American Journal of Industrial Medicine 25:825–836 (1994)

Statistical Model for Prediction of Retrospective Exposure to Ethylene Oxide in an Occupational Mortality Study

Richard W. Hornung, Dr PH, Alice L. Greife, PhD, CIH, Leslie T. Stayner, PhD, N. Kyle Steenland, PhD, Robert F. Herrick, ScD, CIH, Larry J. Elliott, MS, CIH, Virginia L. Ringenburg, BS, and John Morawetz, MSc

Since direct measures of individual exposure seldom exist for the entire period of an occupational mortality study, retrospective exposure estimates are necessary. This is often done in a subjective manner involving a consensus of opinion from a panel of epidemiologists and industrial hygienists. An alternative method utilizing a statistical model provides a more objective procedure for retrospective exposure assessment. The development of a weighted multiple regression model is presented for estimation of exposure levels to ethylene oxide (ETO) for inclusion in a cohort mortality study of workers in the sterilization industry. Three steps in development of the model are described: (1) data acquisition and assessment, (2) model building, and (3) evaluation of the model. The final model explained a remarkable 85% of the variability in 205 average measurements of ETO levels. Exposure factors included in the model were exposure category, product type, size of the sterilization unit, selected engineering controls, days after sterilization, and calendar year. The model was evaluated in two ways: against a set of measurement data not used to develop the model and a panel of 11 industrial hygienists representing the sterilization industry. The model predicted ETO exposures within 1.1 ppm of the validation data set with a standard deviation of 3.7 ppm. The arithmetic and geometric means of the 46 measurements in the validation data set were 4.6 and 2.2 ppm, respectively. The model also outperformed the panel of industrial hygienists relative to the validation data in terms of both bias and precision.

Key words: exposure assessment, regression models, sterilization industry workers, ETO, occupational mortality

Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Cincinnati, Ohio (R.W.H., A.L.G., L.T.S., N.K.S., R.F.H., L.J.E., V.L.R.).
International Chemical Workers Union, Center for Workers Health and Safety Education, Cincinnati, Ohio (J.M.).
Address reprint requests to Dr. Richard Hornung, Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Mailstop R-44, Cincinnati, OH 45226.
Accepted for publication May 13, 1993.

INTRODUCTION

One of the most serious problems confronting epidemiologists and industrial hygienists in conducting a retrospective occupational mortality study is the characterization of past exposures. While it is usually possible to obtain detailed information on many characteristics of workers in the study population, that is, names, ages, job titles, vital status, and cause of death, comparable information on the exposure of interest is rarely available [Stewart et al., 1991].

When direct measures of exposure do not exist, estimates of individual exposures must be made. This can be done in a number of ways, some more difficult than others. One way is for the principal investigators to assign exposure levels for each worker, based upon their subjective judgements concerning levels in past years and the probable effect of changes in operating conditions. Another way is to enlist the help of persons not involved with the study, such as industrial hygienists familiar with current and past exposure levels at the study site(s) [Kromhout et al., 1987].

Each of these approaches has problems which may compromise the interpretation of the results of the study. The principal investigators often have only limited information or anecdotal reports of conditions in the study site(s) in past years. When representatives of the industry being studied who are familiar with past conditions are utilized for providing estimates of exposure levels, there may be a conflict of interest, whether it is perceived or real.

A desirable alternative to these two approaches would be an objective estimation procedure which makes maximum use of existing exposure data. When sufficient industrial hygiene (IH) data are available, one such method is the use of a statistical model developed from data sampled under conditions which may potentially influence exposure levels. The use of such algorithms and statistical models for prediction of occupational exposure levels has seen limited attention to date [Eisen et al., 1984; Dement et al., 1983; Woskie et al., 1988]. This approach can be used to fill in gaps in job/plant/year information, as well as to project exposures to time periods when no measurement data existed. The advantage of this approach is that it permits an objective determination of exposure levels as a function of factors whose effect can be estimated and statistically tested.

Three steps are presented in development of such a model for use in predicting exposure levels of ethylene oxide (ETO) in the sterilization industry. These estimates are used to calculate the cumulative exposure to ETO for each worker in order to conduct an exposure-response assessment in an epidemiologic mortality study conducted by the National Institute for Occupational Safety and Health (NIOSH) [Steenland et al., 1991].

