A indicates that

TABLE I
Navy occupational exposure database: number of samples for selected stressors in the database as of december 1997

CAS number	Stressor	Number of samples	%
7439-92-1	Lead	6169	15.4
TOTAL DUST	Nuisance particulates, total dust	2710	6.8
7440-47-3A	Chromium metal and inorganic cmpds (as Cr)	2368	5.9
7440-43-9	Cadmium and compounds (as Cd)	2120	5.3
12001-29-5B	Asbestos, non-specified	1595	4.0
1309-37-1B	Iron oxide dust and fume (as Fe)	1320	3.3
7439-96-5	Manganese fume (as Mn)	941	2.3
108-88-3	Toluene (toluol)	930	2.3
7440-47-3F	Chromic acid and chromates (as CrO_3)	906	2.3
7440-02-0A	Nickel insoluble compounds (as Ni)	1195	3.0
7440-50-8B	Copper dust and mists (as Cu)	714	1.8
1309-37-1A	Iron oxide dust and fume (as Fe_2O_3)	792	2.0
7440-47-3E	Chromium (VI) insoluble cmpd NOC (as Cr)	769	1.9
1330-20-7	Xylene (o-, m-, p-isomers)	754	1.9
7440-50-8A	Copper fume (as Cu)	783	2.0
WELDING FUME	Welding fumes (NOC), total particulate	732	1.8
630-08-0	Carbon monoxide	601	1.5
7440-66-6	Zinc	529	1.3
75-09-2	Methylene chloride	546	1.4
78-93-3	Methyl ethyl ketone (2-butanone)	537	1.3
Sub-total:		27481	68.5
Remaining 292 stressors, 1–470 measurements per stressor:		12612	31.5
Total measurements in database as of December 1997:		40093	100.0

TABLE II
Navy occupational exposure database: within-between lognormal model: asbestos workers, stressor: asbestos, non-specified, CAS no. 12001-29-5B

	Job title	k	N	${}_{B}R_{.95}$	m_k	GSD	Location
1	Asbestos removal, glove bag	35	93	15.0	0.037	3.63	All Locations
2	Asbestos removal, gasket work	13	28	4.5	0.005	1.79	All Locations
3	Asbestos removal, insul. NEC	48	137	10.9	0.024	2.62	All Locations
4	Motor vehicle brake work	22	49	9.2	0.007	2.29	All Locations
5	Asbestos removal, neg. press.	47	188	22.9	0.108	4.56	All Locations
6	Asbestos removal, neg. press.	17	81	24.9	0.171	6.06	Location A
7	Asbestos removal, neg. press.	8	41	11.0	0.316	6.70	Location A, Task #1
8	Asbestos removal, neg. press.	10	35	6.5	0.033	2.57	Location A, Task #2
9	Asbestos removal, neg. press.	25	82	2.1	0.189	3.81	Location B

k = no. of workers.
N = total no. of measurements.
${}_{B}R_{.95}$ = scalar value or between-worker variability.
m_k = mean of 8-hr TWA (as fibers/cm^3).
GSD = geometric standard deviation, NEC = Not Elsewhere Classified.
Task #1 = removal of friable asbestos material = INSUL.
Task #2 = removal of non-friable asbestos material = TILE.

there is some misclassification of workers in this group.[17] Using a threshold limit value (TLV®) for asbestos, all forms, of 0.1 fibers per cubic centimeter (f/cc), the target categories for the categorical variables were defined as <AL = 0.00 to 0.05 f/cc; AL-OEL = 0.05 to 0.10 f/cc; and >OEL 0.1 to 25.0 f/cc.[28]

The proportional odds model was applied using two variables for Location A: task and local ventilation. These variables were selected because they have complete entries for all workers in the group, and for each variable categorical codes were chosen for one of two events or conditions. The asbestos removal category involves two distinct tasks, removal of friable, highly fiber-producing material, or removal of less friable material (such as roofing or floor tile), which is less likely to produce airborne fibers. Removing friable insulation was coded as Task 1 = INSUL = 0, and removing tile or non-friable asbestos material was coded as Task 2 = TILE = 1. Local ventilation was coded as NPRES for negative pressure ventilation (most likely to lower airborne fiber levels), or GEN for general ventilation. Clearly one would expect the worst case to be removal of friable, fiber-producing asbestos with general ventilation.

