Statistical Models of Exposure Determinants

9

By
M. Abbas Virji, ScD, CIH
Thomas W. Armstrong PhD, CIH

9.0 Introduction

Exposure monitoring has traditionally focused on collecting full-shift time-weighted average (TWA) exposure measurements with the primary objective of demonstrating compliance with regulatory limits. Sampling strategies have relied on identifying and monitoring the highest exposed workers rather than on random monitoring of workers or groups of workers. Additionally, very little contextual information has been collected during monitoring on characteristics of the processes, work environment, control measures, environmental conditions, job tasks and worker activities. For example, the National Institute for Occupational Safety and Health (NIOSH) Occupational Exposure Sampling Strategies Manual identifies specific information on sampling and analysis to be recorded, but provides only minimal guidance on collecting information on work environment characteristics.(1) Resulting databases, such as the Occupational Safety and Health and Administration Integrated Management Information System (OSHA IMIS), consist of little descriptive information beyond job, industry, year of inspection, company size, and other higher order information.(2) As a consequence, such databases serve limited objectives beyond their intended purposes, despite the potential and need for multiple uses. Nevertheless, the IMIS data have been used successfully to illustrate trends in exposure over time or among industry sectors.(3) In practice, even when some contextual information is gathered during monitoring, it is generally in the form of notes and is not systematically collected, hence limiting its utility. The contemporary practice of industrial hygiene is moving away from compliance-based monitoring toward comprehensive exposure assessment.(4–6) There is also the recognition that exposure modeling is an integral part of exposure assessment(7); both monitoring and modeling are required to improve the efficiency of a sampling strategy and to achieve a more complete exposure assessment.(8)

This chapter focuses on statistical models applied to measured exposure data to evaluate the underlying nature of exposures and causes of variability. It describes the mathematical basis for these models and provides some applied examples. Citations within the text do not represent a comprehensive review of the literature. The reader is referred to additional useful references for a thorough and complete treatment of the topic.

9.1 Statistical Models

In industrial hygiene, models are used to describe the relationships between exposure levels and factors affecting/determining exposures via deterministic or statistical model designs. Deterministic models are based on principles of physics and chemical properties of the exposure substance and can range from simple vapor pressure models to more complex models characterizing the generation and transport of exposure.(9) Many deterministic models, based on first principles such as the conservation of mass, energy, or momentum, are discussed in other chapters of this book. Deterministic models are focused on underlying mechanisms involved in the exposure paradigm, including the initial process of contaminant generation to the ultimate step of estimating breathing zone exposures.

Other models arise from using observational and experimental data. In exposure assessment, these empirical models are developed to examine and estimate the relationship between a dependent exposure variable and a set of independent predictor variables related to characteristics of the work environment. These models rely on statistical techniques, using exposure measurements combined with either quantitative (continuous) or qualitative (categorical) data on the work environment collected simultaneously with the monitoring data. Therefore, statistical modeling of exposure determinants requires the collection of both exposure measurements and detailed contextual information on the work environment. One of the impediments of statistical modeling using existing or historical exposure data are the lack of systematically gathered contextual information on the work environment.

In exposure assessment models, the dependent variable is usually an exposure metric such as the full-shift TWA exposure or short-duration task exposure. Independent variables represent one or more factors affecting exposure or surrogates for such factors and may or may not be directly related to the physical and chemical factors used in deterministic models. More often, independent variables include a range of factors that describe exposure-generating processes, the transport of exposure and worker interface with exposure.[10] Hence, the model is often a mathematical expression translating observable or measurable information likely to affect exposures and quantifying the impact of that information on exposure levels. A diversity of determinants-of-exposure models have been published in the literature, making it difficult to summarize and understand exposure factors, thereby underscoring the need for a structured approach to studying exposure determinants. Tielemans et al.,[11] have proposed a conceptual model that describes the processes leading to exposure and suggest a set of broadly defined exposure modifying factors within which specific determinants can be identified and used in either deterministic or statistical models.

A number of statistical techniques have been used for modeling exposure determinants including: 1) linear regression; 2) random- and mixed-effects analysis of variance (ANOVA); 3) logistic regression; and 4) other regression techniques (e.g., Tobit regression). These models are briefly described in the next sections.

9.1.1 Linear Regression Models

Linear regression is typically used to study the determinants of exposure when modeling continuous exposure measurements that are combined with information on characteristics of the work environment. By identifying factors of the work environment that contribute most to exposures, these models can assist in prioritizing interventions for controlling workplace exposures. Multiple linear regression modeling is the most commonly used technique to investigate determinants of exposure, and is of the general form:

$$\text{Ln}\,[C] = \beta_0 + \beta_1\,(\text{Var}_1) + \beta_2\,(\text{Var}_2) + \ldots + \beta_n\,(\text{Var}_n) + E \qquad (9\text{-}1)$$

where

Ln [C] = natural logarithm of the exposure concentration, since most occupational exposure data have a more lognormal than normal distribution,
β_0 = intercept,
β_1–β_n = regression coefficients of the independent predictor variables 1 to n,
Var_1–Var_n = possible independent exposure predictors or modifiers, and
E = error term.[12]

The intercept represents the overall mean exposure level when all independent variables are equal to zero; it is also often considered an estimate of the background exposure level. Regression coefficients represent the change in exposure levels associated with each one-unit change in the independent variables. Dichotomous independent variables are modeled as (0/1), while polychotomous variables are transformed into dummy dichotomous variables (0/1), selecting one level as the reference level. The error term in the statistical model represents the difference between the predicted and observed exposure measurements, but is not used in the mathematical expression to predict exposure levels. However, the variance associated with the error term is used to construct confidence intervals for predictions. Kleinbaum et al.[12] describe in detail the assumptions associated with using multiple linear regression models, and include:

1) Ln [C] is a random variable dependent on Var_i values,
2) Values of Ln [C] are statistically independent,
3) The mean value of Ln [C] is a linear function of the Var_i values, though product terms can be included that allow for interactions,
4) The variance of Ln [C] is the same for any Var_i, and
5) For any value of Var_i, Ln [C] is normally distributed.