METHODS

Development of models for prediction of both occupational and environmental exposures has begun to receive considerable attention. There have been a few examples of statistical prediction models for environmental exposures [Ostro, 1987; Hasabelnaby et al., 1989; Schwartz, 1991]. Most of the interest, however, has centered around mathematical or deterministic models in the areas of both environmental and occupational exposures [Jaylock, 1988; Yu et al., 1990]. The approach reported here is considerably different from those employed to develop deterministic models. Deterministic models are generally built around mathematical principles associated with physics and chemistry. Statistical models, on the other hand, are descriptive in nature and are developed by fitting them to observed data. Of course, in order for any statistical model to have credibility, it must not be at odds with physical principles.

Development of the statistical model can be broken down into three steps: (1) data acquisition and assessment, (2) model building, and (3) evaluation. A description of each of these steps as used in the ethylene oxide (ETO) study follows.

Data Acquisition and Assessment

The first step is determination of the amount of industrial hygiene data available and suitable for use in development of the exposure prediction model. Data for the ETO study was obtained from 20 different facilities subsequent to walk-through surveys of a total of 36 different plants engaged in sterilization of spices and medical supplies. During these surveys, NIOSH industrial hygienists and epidemiologists assessed the quality and availability of historical measurements of ETO, work histories of the employees, and information regarding other factors potentially influencing exposure levels, for example, engineering controls. Sixteen facilities were excluded from the exposure assessment study for one or more of the following reasons: no personal sampling for ETO, no documentation of sampling and analytical methods, or sampling records lacked sufficient detail to link sampling results with job categories. Descriptions of this process have been previously reported [Griefe et al., 1988; Steenland et al., 1987].

Each measure of exposure used in the model development was an arithmetic mean of all samples taken for a given job and/or location within each plant for a given year. The majority of these data was collected after the late 1970s when the health effects of ETO became widely known [Hogstedt et al., 1979; Garry et al., 1979]. Most of the data (approximately 75%) were personal samples taken with charcoal tubes. The remaining data were collected using personal passive dosimeters or area samples. Because of a significantly higher degree of variability in these latter data, as well as NIOSH experience concerning a lack of strong agreement between personal charcoal tube samples and either area samples or passive monitor data [Leidel et al., 1977], only personal charcoal tube data were used in building the model.

When more than one sample was taken for a given plant, location, job, and year; standard deviations were obtained as a measure of the reliability of the recorded exposure level. Both the number of samples taken and the standard deviation were used as weights in fitting the exposure prediction model. The different variables which were potential candidates for inclusion in the model are listed in Table I. These variables were determined during walk-through surveys of sterilization facilities conducted by NIOSH investigators.

Model Building

Model building is an art more than a science. Given a complex set of data, it is doubtful that any two statisticians would arrive at exactly the same model. The steps used in developing this model are described with brief explanations for their utility. Other analysts may have used slightly different approaches but hopefully would have found similar results. The model development phase was initiated after the personal charcoal tube data were edited and any obvious errors were corrected. A weighted multiple linear regression analysis using PROC GLM in the Statistical Analysis

TABLE I. Variables Available for Development of a Model for Predicting Retrospective Exposure to Ethylene Oxide

Calendar year of operation
Job
Location or department
Type of product produced
Annual pounds of ETO gas mixture used
Vessel volume
Engineering controls
Multiple vessel airwashers
Powered rear vessel exhaust
Vessel isolation
Vacuum pump isolation
Localized exhaust over vessel door
Localized exhaust over ETO gas cylinders
Sealed vessel/sanitary drain connection
Increased vessel area dilution ventilation
Administrative control
Use of control room
Process controls
Recirculation of exhausted ETO gas
Determination of ETO concentration in vessel
Decreased ETO concentration in vessel
ETO/diluent gas mixture
Product age (days since product sterilization or treatment with ETO)
Vessel charging—manual
Vessel charging—automatic
Aeration procedure

System (SAS) was used as the basic model. Because of the large number of potential independent variables and in the interest of simplicity, we decided to use no more than a second order model, that is, one with no more complex terms than two-way interactions or quadratic effects.

All independent variables were examined with respect to their range and correlation with other independent variables in order to assess their potential influence on the model. A high degree of correlation among independent variables (multicollinearity) can cause bias in estimating the coefficients when attempting to develop a regression model. This approach revealed that a combination of the variables "job" and "location" would be more appropriate than entering both in the model individually because of their high degree of correlation. Accordingly, the variable "exposure category" was created to roughly correspond to proximity to the source of exposure. This variable took the same values as "location," with the exception of sterilizer operators and maintenance workers who were assigned to separate categories regardless of their recorded location.

