Title: Statistical Analysis of Occupational Exposure Data
Investigator: Rachael M. Jones
Contact Information: School of Public Health, University of Illinois at Chicago, 2121 W Taylor St., Chicago, IL 60612; 312-996-1960; rjones25@uic.edu
Institution: University of Illinois, 1737 W Polk St, Chicago, IL, 60612-7227
State: IL
Sponsor: Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health
Award Number: 5K01OH010537
Start & End Dates: 4/1/2014-3/31/2017
Final Report Date:

Table of Contents

Abstract

Title: Statistical Analysis of Occupational Exposure Data
Investigator: Rachael M. Jones
Contact Information: School of Public Health, University of Illinois at Chicago, 2121 W Taylor St., Chicago, IL 60612; 312-996-1960; rjones25@uic.edu

This K-01 award (5K01OH010537) enhanced the skills and research capacity of the Investigator in the area of statistical analysis of occupational exposure data through training and mentored research activities.

The first research activity in this award characterized variability and determinants of lead exposure during surface preparation activities. The exposure data used in this activity were cross-classified, which means that worker's exposures were measured while the worker participated in one or more groups. Cross-classified designs are an alternative to more traditional hierarchical designs in which one worker is in only one group, and may be particularly advantageous to determine task-based exposures from longer-term exposure measurements (during which a worker performed multiple tasks). Within-worker variability (day-to-day) was the primary driver of variability in workers' personal exposures to lead, and lead concentrations measured at fixed locations within containment were not associated with workers' personal exposures. These findings suggest a worker's own activities and the emission that these activities generate are likely the drivers of exposure levels. We also found the mean exposure measured inside air-supplied blasting hoods of workers was 2.4-fold lower, on average, than the mean exposure measured outside workers' half-mask respirators, which is low relative to the Assigned Protection Factor for these hoods (≥ 25). While an imperfect measure of respirator performance, this finding suggests respirator effectiveness may not always equal the expected performance.

The second research activity in this award focused on Bayesian methods for the analysis of occupational exposure data. Current methods are limited in scope, focusing on the 95^{th} percentile of the exposure profile to evaluate compliance, or extremely complicated. We described and demonstrated three accessible methods for Bayesian analysis that use conjugate prior distributions: With conjugate priors, the posterior distributions for the mean and variance of the logarithm of the exposure profile are analytical expressions that can be readily sampled from in most statistical software packages and in Microsoft Excel. Sample codes are available to the public. As a result, the methods are readily accessible to industrial hygiene researchers and professionals.

The research activities of this award have involved innovative applications of statistical methods for the analysis of occupational exposure data, and contribute to toolbox for occupational exposure analysis.

Section 1

Significant or Key Findings

The first study of this research focused on the analysis of occupational exposures to lead during the preparation of surfaces for painting. While performing this research we described an event of cross-classification in the personal exposure data, which is the first time that this condition has been described in occupational exposure assessment, to our knowledge. With respect to workers' exposure to lead, we found that within-worker variability was the primary contributor to exposure variability. We did not identify lead concentrations measured at fixed locations to be a determinant of workers' personal exposures. We found that the lead exposure of workers wearing supplied-air blasting helmets was 2.4-fold lower, on average, than the lead exposure measured in the breathing zone of workers not wearing blasting helmets.

The second study of this research focused on Bayesian methods for analysis of occupational exposure data. Specifically, we explained and then demonstrated the utility of Bayesian methods with conjugate priors for analysis of lead exposures in a foundry.

Translation of Findings

The first important implication from our study of occupational exposures to lead is with respect to exposure study design. Most exposure studies are designed to characterize the exposure of groups of workers, and employ a hierarchical design in which each worker (and all of his or her exposure measurements) is nested within the group. Alternative designs, however, can be useful for exposure analysis. Cross-classified exposure studies include exposure data for workers that participate in multiple groups. In our example, a worker may have had exposure measurements collected while he or she worked at two different worksites. Another relevant example is the collection of biomarkers that represent exposure over 24-hours or more, during which time the worker performed multiple tasks. The cross-classified framework enables the exposure to be attributed differentially to each task during the exposure period.

The second important implication from our study of occupational exposure to lead is that supplied-air blasting helmets may not perform as well as suggested by the Assigned Protection Factor (APF) when used in worksite. We found that the lead concentration inside blasting helmets was 2.4-fold lower, on average, than the lead concentration outside half-mask respirators. While this is an imperfect measure of respirator performance, the APF for supplied-air blasting helmets is ≥ 25, depending on design, which is substantially larger than was observed in our analysis. There are a number of reasons that the effective performance is lower than anticipated by the APF, including removal of the blasting helmet by workers while still in the containment structure. But, this speaks to the need to consider both respirator performance and worker use of respirators in actual worksites when defining effectiveness.

The key implication of our Bayesian analysis study is the availability of an accessible, flexible tool for Bayesian analysis of occupational exposure data. The methods we demonstrated provide posterior distributions for the sufficient statistics of the logarithm of the exposure profile, from which posterior distributions about other statistics of interest can be obtained. Sample statistical code has been shared, and we have described implementation of one method in Microsoft Excel.