It is noteworthy that the linear regression model is not appropriate for repeat measurements on the same worker or when the data are highly correlated within a job group or worksite. A mixed model (described in the next section) is required for such data. In addition, measurement error in the explanatory variables (Var_i) can lead to biased estimates of the regression coefficients and thus it is important to attempt to understand the sources of measurement error in the explanatory variables. A summary of linear regression models used for studying the determinants of exposure, including examples, is contained in texts by Boleij et al.[13] and Burdorf.[14]

9.1.2 Mixed Models

A large number of published studies continue to show considerable variation in exposure between workers performing the same job, as well as differences in exposure levels from day-to-day.[15] Random effects ANOVA models have been used to model exposure data with repeated measurements on workers to obtain estimates of within- and between-worker components of exposure variability.[13] Understanding of variance components can have important implications for developing strategies for sampling, exposure control, and epidemiological grouping. The more recent use of mixed models permits the identification of important factors affecting exposures, while providing additional information on within- and between-worker variability.[6] The natural logarithm of exposure is modeled as the dependent variable, with person effects declared as random and workplace characteristics as fixed effects. The model is of the general form:

$$Y_{ij} = Ln(X_{ij}) = \mu + \beta_{1-k} + X_i + \varepsilon_{ij} \quad (9\text{-}2)$$

where

Y_{ij} = natural logarithm of the exposure of the i^{th} person on the j^{th} day,
μ = overall mean of log-transformed exposure,
β_1–β_k = fixed effects of exposure determinants,
X_i = random effect of the i^{th} worker, and
ε_{ij} = random error.

Parameter estimates of the significant fixed effects (β_{1-k}) provide estimates of the magnitude of the change in average exposure associated with each factor (exposure determinants). From the model, estimates of the within-worker ($_{ww}S^2$) and between-worker ($_{bw}S^2$) variances are obtained. Assumptions of the mixed models are described in detail by Kleinbaum et al.[12] and Rappaport and Kupper.[6] Mixed models are now commonly used to understand and determine factors affecting exposure and can include any of the explanatory variables included in the regression model (9-1). In addition, mixed models can account for correlation between repeated measurements on the same worker and among samples collected from the same work area of facility. Moreover, mixed models can accommodate a grouping variable to evaluate whether workers in different groups have different exposures and different variance components. Information on variance components can be used to identify types of control measures that are most beneficial, as well as to determine exceedance probabilities and average exposure for individual workers.[6] A summary of mixed models, including examples, is also provided by Burdorf.[14]

9.1.3 Logistic Regression Models

Modeling using dichotomous exposure data (i.e., yes/no) obtained from survey questionnaires, exposure measurements above/below the limit of detection (LOD), or exposure measurements above/below some identified threshold value (e.g., PEL) requires special consideration. Logistic regression models have been used to identify factors affecting exposure from severely censored (>50% LOD) datasets[16] and from data exceeding regulatory limits.[17] Logistic regression models can be considered a special case of linear regression for dichotomous dependent exposure variables. The exposure measurements or scores are dichotomized (1/0) to mark an exposure event. The unconditional logistic regression model is of the general form[12]:

$$\text{Logit}\ (\text{pr}\ Y{=}1) = \alpha + \beta_1\ (Var_1) + \beta_2\ (Var_2) + \ldots + \beta_n\ (Var_n) + E \quad (9\text{-}3)$$

The Logit (pr Y=1) is Ln (Pr Y=1/Pr Y=0). The logit model above describes the ratio of the probabilities for and against an exposure event (the outcome variable) as a function of:

a = constant term,
β_1–β_n = parameter estimates for the exposure factors (Var_1–Var_n),
Var_1–Var_n = possible independent exposure predictors or modifiers, and
E = error term.

From the model described above, odds ratios and confidence intervals for exposure determinants are available. If the probability of an exposure event is high (not rare), then the odds ratios from the logistic model provides biased estimates of the relative risk (in this case of exposure event).[18] Methods have been developed to convert the odds ratio to prevalence ratio[18] or to directly estimate prevalence ratio from a Cox regression model.[19] For repeated measurements on the binary outcome variable, general estimating equations can be used to model exposure determinants.

9.1.4 Other Regression Models

Modeling highly censored exposure data requires special consideration to adjust for truncation of the exposure distribution. Modeling truncated continuous exposure measurements can be achieved through Tobit regression, which uses the maximum likelihood method to estimate parameters (i.e., exposure determinants). The model uses log-transformed exposure measurements as the outcome variable with the assumption that the data below the LOD follow the same distribution as the observed data. The log-likelihood function has two components, one for observed data and the other for data below the LOD.[20] Maximum likelihood estimates of β_1 through β_n are then obtained by maximizing the log likelihood function. Woskie et al.[16] used this approach to model determinants of isocyanate exposures in automobile body repair shops. Methods have also been developed to model highly censored data with repeated measurements using mixed models.[21]

9.2 Model Development Strategies

Many studies have applied the principles of multiplicative models, including the following reviews and examples.[22–25] A review by Burstyn and Teschke[10] illustrates the range of approaches and methods used to develop exposure determinants models. The authors also show that a large number of predictor variables have been used in these models that are often specific to particular workplace circumstances. Hence, a universal model of exposure determinants is not feasible, and would not yield useful results. An important initial step before modeling is to develop a conceptual framework of how various factors may impact exposure levels and the inter-relationships among factors.