Therefore, all job titles at a given location were assumed to have homogeneous exposure levels. This pooling of job titles resulted in eight different exposure categories as shown in Table II. In order to assure that the residuals of the model were normally distributed, a transformation of the dependent variable (mean exposure level) was desirable. Since the exposure measurements were highly skewed, a logarithmic transformation was used. However, it should be noted that the normality

TABLE II. Exposure Categories and Product Types Used in Retrospective Exposure Model

Exposure categories	Product types
1. Laboratory	1. Spore strips
2. Sterilizer area (except operator)	2. Plastic
3. Warehouse	3. Gauze
4. Production area	4. Spices
5. Clean room	5. Glass or metal
6. Quarantine room	
7. Sterilizer operator	
8. Maintenance	

assumption is not as important to a prediction model as to a model used for hypothesis testing.

Desirable features of the log transformation are that it produces a multiplicative effect for each of the independent variables and permits no negative estimates of exposure level; that is, each exposure factor included in the final model produces a proportional increase or decrease in the estimated ETO level. This is generally preferable to an additive model, which may produce negative exposure estimates under certain combinations of exposure factors. The weight assigned to each value of the dependent variable was the inverse of the variance of that value (the log of the mean of several annual samples). The weight is a function of the coefficient of variation (CV) and sample size (n).

The weighting function was of the form:

$$WT = 1/[\log(CV^2/n + 1)],$$

where CV = standard deviation/mean. When no estimate of the standard deviation was available, the value 0.426 was substituted for the CV, since this was the average CV across all data used in developing the model.

Before the model was actually developed, data from six randomly selected plants (representing approximately 20% of the total data) were removed from consideration, to be used later in evaluating the final model. Two other plants were eliminated since they had exposure data taken only with passive monitors. This left 205 annual arithmetic means based upon 2,350 full-shift charcoal tube measurements from 12 different plants to be used in developing the model. Since there were no measurement data available before 1976, the majority of cells in a job-exposure matrix dating back as early as 1943 were missing and would require estimation, using the model.

The last step before fitting the model was determination of availability of each of the potential independent variables for every plant used in the epidemiology study. If data for some independent variables were not available for all or nearly all plants in the mortality study, then they had to be removed from consideration for entry into the model. One such casualty was "pounds of ETO" used each year, which was a surrogate measure of potential level of exposure. Since these data were not available for all plants in the study, the size of the sterilizer units (in cubic feet of capacity) was substituted after we determined that there was a high degree of correlation between these two variables.

The model was developed by forcing several key variables into all runs and then entering and assessing other variables one at a time. Because of their high probability for impact on ETO levels, exposure category and product type were two factors that were used in all potential models. Variables were kept in the model if their level of significance (p value) was less than or equal to 0.10. Since this was an exercise in estimation rather than hypothesis testing, tight control of type I errors (e.g., $p < .05$) was not essential. The overall goal was production of a stable model which explained a high degree of variability in ETO exposure levels with a minimum number of terms.

Evaluation of the Model

After developing a stable model which explains an acceptable amount of variability in exposure levels, it is necessary to verify that this model can be used reliably in prediction of historical exposures. Since no data were available prior to the mid-1970s, we chose a three-phase evaluation procedure.

We first applied the model to a subset of our original data file not used to develop the model. This is a variation of a statistical technique known as cross-validation or data splitting [Picard and Berk, 1990]. This subset consisted of 46 arithmetic means based upon 350 personal charcoal tube samples from six different plants between the years 1979 and 1985. We then compared the model's predictions to the actual means of measured ETO exposure levels. In this way we could assess both the degree of bias (average difference) and precision (standard deviation of differences) associated with the model's predictions.

Another way of evaluating the performance of the model was by comparing its predictions to those of a panel of industrial hygienists familiar with ethylene oxide levels in the sterilization industry. NIOSH assembled a panel of 11 such industrial hygienists from the industries involved in the study. Each industrial hygienist was provided data on the same set of independent variables used to develop the model. Values for all independent variables were those associated with the 46 samples comprising the evaluation data set. In order to have a baseline for estimation, the industrial hygienists were also provided with the annual average ETO level in the earliest year in which a given set of exposure conditions existed. They were then instructed to estimate the ETO exposure for each set of conditions in a subsequent year for the 46 samples of the evaluation data set. The amount of bias and precision for each set of conditions were then computed and compared to those of the model. In order to calculate bias and precision, we used the measured level of ETO in the evaluation data set as the "true" measure of exposure. We discuss the implications of this assumption later.