Ordinary least squares regression was performed for Location A, using task and ventilation as dependent variables. As can be seen in Table III, both variables are significant; since this is a logit model, an odds ratio of 9.84 for task is an indication that for a unit change in the variable task (i.e., moving from friable to non-friable) the odds of the exposure level being in a lower category are exp(2.2868), or 9.84 times as great moving from friable to non-friable material. Such a high ratio is a result of the clearly different exposure levels recorded for workers performing the two tasks.

Probabilities obtained from this model are shown in Table III. It can be seen that as expected, the highest probability of being below the action level is for TILE and NPRES, representing removal of non-friable asbestos with negative pressure ventilation. The highest probability of being above the occupational exposure limit is INSUL and GEN, removal of friable asbestos material without negative pressure, using general or room air ventilation.

To seek validation, additional data not used to build the model were examined. Within the NOED database, the last 8-hr TWA entry for a worker performing asbestos removal is 9/5/96. Additional worker measurements, taken up to 4/29/99 were obtained from the Navy Environmental Health Center. All available measurements for workers in the job operation codes for asbestos removal after 9/5/96 were examined. A total of 37 records were formed and all available variable fields were matched to each worker TWA. Two workers were listed as "setting, up containment area," and these two jobs were coded as likely to produce low-level exposures. The other 35 records were listed as "ripout," and no further description of the type of asbestos material was provided. All ventilation was listed as general, since no negative pressure enclosures were indicated.

Applying the proportional odds model to these additional measurements, Table III shows the probabilities obtained for the worker tasks from data collected during 1997–1999 compared to the data described above, collected during 1987–1997.

It appears that for the tasks of setup of containment and ripout of asbestos material, the derived probabilities support the model. The workers performing setup may have been exposed to some small levels of asbestos material present in the work area, explaining why this group shows a probability of being below the AL of 0.62 compared with a probability of 0.92 when removing non-friable tile. On the other hand, for workers performing ripouts, a higher probability of being greater than the OEL,

TABLE III

Navy occupational exposure database, location A; variables: task and ventilation

Ordinary least squares regression:

$$\ln\text{TWA} = -1.59 - 1.54\ \text{Task} - 1.23\ \text{Vent}$$

Predictor	Coef	StDev	T	P
Constant	−1.5881	0.2547	−6.23	0.000
Task	−1.5437	0.3562	−4.33	0.000
LocVent	−1.2286	0.3600	−3.41	0.001

Logistic regression:

Predictor	Coef	StDev	Z	P	Odds ratio	95% lower	CI upper
Const(1)	−1.9713	0.4741	−4.16	0.000			
Const(2)	−1.4256	0.4403	−3.24	0.001			
Task	2.2868	0.6182	3.70	0.000	9.84	2.93	33.07
Vent	2.1596	0.5653	3.82	0.000	8.67	2.86	26.25

Probabilities using the proportional odds model:

Task	Vent	$<$AL	AL-OEL	$>$OEL
TILE	NPRES	0.92	0.03	0.05
TILE	GEN	0.58	0.12	0.30
INSUL	NPRES	0.55	0.13	0.32
INSUL	GEN	0.12	0.07	0.81

Probabilities for data collected during 1987–1997, compared to data obtained during 1997–1999

Years	Task	Vent	$<$AL	AL-OEL	$>$OEL
87–97	TILE	NPRES	0.92	0.03	0.05
87–97	TILE	GEN	0.58	0.12	0.30
87–97	INSUL	NPRES	0.55	0.13	0.32
87–97	INSUL	GEN	0.12	0.07	0.81
97–99	SETUP	GEN	0.62	0.13	0.25
97–99	RIPOUT	GEN	0.27	0.14	0.59

almost 60 percent, seems to show that at least some of them may have been exposed to friable asbestos material. Because the probability for the 1997–1999 “ripout” group is not as high as for the 1987–1997 group identified as removing friable insulation material, it may be that one or more of the workers in the former group were removing non-friable material that was not described as such, and thus the tasks were incorrectly coded. Still, it is clear that there is a higher probability of being above the OEL in this group.