The following steps are provided as a framework for readers to develop an exposure model based on specific workplace determinants. Statistical model development is well described by Kleinbaum et al.[12]

9.2.1 Develop a Conceptual Model

Starting with a conceptual model, such as the source-receptor model, allows for a more consistent approach to collecting information on potential classes of exposure determinants. A source-receptor model describes the physical pathway of exposure from its source through different compartments and mechanisms to the route of entry into the body. Figure 9.1 illustrates the three components of a source-receptor model and provides some examples of general factors that may affect exposure (adapted from Armstrong et al.[23]); note that specific operations factors may include additional sub-factors to consider. Additionally, there are higher order factors related to workplace, industry, and calendar time that may also directly or indirectly impact determinants of exposure, thus affecting exposure levels.

Developing a conceptual model requires an understanding and knowledge of: processes, raw materials, by-products and end-products; mechanisms and factors affecting exposure generation; mechanisms and factors affecting transport of exposure; individual activities and opportunities for direct or indirect contact with exposure; and routes of entry and associated factors. A joint ACGIH®-AIHA task group on occupational exposure databases issued a report that identified over 100 data elements to be collected with exposure data.[26] These data elements are exposure determinants that can serve as a starting point toward building the conceptual model. Tielemans et al.[11] have recently proposed a conceptual model for estimating inhalation exposures based on the source-receptor model that describes the source, various transport compartments and mechanisms, and the receptor. A set of nine broadly defined, mutually independent modifying factors are outlined, chosen according to first principles. Each of the modifying factors can include a series of additional factors that could be considered as exposure determinants, and that can be applied to either deterministic or statistical models. Such a structured approach permits for a consistent consideration of compartments and mechanisms of exposures, and at the same time allows for more specific exposure determinants to be developed. It also provides a structured framework for characterizing the workplace environment and collecting contextual information on the work environment characteristics.

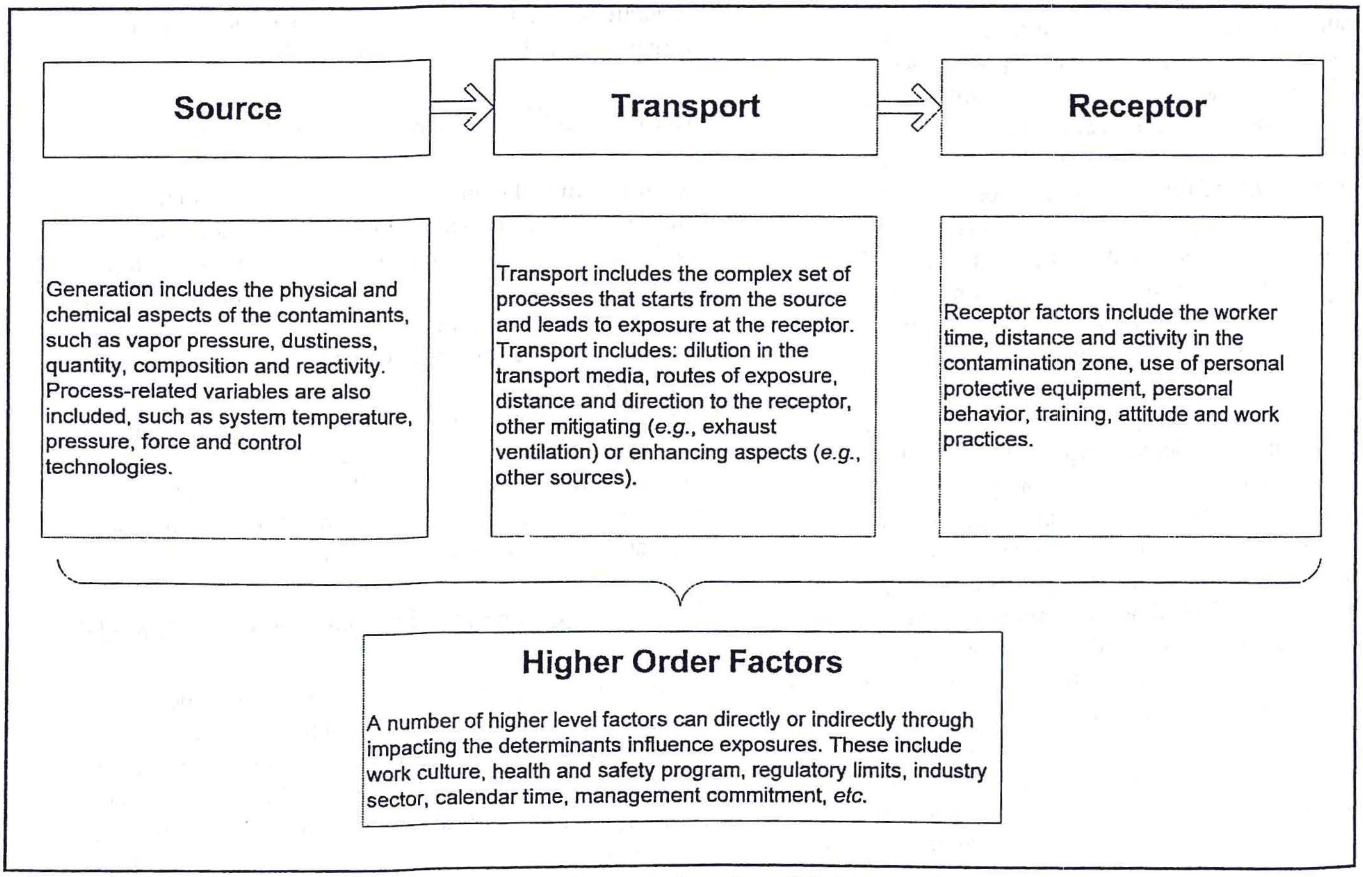


Figure 9.1 — Source-receptor model components with examples of possible exposure determinants.