The last test of the reliability of the model was its actual use in generating an exposure matrix for the occupational mortality study. If no obvious errors in prediction were discovered, that is, predicted values outside the broad range of possible exposure levels, the model could then be used to predict annual exposure levels for each worker in the epidemiologic study. Although this phase may seem trivial, it is very important to examine the final exposure matrix, since some conditions encountered in the actual study may involve extrapolations from situations when measurements were taken. However, it is also necessary to avoid second guessing the model's prediction without strong evidence that the predicted value is unreasonable.

RESULTS

After a number of generations of models, a working model containing exposure category, product type, age of product, calendar year, rear exhaust, aeration procedure, and sterilizer volume was selected based upon criteria previously described in step 2.

After the first order or main effects terms were identified, all possible second order terms (quadratic effects and two-way interactions) were examined. A model containing quadratic terms for calendar year and sterilizer volume, as well as the interactions of exposures category with product type and also with calendar year, was selected. This model explained a remarkable 90% of the variation in ETO exposure levels. Of the 17 different engineering, administrative, and process controls examined for potential inclusion in the model, only aeration of the product after sterilization and use of a powered rear exhaust valve in the sterilizer unit provided consistent and statistically significant reductions in ETO levels. A decreasing trend in ETO exposures with calendar year of operation was a highly significant factor.

We had hoped that some combination of engineering controls would eliminate the need for including calendar year in the model, since we assumed introduction of such controls was primarily responsible for annual decreases in ETO levels since the late 1970s. However, no combination of variables could be found to allow removal of calendar year from the model. We attributed this finding to calendar year acting as a surrogate for improvement in work practices due to increased awareness of the potential health effects of ETO.

Since little if any data were available prior to the late 1970s (7 means based upon 23 samples in 1976–1978), use of calendar year in the model for predicting earlier exposures necessitated extrapolation beyond the range of the data. However, the effect of a calendar year was nonlinear and resulted in a trend which attained a maximum at approximately 1978, as seen in Figure 1. Since we felt that the decrease in ETO levels after 1978 (independent of engineering controls) was explained by improved work practices after ETO was identified as a potential carcinogen, we set each predicted ETO level prior to 1978 equal to the predicted level in 1978. Variation in exposure levels prior to 1978 were modeled as a function of the remaining terms in the model with the calendar year effect fixed at 1978. Therefore, there was no extrapolation by calendar year prior to 1978.

Based upon the results of the validation procedure described earlier, we made further modifications to the model. By observing the pattern of differences between the model's predictions and the actual measurements (Table IV), we determined that the interaction terms between exposure category and product type, as well as exposure category and calendar year, should be removed. These interactions caused the model to be overly specific for the model development data. Also, a number of the possible combinations of exposure category and product type resulted in empty cells.

When the interaction terms were removed, the model still explained a very high 85% of the variability in ETO exposure levels (Table III). Both the interaction model and the model with interactions removed were compared to the 46 samples in the evaluation data set. This was done by taking the difference between the measured ETO level and the ETO level predicted by the model from the set of independent variables under which each sample was taken, that is, exposure category, product, calendar year, etc. Table IV demonstrates that the reduced model resulted in a sub-