Ordinary least squares regression determined the following predicted exposure levels and 95 percent confidence intervals for the 1987–1997 data:

Task	Vent	f/cc	95% C.I.
TILE	NPRES	0.013	(0.008, 0.020)
INSUL	GEN	0.204	(0.123, 0.339)

The average for the 35 workers performing ripout with general ventilation during 1997–1999 is 0.511 f/cc, much higher than the predicted value above. However, it was noted that on one day in particular, eight workers had very high fiber measurements, and it seems all were working at the same location. If the eight results from this one day are removed from the group, then the average for the remaining 27 workers is 0.226, within the predicted confidence interval derived from the 1987–1997 data. The average for the eight worker measurements from the one day in question was 1.48 f/cc, significantly higher than the predicted value as well as allowable exposure limits. Some unusual occurrence or work procedure may have contributed to these high values for this work day. The average for the two workers performing setup of a containment area was 0.039 f/cc, slightly higher than the confidence interval predicted for working with non-friable material with negative pressure, and is probably due to performing the work using general ventilation.

DISCUSSION

The most important and useful characteristic of the database described in this research has been the extensive number of entries on worker tasks and ventilation status in the work environment during sampling; this level of detail has not been encountered for databases that have been described in the literature. Another major characteristic that facilitated this research has been the multiple entries for individual workers performing the same job over time. These multiple entries can be separated in time by a few days, to months or years. Depending on the worker group selected for analysis, from 21 percent to 53 percent of workers had two or more measurements, and could thus be analyzed using the within-between lognormal model. The NOED was not specifically designed for this type of research, or for use of the within-between lognormal model as such, as has been the case for at least one other set of data.[29] Still, an overall assessment is that the NOED represents an excellent source of information, and these models have provided a good way to search the data. At the time of this writing, enough additional measurements may have been added to the NOED to make re-evaluation of the groups listed in Table II a worthwhile endeavor.

Confidence in the validity of this model is based on other recorded values for asbestos removal, and industrial hygiene professional experience. Previously collected data have established the fact that removal of friable asbestos without adequate ventilation will result in much higher exposure levels, often over the OEL. The large difference in probability noted between the two tasks, and two levels of ventilation, validate this method for evaluation of other variables from other selected observational groups.

CONCLUSIONS

The method can help predict probability of worker membership in groups that present much lower worker exposure values, and less distinct differences in exposure levels between two different work conditions. The models can be applied to any combination of job operations and tasks, particularly in those cases where the difference between tasks may not be as clear or well documented as in the case of asbestos. Given the usefulness of these models for asbestos, it is expected they would prove equally useful for other stressors and other tasks.

Based on these interpretations, the predictions and exposure estimates produced by the models based on the 1987–1997 NOED data would seem to be supported by the additional collected measurements from the period 1997–1999. The results obtained are not inconsistent with the hypothesis that persons performing the same job under the same conditions will have a similar probability of exposure to similar concentrations of contaminants. The analysis also shows the importance of accurately recording the task performed, and its relationship to the modeling variables.

What makes the method of applying these two models promising is that they can be applied to any extensive data set, using any combination of variables, as long as the coding of the variables is consistent within that database. Although two or more databases may differ greatly in design, even if coding has been done differently, the models could conceivably still provide the same SEGs containing the same variables, tasks, and so on for an identified worker group. Similarities between different databases that have collections of data on essentially the same types of jobs would become apparent after applying the method described here.

In an editorial reply related to the discussion of the within-between lognormal model for determining variance components, Manuel Gómez stated:

> "...The information about exposure determinants that is currently collected with most exposure measurements is woefully inadequate, and a strong argument could be made that better collection of such information would allow far better grouping, analysis, and utilization of exposure data. It would also allow better modeling and many other improved uses of exposure data. One step in that direction would be to agree on the determinants we need to collect, how we will define them, and how we will code them. It is worth a try."[30]

This examination of data collected in the NOED has supported this statement. It appears that the proportional odds model (using ordinal categories) is a reliable method of highlighting combinations of variables that may contribute to higher worker exposures. Other research groups have used the within-between lognormal model, and examined the values for ${}_{B}R_{.95}$ in comparing between-worker variability when moving from larger observational groups to more narrowly defined groups.[19]

The models can indicate where conditions of higher potential worker overexposures are likely to occur within the sampled areas, using all available information, in situations that may not be so clear-cut. A job with many different tasks or a job process composed of a sequence of tasks having only a relatively small difference in exposure levels between tasks, would be much easier to detect. In cases where only small differences in exposure level are very important (chronic exposures, or toxic material), an ability to determine which tasks or combination of tasks give a greater probability of being higher than a chosen exposure level would be very valuable. The implications are that such groups, when identified, can be more closely observed over time to detect possible health effects related to the stressor in question.