9.2.2 Collect Information on Exposure Determinants

A prerequisite of statistical modeling is the availability of information on exposure determinants describing different aspects of the work environment. This information is generally collected via questionnaires, worker diaries or direct observation by exposure assessors during exposure monitoring. A conceptual model and initial list of exposure determinants are generally constructed after an initial assessment that includes a walk-through of the facility and a review of the processes, materials, etc., as described in the AIHA strategy. Much has been covered on individual versus group measurements, with well-formed and evaluated Similar Exposure Groups (SEGs) widely used for efficient workplace evaluations. Exposure monitoring records for SEGs may be particularly informative if the underlying exposure determinants are documented or can be reconstructed. Details on exposure assessment strategies and construction or validation of SEGs is available elsewhere.[4,5] Additional information on the integration of professional judgment of industrial hygienists with the available exposure data are provide elsewhere in this book. Further information on the assembly of sources for and methods of collecting exposure determinant for exposure reconstruction is available from Viet et al.[27]

9.2.3 Evaluate Exposure Data

Exposure monitoring is conducted for a variety reasons such as compliance assessment or exposure evaluation using different strategies, which may result in variable data quality, potential biases and limitations to data interpretation. For example, Lavoué et al.[28] investigated the determinants of formaldehyde exposures in the reconstituted wood panel industry in Québec by combining two exposure datasets collected using different sampling strategies. Results indicated significant differences in the estimated yearly geometric means for jobs and areas between exposure data collected by the government as compared to data collected by researchers. Likewise, in another study, Lavoué et al.[3] reported some differences in formaldehyde exposures in the OSHA IMIS dataset between data collected for programmed versus non-programmed inspections. Aspects of an exposure dataset that may impact its utility in exposure assessment include but are not limited to the purpose of sampling; the sampling strategy including the manner in which samples were collected, whether specific jobs or work sites were intentionally included or excluded, and details of sampling and analysis; and the extent of contextual information recorded or extractable from notes or other information sources. Depending on the specific purpose for modeling exposure determinants, these factors can impact or limit the utility of exposure datasets in achieving the desired objectives. Tielemans et al.[29] proposed guidelines for evaluating the quality of exposure data to assess the relative value of data for risk assessment. Such a procedure ensures systematic evaluation of exposure datasets and identifies biases or limitations. Furthermore, such a procedure ensures consistency among assessors as well as in pooling or aggregating exposure datasets from different sources. Several researches have provided details of the evaluation of data quality and the development of datasets used in their studies.[30–33] Before performing exposure modeling, it is critical to understand data quality as well as the representativeness of the data to enhance interpretation of modeling results.

9.2.4 Evaluate Exposure Determinants

After evaluating the exposure distribution and identifying the appropriate models to use, exposure determinants are investigated in univariate models to identify factors that significantly affect exposure. Factors that are significant in explaining the variability in exposure are further evaluated to investigate their correlations and to develop a strategy to handle correlated variables. Several approaches can be used to handle correlated variables including selecting only a single variable that is the easiest to interpret or that has the greatest impact on exposure; developing separate models for each of the correlated variables; categorizing continuous variables; and creating new variables by combining the correlated ones.[10,34] Principle components analysis or latent variable analysis can also be used to generate factors that can then be used as independent variables in multiple regression models.

9.2.5 Build Multiple Regression Models

A number of techniques have been used to build the multiple regression models, including forward selection, backward elimination, or step-wise process with different criteria for model entry significance level, and significance level to remain in the model. The statistic C_p is also a useful means of comparing regression models with many explanatory variables. Details of model building and variable selection criteria are described by Kleinbaum et al.[12] The following studies serve as examples of model building strategies.[34,35] Generally, automatic model building is not recommended. Some of these techniques have application to mixed models as well.

9.2.6 Use Standard Techniques to Evaluate Models

An integral component of model building is model evaluation. Standard regression diagnostic techniques to evaluate the model have been described by Kleinbaum et al.[12] and include assessing goodness of fit of the model, assessing distributional assumptions, examining residuals and standardized residuals, and detecting outliers and points of influence, including sensitivity analysis.

9.3 Model Validation

Several problems can be associated with predictive models that cannot be detected using standard regression diagnostics identified above. These problems can include variable selection, over-fitting the model to the data, estimation of the parameters, and prediction error.[36,37] Hence, validation using data not used in model building is necessary. However, such data are often unavailable and alternative model evaluation strategies using bootstrap or jackknife methods can be used. Three options are briefly presented.