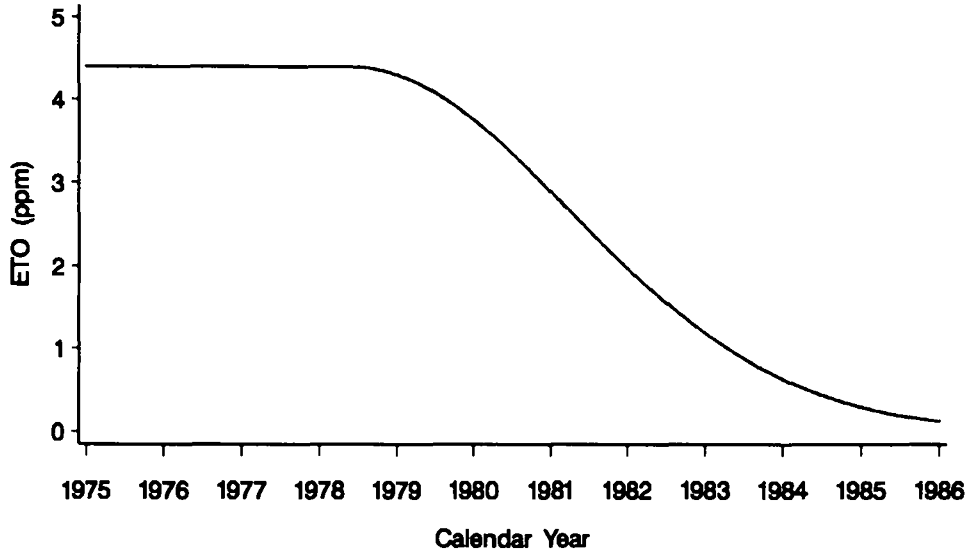

Fig. 1. Example of predicted ETO levels vs. calendar year. Actual ETO level (height of curve) will depend on other exposure factors in the model.

TABLE III. Weighted Multiple Regression Model for Estimation of Retrospective Exposure to Ethylene Oxide

Exposure factor	d.f.	F test	p value
Exposure category	7	3.40	0.002
Product type	4	17.41	<.001
Age of product	1	68.31	<.001
Year	1	124.02	<.001
$(\text{Year})^2$	1	31.91	<.001
Exhaust	1	4.02	0.046
Aeration	1	2.71	0.101
Cubic feet	1	77.35	<.001
$(\text{Cubic feet})^2$	1	55.71	<.001
Error	186		

$R^2 = 0.85$.

stantial increase in precision when compared to the 46 measured values, although there was an increase in bias (−0.44 vs. −1.13 ppm). The overall accuracy as measured by the mean squared combination of bias and precision was better for the reduced model (3.8 ppm vs. 5.4 ppm).

As shown in Table V, when the model was compared to the predictions of the panel of industrial hygienists, the model resulted in less bias than 9 out of 11 industrial hygienists and resulted in better precision than all 11. When the estimates for all 11 hygienists were averaged, the model once again showed less bias and more precision than the panel.

After the results of the validation procedure indicated a reliable model, it was used to predict annual exposures for each worker in the epidemiologic study. For those working in the warehouse, predicted exposures appeared to be relatively high

TABLE IV. Comparison of Predictions of the Model to 46 Ethylene Oxide (ETO) Measurements at Six Plants Between 1979 and 1985

	Measured ETO Level (ppm)[a]	Model 1[b]	Model 2[c]
Geometric mean	2.22	1.18	1.50
Geometric SD	3.80	6.36	5.09
Arithmetic mean	4.62	4.19	3.50
SD	5.76	6.64	3.79
Range	0.1–32.0	0.05–26.0	0.05–15.7
Bias (ppm)	—	−0.44	−1.13
Precision (ppm)	—	5.40	3.66

[a]Based on 46 measurements not used to develop the model.
[b]Model with interaction terms for exposure category by product type and exposure category by year.
[c]Model 1 with interaction terms removed.
SD = standard deviation.

TABLE V. Comparison of Predictions of the Model With a Panel of 11 Expert Industrial Hygienists (IH) for 46 Sampling Conditions in Six Excluded Plants

	Arithmetic mean (ppm)	Bias (ppm)[a]	Precision (ppm)[b]
IH1	9.71	+4.39	11.55
IH2	1.44	−3.88	5.93
IH3	7.60	+2.28	6.67
IH4	2.23	−3.10	5.41
IH5	4.09	−1.24	4.68
IH6	2.43	−2.89	4.86
IH7	1.63	−3.69	6.13
IH8	2.61	−2.71	5.03
IH9	3.35	−1.97	4.63
IH10	4.87	−0.45	5.14
IH11	4.26	−1.07	5.85
IH AVG	4.02	−1.30	4.54
Model	3.50	−1.13	3.66

[a]Bias = $\sum_{i=1}^{46} d_i/46 = \Sigma\,(P_i - M_i)/46$,
where M_i = measured ETO on i^{th} sample and P_i = ETO level predicted by IH or the model for i^{th} sample.
[b]Precision = standard deviation of d_i.

compared to measured ETO levels under certain exposure conditions. For this reason, the original 205 data values used in developing the model were closely reviewed. It was then determined that eight values previously coded as "office" or "supervisor" should actually have been assigned to warehouse locations. The data were corrected and the model refit.