RECOMMENDATIONS

1. For worker groups identified as similar exposure groups (SEGs), a mean TWA for a group appreciably below the OEL is a clear indication that workplace measures taken to limit worker exposures are effective. Industrial hygiene and occupational health personnel should continue to monitor such operations, but further routine collection of sampling measurements may not be necessary, and would not be recommended unless the process changes. This type of decision involves a reallocation of often scarce resources, and must be supported by workplace measurements. Results from the type of research discussed here would provide additonal support for this recommendation.

2. For other major job/stressor combinations in the NOED for which SEGS could not be defined (i.e., welding/cadmium fumes, or spray painting/toluene), additional measurements should be collected to specifically support an extension and refinement of the models discussed here. Such sampling would entail obtaining two or more measurements for several (five or more) identified workers for each job, in an effort to determine when the job/stressor combination meets the requirements for a SEG.
3. An exhaustive analysis of this data set, focussed particularly on task data, could not only identify SEGs, but most likely would highlight groups previously unrecognized as performing tasks with higher probabilities for overexposure. Such groups could be identified for referral to the occupational health physicians, and also be included in the technical manual listing the recommended medical surveillance procedures for various stressors encountered in the U.S. Navy.
4. It would be possible to recommend to field industrial hygiene personnel that every effort should be made to collect at least two measurements per worker for a period of time, perhaps six months to a year, in an effort to help provide more measurements for analysis and review using these models. Such coordinated, concerted action to sample for one job/stressor combination throughout the Navy could provide sufficient data for analysis using these models.
5. The implications of these models make clear the importance of submitting data with the Task and Ventilation fields correctly filled out, using menu selections from the IHIMS software. Quality assurance checks could be initiated to ensure that at least these two fields are standardized, so that coding can be done without having to review each separate entry. Quality assurance support can ensure that the data are entered using menu entries that will make coding of variables much easier, to better allow the use of statistical software in the application of these models.

DISCLAIMER

The research described in this thesis reflects solely the views and opinions of the author, and does not reflect either the views or policies of the Navy Environmental Health Center (NEHC), the Bureau of Medicine and Surgery (BUMED), the U.S. Navy, or the Department of Defense.