9.3.1 Internal Validation by Splitting Data

When large datasets are available, exposure data can be randomly split into model building and model validation datasets. The predictions from the final model in the model building dataset are then compared to the exposure measurements in the validation dataset. Hornung et al.[35] provide an example of data splitting for validation. Several measures can be used to assess the degree of agreement between the model estimations and observed exposure measurements. These measures include bias, precision and accuracy.[35] Additional measures of agreement that provide information on the degree of concordance and sources of disagreement between two measurements include the concordance correlation coefficient and the intra-class correlation coefficient; a summary of these approaches is described in Virji et al.[38]

9.3.2 External Validation using Different Data Source

The model-generated exposure estimates can be compared to external data sources such as comparison to additional exposure data from the same or different workplaces within the same industry; comparison to data reported in the literature; comparison to exposures predicted by deterministic models; and/or comparison to expert assessment of exposure. An example from the published literature includes a study of formaldehyde exposures among embalmers that used measurements from other surveys for validation.[39]

9.3.3 Evaluation Using Monte Carlo Methods

The traditional internal and external validation approaches require large amounts of data and are not immune to errors associated with over-fitting the model to the data. However, computer-intensive methods, referred to as re-sampling methods, are available to address these issues. These methods are well described in a text entitled "Randomization and Monte Carlo Methods in Biology."[40] The bootstrapping technique has consistently yielded optimum results and is one of several computer-intensive approaches used to develop and evaluate models; it is ideal for small sample sizes.[41] This approach is based on sampling with replacement from the original dataset of measurements. The original sample is treated as a population from which typically 1000 datasets are drawn with replacement, each the same sample size as the original data. The 1000 datasets can then be used to develop predictive models and identify the important factors affecting exposures as well as to estimate parameters.[41,42] The results from each of the 1000 datasets are combined to generate a distribution of parameter estimates and their confidence intervals. Recently, Hein et al.[33] used a combination of the split data validation and bootstrap methods to validate estimates of benzene, toluene and xylene exposures from statistical models. For a selected number of operations, they drew 1000 datasets each of which was split into model building and validation subsets. The predicted exposures from the model-building subset were compared to the observed data in the validation subset via correlation coefficients, and summarized over the 1000 datasets to obtain their distributions.

9.4 Examples of Application

The objective of modeling in industrial hygiene is to investigate and describe the relationship between exposure levels and workplace factors to identify the sources of exposure variation. In workplace environments, exposures tend to vary significantly from day-to-day, location-to-location, job-to-job, and between workers due to differences in a host of factors including workplace conditions, environmental factors, and personal work practices. Statistical models have been developed to identify factors that are most influential in determining exposures, thereby providing opportunities to identify and prioritize control measures, to estimate current exposures for un-sampled workers or historical exposures of individuals, or to assess time trends in exposures. Such statistical models have previously been applied to a variety of industrial and environmental situations. While an exhaustive literature review is not the focus of this chapter, readers are urged to conduct a literature search using the phrase "exposure determinants" to access more studies. Several examples are discussed in this section.

9.4.1 Identifying Exposure Factors for Control Measures

A study was conducted to examine the determinants of bulk and airborne microbial exposures, and to assess the impact of changes in exposure determinants on bulk and airborne microbial levels.[34] These investigators used modeling to identify factors affecting exposure and to develop effective exposure control strategies. They developed three separate models to evaluate the effects of environmental, process and personal factors on airborne viable microbial levels, and three additional models to evaluate the effects of fluid characteristics, process and environmental factors on bulk material viable microbial levels. The determinants evaluated were based on a conceptual framework of exposure to metal working fluids (Figure 9.2).

The final full model for airborne microbial levels accounted for 24% of the total variability, whereas the final full model for bulk material microbial levels captured 45% of the total variability. To evaluate the impact of each independent factor on exposure levels, the value of one variable was changed at a time from its minimum to maximum value while holding the remainder of the variables at the median value (Figure 9.3). The airborne model showed that (1) decreasing bulk microbial concentrations, machine density and production rate, (2) increasing worker distance, and (3) use of enclosures were all associated with reducing airborne microbial levels (Figure 9.3a). The bulk material microbial model showed that increasing pH and decreasing machine RPM, percent tramp oil, and sump size of the fluid were all associated with decreasing bulk material microbial contamination levels (Figure 9.3b). Results from these models can be used to estimate potential reduction in worker exposure or fluid contamination due to introduction of certain control measures.