The fit of the model, as measured by R^2 and the degree of bias and precision when compared to the validation data set, remained essentially unchanged after this correction. This illustrates that examination of model predictions in all cells of the job exposure matrix is very important. This finding also required that all data regarding

TABLE VI. Using Multiple Linear Regression Model: An Example

Exposure category	=	Sterilizer operator (category #7)
Product type	=	Gauze (product #3)
Age	=	Freshly sterilized (age = 1)
Year	=	1980
Rear exhaust	=	Yes
Aeration	=	No
Cubic feet	=	2,800
ℓn ETO	=	−0.946 − 0.289 AERATION − 0.181 EXP1 − 0.88 EXP2 − 0.188 EXP3 − 0.606 EXP4 − 0.207 EXP5 − 0.087 EXP6 + 0.292 EXP7 + 0.279 PROD1 + 0.939 PROD2 + 0.688 PROD3 + 2.059 PROD4 − 0.233 (AGE − 4) − 0.446 (YEAR − 82) − 0.062 (YEAR − 82)^2 − 0.624 EXHAUST + 0.114 (CUBICFT − 1000)/100 − 0.0021 ((CUBICFT − 1000)/100)^2
ℓn ETO	=	−0.946 − 0.289 (0) + 0.292 (1) + 0.688 (1) − 0.233 (1−4) − 0.446 (80−82) − 0.062 (80−82)^2 − 0.624 (1) + 0.114 (2800−1000)/100 − 0.0021 ((2800−1000)/100)^2
ETO	=	exp (2.12) = 8.4 ppm

job classifications and locations within the plants be carefully reviewed before use of the model for exposure estimation in the mortality study. When all corrections were made, ETO values for workers in the mortality study seemed to be within a reasonable range of possible levels. Table VI gives an example of how the model was used to estimate annual exposure levels for a typical worker in the study.

DISCUSSION

Developing a statistical model for prediction of historical exposures is considerably more involved than the usual regression analysis, where one is often primarily interested in hypothesis testing. For this reason, significance levels are not as important for prediction models as the stability of the model and its ability to explain a substantial amount of variability in the exposure measurement data. Careful attention must be given to the availability of the predictor variables for all conditions encountered in the epidemiologic study for which exposures are being predicted.

Once the candidate variables have been determined, the model must be developed in such a way that terms are not introduced which substantially alter the coefficients of other terms in the model. This is commonly referred to as model stability. After an acceptable model has been determined using available data, it must be validated using an independent source. In this study, a set of data not used in developing the model was used to check the reliability of the model's predictions. As an added precaution, the model was compared to the estimates made by an expert panel of industrial hygienists.

Since the model resulted in less bias and higher precision than the average prediction of the panel, we felt confident that its use was an objective and reliable tool in reconstructing past exposures. Although the reduction in bias and precision compared to the IH panel is not great, an additional benefit of using a prediction model was that exposure estimation could be automated by use of computer software to assign exposure levels to individual workers in the job-exposure matrix.

There are, however, a few limitations to this approach to exposure assessment. First, a reasonably large set of industrial hygiene measurements sampled under a variety of exposure conditions must be available. These measurements must have been taken both before and after the introduction of any significant changes in exposure factors, for example, introduction of an engineering control. This permits use of the model to quantify the percent change in exposure levels attributed to any engineering control identified in the model development as causing a significant reduction.

Second, information on all independent variables (exposure factors) used in the prediction model must be available for the entire period of the epidemiologic study. For example, it is of no use to develop a model which includes production rates as an independent variable if data on such rates are not available during the early years of the retrospective study. Third, information on job titles or departments must be reduced to a manageable number (say, 10–12 or less). If thousands of such categories exist and the industrial hygienists are unable to pool them, the modeling approach is of limited value. If data exist on most job titles, statistical methods such as cluster analysis might be used to pool the data into a reduced number of job categories. The use of statistical methods for pooling also has an additional advantage of adding more objectivity. The data in this study were not extensive enough to permit a cluster analysis. Accordingly, while the statistical modeling provides a more objective approach, it still depends to some degree upon the subjective determination of exposure categories.