REFERENCES

1. Armstrong, T.W.: Past and Current Approaches to Occupational Exposure Databases in the Private Sector. Appl Occup Environ Hyg 10(4):257–263 (1995).
2. Lippmann, M.: Research Needs: Retrospective Exposure Assessment for Occupational Epidemiology. Appl Occup Environ Hyg 6(6):550–552 (1991).
3. Gómez, M.R.; Rawls, G.: Conference on Occupational Exposure Databases: A Report and Look at the Future. Appl Occup Environ Hyg 10(4):238–243 (1995).
4. Horning, R.W.; Greife, A.L.; Stayner. L.T.; et al.: Statistical Model for Prediction of Retrospective Exposure to Ethylene Oxide in an Occupational Mortality Study. Am J Ind Med 25:825–836 (1994).
5. Harris, R.L.: Guideline for Collection of Industrial Hygiene Exposure Assessment Data for Epidemiologic Use. Appl Occup Environ Hyg 10(4):311–316 (1995).
6. Goldberg, M.; Hemon, D.: Occupational Epidemiology and Assessment of Exposure. Int J Epidemiol 22(6)(Suppl 2):S5–S9 (1993).
7. Industrial Hygiene Information Management System (IHIMS) Technical Manual. Navy Environmental Health Center, Norfolk, VA (October 1994).
8. Pugh, W.M.; Beck, D.D.; Ramsey-Klee, D.M.: An Overview of the Navy Occupational Health Information Monitoring System (NOHIMS), Naval Health Research Center Report No. 83-8. Naval Medical Research and Development Command, Bethesda, MD (1984).
9. Goldberg, M.; Kromhout, H.; Guenel, P.; et al.: Job Exposure Matrices in Industry. Int J Epidemiol 22(6)(Suppl 2):S10–S15 (1993).
10. Bouyer, J.; Hemon, D.: Retrospective Evaluation of Occupational Exposures in Population-Based Case-Control Studies: General Overview with Special Attention to Job Exposure Matrices. Int J Epidemiol 22(6)(Suppl 2)S57–S64 (1993).
11. Bouyer, J.; Hemon, D.: Studying the Performance of a Job Exposure Matrix. Int J Epidemiol 22(6)(Suppl 2):S65–S71 (1993).
12. 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(2):115–119 (1998).
13. Hornung, R.W.; Herrick, R.F.; Stewart, P.A.; et al.: An Experimental Design Approach to Retrospective Exposure Assessment. Am Ind Hyg Assoc J 57:251–256 (1996).
14. Esmen, N.: Retrospective Industrial Hygiene Surveys. Am Ind Hyg Assoc J 40:58–65 (1979).
15. Special Report by Joint ACGIH-AIHA Task Group on Occupational Exposure Databases: Data Elements for Occupational Exposure Databases: Guidelines and Recommendations for Airborne Hazards and Noise. Appl Occup Environ Hyg 11(11):1294–1311 (1996).
16. Gómez, M.R.: Factors Associated with Exposure in Occupational Safety and Health Administration Data. Am Ind Hyg Assoc J 58:186–195 (1997).
17. Mulhausen, J.R.; Damiano, J.: A Strategy for Assessing and Managing Occupational Exposures, 2nd Edition. AIHA Press, Fairfax, VA (1998).
18. Rappaport, S.M.: Selection of the Measures of Exposure for Epidemiology Studies. Appl Occup Environ Hyg 6(6):448–457 (1991).
19. Heederik, D.; Boleij, J.S.M.; Kromhout, H.; Snid, 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(6):458–464 (1991).
20. Rappaport, S.M.; Lyles, R.H.; Kupper, L.L.: An Exposure-Assessment Strategy Accounting for Within- and Between-Worker Sources of Variability. Ann Occup Hyg 39(4):469–495 (1995).

21. Tornero-Velez, R.; Symanski, E.; Kromhout, H.; et al.: Compliance Versus Risk in Assessing Occupational Exposures. Risk Anal 17(3):279–292 (1997).
22. Rappaport, S.M.: Assessment of Long-Term Exposures to Toxic Substances in Air. Ann Occup Hyg 35:61–121 (1991).
23. Rosner, B.: Fundamentals of Biostatistics, 4th ed., pp. 299–329. Duxbury Press, Belmont, CA (1995).
24. Hosmer, D.W.; Lemeshow, S.: Applied Logistic Regression, pp. 217–241. John Wiley & Sons, New York (1989).
25. Demaris, A.: Logit Modeling Practical Applications, pp. 61–79. Sage Publications, Newbury Park, CA (1992).
26. Francis, M.; Selvin, S.; Spear, R.; Rappaport, S.: The Effect of Autocorrelation on the Estimation of Worker's Daily Exposures. Am Ind Hyg Assoc J 50(1):37–43 (1989).
27. Formisano, J.A., Jr.: Analyses of the U.S. Navy's Occupational Exposure Database. Doctoral dissertation, New York University School of Medicine, Tuxedo, NY (May 2000).
28. American Conference of Governmental Industrial Hygienists (ACGIH): 1999 TLVs® and BEIs®, Threshold Limits Values for Chemical Substances and Physical Agents, Biological Exposure Indices. ACGIH, Cincinnati, OH (1999).
29. Kromhout, H.; Symanski, E.; Rappaport, S.: A Comprehensive Evaluation of Within- and Between-Worker Components of Occupational Exposure to Chemical Agents. Ann Occ Hyg 37(3):253–270 (1993).
30. Gómez, M.R.: Letters to the Editor. Am Ind Hyg Assoc J 55:873–877 (1994).