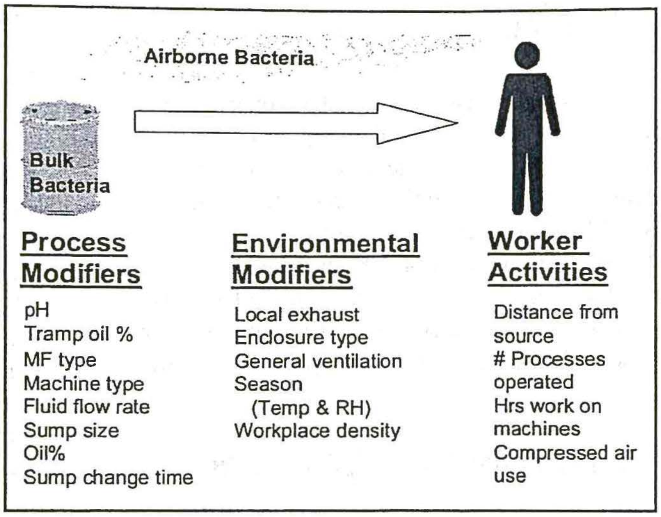


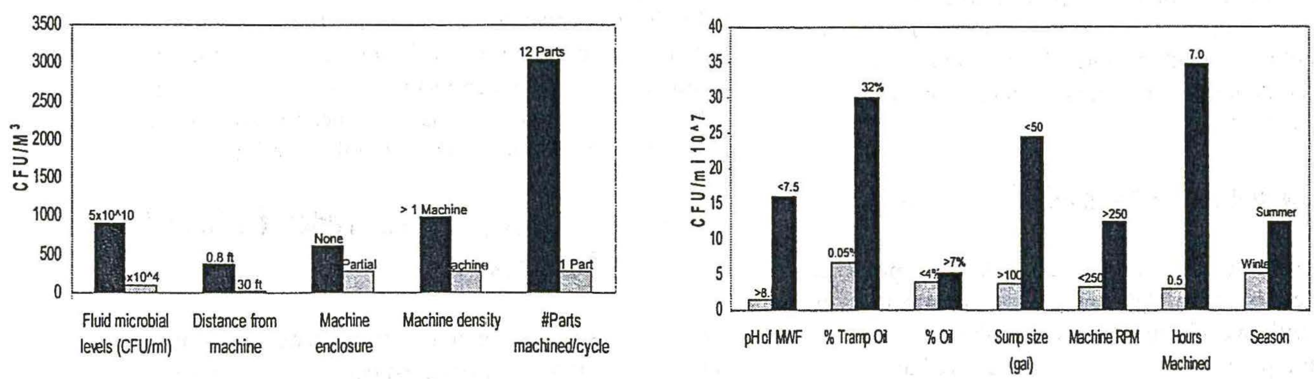
Figure 9.2 — Illustration of metal working fluid exposure model.

9.4.2 Predicting Exposures

A weighted linear regression model was developed as part of an epidemiologic study to estimate historical worker exposure to ethylene oxide (EtO) in the sterilizing industry.[35] The model considered independent variables identified as possible exposure determinants based on walkthrough surveys of the operations by experienced investigators. The individual 8-hr TWA exposure data were grouped by plant, year, and type of sampling media for each job and location, resulting in 205 average measurements. A multiple regression model was developed that included the significant effects of exposure category, product type, size of the sterilization unit, selected engineering controls, days after sterilization and calendar year. The model explained a remarkable 85% of the variability and showed a steady decline in exposures over time. The model predicted EtO exposures were validated based on a fraction of the data not used in model building (validation by splitting data). Comparing the predicted exposures with measured data showed a remarkably small bias of 1.1 ppm. The model was used to predict historical EtO exposures for workers based on the combinations of the independent variables in the final model.

9.4.3 Evaluating Time Trends

Among the assumptions of retrospective exposure assessment is that exposure levels change over time due to



Microbial Airborne Levels (a)

Bulk Material Microbial Levels (b)

Figure 9.3 — Evaluating the impact of changes in exposure determinants on microbial levels.

factors that include changes in processes, ventilation controls, and regulations among others. Therefore, to more accurately estimate past exposures, current exposure estimates need to be adjusted for such changing factors. A number of studies have quantitatively documented time trends in exposures using statistical models.[43–45] These studies have not only reported trends in the reduction of exposures over time, but have also demonstrated the need to understand factors affecting exposure. For example, Burstyn et al. (2000)[45] constructed a database of bitumen fume and vapor and benzo(a)pyrene exposures among asphalt workers from numerous sources and several countries. Using a mixed-effects model, they showed a steady decline in exposure levels of 6–14% per year for the three exposure agents over a 20-year period between 1970 and 1990. Their statistical models explained 36–43% of the total variability for the three exposure agents and included significant effects of various production factors and asphalt temperature.

Several other studies have reported such declines in exposure levels over time, including decline of 12% per year in toluene exposure among commercial painters.[46] Symanski et al.[44] reported declining total nickel exposures of 7 to 9% per year in nickel mining, smelting and refining sectors, but an increasing trend in exposures (4% per year) within the milling sector. Yu et al.[47] used data on formaldehyde exposures among foundry workers to point out the need for step-change considerations in deterministic exposure modeling, noting that changes in exposure can be sudden and may occur at a given point in time.

9.5 Summary

In this chapter, we have attempted to describe statistical models of exposure determinants and demonstrate their utility in understanding factors affecting exposure. These models generally require large quantities of exposure measurements along with contextual information on characteristics of the work environment that are collected during monitoring. Additionally, some of the exposure determinants models utilizing full-shift exposure measurements capture only a modest fraction of exposure variability and identify few specific activities or control measures.[48] Perhaps developing these models using exposure determinants from a conceptual model based on the first principle of mass conservation may improve the fraction of exposure variation explained by the models. Additionally, consideration of time-varying exposure determinants with real-time exposure measurement can identify specific activities and control measures and can have a significant impact on intervention strategies.[48]

Improvements in exposure assessment methodologies require both exposure monitoring and modeling, both statistical and deterministic. This chapter provides only a brief overview of techniques available for using exposure data to generate models to estimate historical exposures, to identify exposure factors contributing to increased or decreased exposures, and to develop a targeted control strategy. Consultation with statistical texts and software packages are recommended prior to proceeding. However, with this overview information and additional reading of studies in journals, the reader is encouraged to evaluate exposure databases to determine if exposure determinants can be evaluated. While this approach requires significantly more data than may exist in small worksites, multi-site industries or multi-industries with similar processes may aggregate exposure data and develop models to determine key workplace environmental factors that affect exposures. While a mixture of factors and branches of logic may be used in different situations, a conceptual framework can help to organize the many considerations. Efforts by regulators, researchers, and industry groups are encouraged to pursue data aggregation, with clearly and uniformly defined datasets including information on workplaces, materials, work methods, and environmental data that may contribute to task- and industry-specific workplace exposure determinant models. Using data to identify key contributors to exposure is critical to understanding and controlling exposure risk factors, and to providing evidence-based methods for prioritizing intervention strategies.