Fourth, the method of evaluating the model presented here regards the IH measurements in the evaluation data set as the gold standard. Obviously, these measurements are subject to sampling variability and therefore do not represent "truth." However, we know of no other practical way to evaluate the modeling approach or any other quantitative exposure assessment method. One possibility to reduce sampling error in the evaluation data set is to use only exposure levels based on means of some minimum number of samples. We decided not to adopt that approach in this study since a meaningful minimum number of samples for each exposure condition (e.g., five or more) would have substantially reduced our evaluation data set. However, if larger sets of industrial hygiene measurements are available, basing the evaluation only on means of a selected minimum number of samples would enhance reliability of the final model.

Finally, although we feel that this model produces relatively accurate estimates of ETO exposure levels for the NIOSH epidemiologic study, there is a broader problem which we could not address. The accuracy of this model (or any other method of exposure assessment relying on IH measurements) depends heavily upon the representativeness of the measured data. If the industrial hygienists who collected the original data used a sampling strategy weighted toward identifying overexposure problems, exposure estimates will probably be biased on the high side. This, in turn, would bias risk estimates based on such data toward the null. Whenever possible, we eliminated sample results from consideration when it was noted that sampling was performed under unusual conditions, for example, a spill. The assumption we, or any other exposure assessors, make when using measured data to generate exposure estimates is that the set of samples used is representative of conditions typically encountered by members of the study population. The extent to which such sets of monitoring results depart from this assumption is in need of additional study.

ACKNOWLEDGMENTS

The authors thank the clerical staff at NIOSH who did an outstanding job of coding, abstracting, and manuscript preparation. We also thank the companies and unions involved in this study for helpful comments.

REFERENCES

Dement JM, Harris RL, Symons MJ, Shy CM (1983): Exposures and mortality among chrysotile asbestos workers. Part I: Exposure estimates. Am J Ind Med 4:399–419.

Eisen EA, Smith TJ, Wegman DH, et al. (1984): Estimation of long-term dust exposures in Vermont granite sheds. Am Ind Hyg Assoc J 45:89–94.

Garry VF, Hozier J, Jacobs D, Wade RL, Gray DG (1979): Ethylene oxide: Evidence of human chromosomal effects. Environ Mutagen 1:375–382.

Griefe AL, Hornung RW, Stayner LG, Steenland NK (1988): Development of a model for use in estimating exposure to ethylene oxide in a retrospective cohort mortality study. Scand J Work Environ Health 14:29–30.

Hasabelnaby N, Ware JH, Fuller WA (1989): Indoor air pollution and pulmonary performance: Investigating errors in exposure assessment. Stat Med 8:1106–1126.

Hogstedt C, Malmqvist N, Wadman B (1979): Leukemia in workers exposed to ethylene oxide. JAMA 241:1132–1133.

Jaylock MA (1988): Assessment of inhalation exposure potential from vapors in the workplace. Am Ind Hyg Assoc J 49:380–385.

Kromhout H, Heederik Y, Boley D (1987): Agreement between qualitative exposure estimates and quantitative exposure estimates. Am J Ind Med 12:551–562.

Leidel NA, Busch K, Lynch J (1977): "Occupational Exposure Sampling Strategy Manual." Washington, D.C.: National Institute for Occupational Safety and Health (DHEW/NIOSH publication 77–173), pp 75–76.

Ostro BD (1987): Air pollution and morbidity revisited: A specification test. J Environ Econ Manage 14:87–98.

Picard RR, Berk KN (1990): Data splitting. Am Stat 44:140–47.

Schwartz J (1991): Particulate air pollution and daily mortality in Detroit. Environ Res 56:204–213.

Steenland NK, Stayner LG, Griefe AL (1987): Assessing the feasibility of retrospective cohort studies. Am J Ind Med 12:419–30.

Steenland NK, Stayner LG, Griefe AL, Halperin W, Hayes R, Hornung RW, Nowlin S (1991): Mortality among workers exposed to ethylene oxide. N Engl J Med 324:1402–1407.

Stewart PA, Herrick RF (1991): Issues in performing retrospective exposure assessment. Appl Occup Environ Hyg 6:421–427.

Woskie SR, Smith TJ, Hammond SK, Schenker MB, Garshick E, Speizer FE (1988): Estimation of the diesel exhaust exposures of railroad workers: I. Current exposures. Am J Ind Med 13:381–394.

Yu RC, Tan WY, Mathew RM, Andjelkovich DA, Levine RJ (1990): A deterministic mathematical model for quantitative estimation of historical exposure. Am Ind Hyg Assoc J 51:194–201.