9.6 Acknowledgement

The authors would like to acknowledge Gregory Day, Susan Woskie, and Stanley Shulman for their review of the chapter and their thoughtful comments and suggestions which improved the quality of the chapter.

9.7 References

1. **Leidel, N. A., K.A. Busch, and J.R. Lynch:** *Occupational Exposure Sampling Strategy Manual.* U. S. Department of Health, Education, and Welfare. Public Health Service. Center for Disease Control. National Institute for Occupational Safety and Health. DHEW (NIOSH) Publication No 77-173. Cincinnati, OH: NIOSH, 1977.
2. **Gómez, M.R.:** Factors associated with exposure in Occupational Safety and Health Administration data. *Am. Ind. Hyg. Assoc. J. 58(3)*:186–195 (1997).
3. **Lavoué, J., R. Vincent, and M. Gérin:** Formaldehyde exposure in U.S. industries from OSHA air sampling data. *J. Occup. Environ. Hyg. 5(9)*:575–587 (2008).
4. **Mulhausen, J.R. and J. Damiano:** *A Strategy for Assessing and Managing Occupational Exposures*, 2nd edition. Fairfax, VA.: American Industrial Hygiene Association, 1998.

5. **Ignacio, J.S. and W.H. Bullock (eds.):** *A Strategy for Assessing and Managing Occupational Exposures*, 3rd edition. Fairfax, VA.: American Industrial Hygiene Association, 2006.
6. **Rappaport, S.M. and L.L. Kupper:** *Quantitative Exposure Assessment*. El Cerrito, CA: Stephen Rappaport, 2008.
7. **Jayjock, M.A., C.F. Chaisson, S. Arnold, and E.J. Dederick:** Modeling framework for human exposure assessment. *J. Expo. Sci. Environ. Epidemiol. 17(1)*:S81–S89 (2007).
8. **Schneider, T.:** Improving exposure assessment requires measurements and modeling. *Scand. J. Work Environ. Health. 28(6)*:367–369 (2002).
9. **Tan, Y., M.R. Flynn:** Experimental evaluation of a mathematical model for predicting transfer efficiency of a high volume/low pressure air spray gun. *Appl. Occup. Environ. Hyg. 15(10)*:785–793 (2000).
10. **Burstyn, I. and K. Teschke:** Studying the determinants of exposure: A review of methods. *Am. Ind. Hyg. Assoc. J. 60(1)*:57–72 (1999).
11. **Teilemans, E., et al.:** Conceptual model for assessment of inhalation exposure: Defining modifying factors. *Ann. Occup. Hyg. 52(7)*: 577–568 (2008).
12. **Kleinbaum, D.G., L.L. Kupper, K.E. Muller, and A. Nizam:** *Applied regression analysis and other multivariable methods*, Third Edition, Pacific Grove, CA: Duxbury Press, 1998.
13. **Boleij, J., E. Buringh, D. Heederik, and H. Kromhout:** *Occupational Hygiene of Chemical and Biological Agents*. Amsterdam, The Netherlands: Elsevier Sciences, 1995.
14. **Burdorf, A.:** Analysis and modeling of personal exposure. In *Exposure Assessment in Occupational and Environmental Epidemiology*. Nieuwenhuijsen, M.J. (ed.). New York, NY: Oxford University Press, Inc. (2003).
15. **Symanski, E., S. Maberti, and W. Chan:** A meta-analytic approach for characterizing the within-worker and between-worker sources of variation in occupational exposure. *Ann. Occup. Hyg. 50(4)*:343–357 (2006).
16. **Woskie, S.R., et al.:** Determinants of isocyanate exposures in auto body repair and refinishing shops. *Ann. Occup. Hyg. 48(5)*:393–403 (2004).
17. **LaMontagne, A.D. and K.T. Kelsey:** Evaluating OSHA's ethylene oxide standard: Exposure determinants in Massachusetts hospitals. *Am. J. Public Health. 91(3)*:412–417 (2001).
18. **Thompson, M.L., J.E. Myers, and D. Kriebel:** Prevalence odds ratio or prevalence ratio in the analysis of cross sectional data: What is to be done. *Occup. Environ. Med. 55*:272–277 (1998).
19. **Lee, J.:** Estimation of prevalence rate ratios from cross-sectional data: A reply. *Internatl. J. Epidemiol. 24(5)*:1066–1067 (1995).
20. **Lubin, J.H., et al.:** Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ. Health. Perspect. 112(17)*: 1691–1696 (2004).
21. **Thiébaut, R. and H. Jacqmin-Gadda:** Mixed models for longitudinal left-censored repeated measures. *Comput. Methods Programs Biomed. 74(3)*:255–260 (2004).
22. **Dodgson, J., J. Cherrie, S. Groat:** Estimates of past exposure to respirable man-made mineral fibres in the European insulation wool industry. *Ann. Occup. Hyg. 31(4B)*:567–582 (1987).
23. **Armstrong, T.W., et al.:** Retrospective benzene and total hydrocarbon exposure assessment for a petroleum marketing and distribution worker epidemiology study. *Am. Ind. Hyg. Assoc. J. 57(4)*:333–343 (1996).
24. **Lewis, S.J., et al.:** Retrospective estimation of exposure to benzene in a leukaemia case-control study of petroleum marketing and distribution workers in the United Kingdom. *Occup. Environmental Med. 54(3)*:167–175 (1997).
25. **Glass, D.C., G.G. Adams, R.W. Manuell, and J.A. Bisby:** Retrospective exposure assessment for benzene in the Australian petroleum industry. *Ann. Occup. Hyg. 44(4)*:301–320 (2000).
26. **American Conference of Governmental Industrial Hygienists® (ACGIH®):** Data elements for occupational exposure databases: Guidelines and recommendations for airborne hazards and noise. *Appl. Occup. Environ. Hyg. 11(11)*:1294–1311 (1996).
27. **Viet, S. M., M. R. Stenzel, C.P. Rennix, T.W. Armstrong, and J.R. Couch (eds.):** *Guideline on Occupational Exposure Reconstruction*. Fairfax, VA: American Industrial Hygiene Association, 2008.
28. **Lavoué, J., C. Beaudry, N. Goyer, G. Perrault, M. Gérin:** Investigation of determinants of past and current exposures to formaldehyde in the reconstituted wool panel industry in Quebec. *Ann. Occup. Hyg. 49(7)*:587–602 (2005).
29. **Tielemans, E., H. Marquart, J. De Cock, M. Groenewold, J. Van Hemmen:** A proposal for evaluation of exposure data. *Ann. Occup. Hyg. 46(3)*:287–297 (2002).
30. **Vincent, J.H. and M.A. Werner:** Critical evaluation of historical occupational aerosol exposure records: Applications to Nickel and Lead. *Ann. Occup. Hyg. 47(1)*:49–59 (2003).
31. **Burstyn, I., H. Kromhout, P.J. Cruise, and P. Brennan:** Designing an international industrial hygiene database of exposure among workers in the asphalt industry. *Ann. Occup. Hyg. 44(1)*:57–66 (2000).
32. **De Vocht, F., et al.:** A database of exposures in the rubber manufacturing industry: design and quality control. *Ann. Occup. Hyg. 49(8)*:691–701 (2005).
33. **Hein, M. J., M.A. Waters, E. van Wijngaarden, J.A. Deddens, and P.A. Stewart:** Issues when modeling benzene, toluene, and xylene exposures using a literature database. *J. Occup. Environ. Hyg. 5(1)*:36–47 (2008).

34. **Virji, M. A., S.R. Woskie, S. R. Sama, D. Kriebel, and D. Eberiel:** Identifying the determinants of viable microorganisms in the air and bulk metalworking fluids. *Am. Ind. Hyg. Assoc. J. 61(6)*:788–797 (2000).
35. **Hornung, R.W., et al.:** Statistical model for prediction of retrospective exposure to ethylene oxide in an occupational mortality study. *Am. J. Ind. Med. 25(6)*:825–836 (1994).
36. **Schumacher, M., N. Holländer, and W. Sauerbrei:** Resampling and cross-validation techniques: a tool to reduce bias caused by model building? *Stat. Med. 16(24)*:2813–2827 (1997).
37. **Steyerberg, E.W., et al.:** Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J. Clin. Epidemiol. 56(5)*:441–447 (2003).
38. **Virji, M.A., et al.:** Agreement between task-based estimates of the full-shift noise exposure and the full-shift noise dosimetry. *Ann. Occup. Hygiene.* [In press] (2009).
39. **Hornung, R.W., et al.:** An experimental design approach to retrospective exposure assessment. *Am. Ind. Hyg. Assoc. J. 57(3)*:251–56 (1996).
40. **Manly, B.F.J.:** *Randomization and Monte Carlo Methods in Biology*. New York, NY: Chapman and Hall, 1991.
41. **Jensen, R. L. and G.M. Kline:** The resampling cross-validation technique in exercise science: modelling rowing power. *Med. Sci. Sports. Exerc. 26(7)*: 929–933 (1994).
42. **Molinaro, A.M., R. Simon, and R.M. Pfeiffer:** Prediction error estimation: a comparison of resampling methods. *Bioinformatics. 21(15)*: 3301–3307 (2005).
43. **Vermeulen, R., J. de Hartog, P. Swuste, and H. Kromhout:** Trends in exposure to inhalable particulate and dermal contamination in the rubber manufacturing industry: effectiveness of control measures implemented over a nine-year period. *Ann. Occup. Hyg. 44(5)*:343–354 (2000).
44. **Symanski, E., C.C. Chang, W. Chan:** Long-term trends in exposures to nickel aerosols. *Am. Ind. Hyg. Assoc. J. 61(3)*:324–333 (2000).
45. **Burstyn, I., H. Kromhout, T. Kauppinen, P. Heikkilä, and P. Boffetta:** Statistical modelling of the determinants of historical exposure to bitumen and polycyclic aromatic hydrocarbons among paving workers. *Ann. Occup. Hyg. 44(1)*:43–56 (2000).
46. **Burstyn, I. and H. Kromhout:** Trends in inhalation exposure to hydrocarbons among commercial painters in The Netherlands. *Scand. J. Work Environ. Health 28(6)*:429–438 (2002).
47. **Yu, R. C., W.Y. Tan, R.M. Mathew, D.A. Andjelkovich, and R.J. Levine:** A deterministic mathematical model for quantitative estimation of historical exposure. *Am. Ind. Hyg. Assoc. J. 51(4)*:194–201 (1990).
48. **Meijster, T., E. Tielemans, J. Schinkel, and D. Heederik:** Evaluation of peak exposures in the dutch flour processing industry: implications for intervention strategies. *Ann. Occup. Hyg. 52(7)*:587–596 (2008).

Mathematical Models for Estimating Occupational Exposure to Chemicals

2nd edition

Edited by Charles B. Keil, PhD, CIH
Catherine E. Simmons, CIH
T. Renée Anthony, PhD, CIH, CSP

AIHA™
Protecting Worker Health

A Publication from the
American Industrial
Hygiene Association