Research Outcomes/Impact

The research performed demonstrated the application of new concepts and methods for the statistical analysis of occupational exposure data. Cross-classification of occupational exposure data may occur accidentally in exposure studies, or by design. This method is particularly useful when an exposure measurement, including a biological marker of exposure, represents exposure that occurs while an individual has different exposure experiences (e.g., performs different tasks) as the technique can identify the contribution of each task to the total exposure. Bayesian analysis with conjugate priors is an accessible approach that yields posterior distributions about sufficient statistics for the characterization of an exposure profile (e.g., the mean and variance of the logarithm of the exposure data) from which each statistic of interest about an exposure profile can be described. This research expands the toolkit for statistical analysis of exposure data by occupational health researchers and practitioners.

Section 2

Scientific Report

Study 1

The first research study in this award involved analysis of lead exposure data previously collected during surface preparation of bridges for painting: Analysis objectives included characterization of variance, identification of determinants of exposure and evaluation of multiple imputation methods for missing data. These research activities were designed to utilize new skills gained through the training aspect of this award.

During the characterization of variance in personal lead exposure measurements, it was discovered that the sampling design was cross-classified, which is rare in occupational exposure assessment. A hierarchical design is more common for repeated measures when groups of workers are defined a priori. In a hierarchical design for occupational exposure assessment, each worker is a member of one group; while in the cross-classified design, each worker is a member of one or more groups. When repeated exposure measures are collected for a worker, in a hierarchical design all exposure measurements for a worker belong to one group; while in a cross-classified design exposure measurements for a worker may belong to different groups.

In the collection of these lead exposure data, the cross-classification design arose accidentally. A convenience sampling approach had been taken, in which workers were selected by convenience at the worksite on the day of sample collection. Since multiple days of sample collection occurred at each worksite, the exposure of many workers was repeatedly measured, and the workers performed multiple job tasks during repeated exposure measurements. In addition, the exposure of some workers was measured repeatedly at when workers were at different worksites. Thus, exposure measurements were cross-classified among worksites, and among work tasks.

It is important to recognize cross-classification designs as distinct from hierarchical designs because of the conceptual implications for exposure. In the hierarchical design, the exposure and risk patterns in the group are assumed to apply to all members of the group. However, workers who participate in multiple groups may be unique, and ignoring this quality may result in incorrect characterization of the group. Fortunately, however, cross-classified data can be analyzed using many existing software packages and functions designed for analysis of hierarchical data using random- and mixed-effects regression models. In this study we used the *lme* package for the R Project for Statistical Computing.

To characterize variance in personal lead exposures, we applied a random-effects regression model: We determined that day-to-day variability within workers was the primary driver of exposure variance. More specifically, within-worker variability ($\widehat{\sigma_w} = 1.36$) was greater than between worker variability ($\widehat{\sigma_B} = 0.454$). Within-worker variance contributed 79% of the total day-to-day variance in exposures, with between-worker and between-group (worksite) variance having similar, minor contributions. The range of exposures experienced day-to-day within workers (${}_{W}R_{0.95} = 208$) was high relative to values observed in other studies. This level of within-worker variance is unusual, and suggests that worker activities (and the level of lead dust generation associated with worker activities) are highly variable from day to day. In the context of engineering controls, capacity should be targeted for high-exposure days.

To characterize determinants of personal lead exposure, we applied a mixed0effects regression model to explore specific objectives.

1. We were interested in determining whether lead concentrations at fixed locations in containment structures were predictive of lead concentrations in the personal breathing zone as this would simplify an exposure monitoring strategy. The logarithm of lead concentrations measured at fixed locations in containment was not statistically significantly associated with logarithms of the lead concentrations measured in workers' breathing zones. Thus, lead concentrations measured at fixed locations are not good indicators of personal exposures, which makes sense due to the proximity of workers to the source of lead dust.
2. We were interested in determining whether the lead content of paint was predictive of lead concentrations in the personal breathing zone as this could help to anticipate exposure levels and select appropriate controls. This analysis was not feasible as there were too few unique lead paint measurements (only four worksites).
3. Some workers wore supplied-air blasting helmets while working in containment, and others wore half-mask elastomeric respirators; exposures were measured outside of half-mask respirators and inside blasting helmets. We were interested to determine the level of protection offered by the blasting helmets. We found the mean difference in the logarithm of the lead concentration in the breathing zone outside the half-mask respirator and inside the blasting helmet to be, on average, 0.892 (95%CI 0.261, 1.56), which means that the lead concentration outside a half-mask respirator was 2.4-fold higher than inside the blasting helmet, on average. While this is an imperfect measure of respirator performance, it suggests that the blasting helmets may not perform as anticipated from the Assigned Protection Factor ($\geq$ 25).

With respect to evaluating multiple imputation, this work was not completed. The data set was determined not to have value for this purpose. Instead, we began to prepare a literature review describing multiple imputation methods and their value to occupational exposure assessment; the method is rarely applied in this field but is commonly used in occupational and environmental epidemiology. This manuscript is in preparation.

Study 2

In the proposal, study 2 was to involve development and evaluation of a likelihood method for use of individual-level exposure estimates for epidemiologic exposure-response analysis. With the addition of Igor Burstyn, PhD to the mentoring team, Study 2 was reframed to explore Bayesian methods, which was more consistent with the proposed training goals. Ultimately, we explored the application of conjugate priors for Bayesian analysis of occupational exposure data.

Our focus on conjugate priors was motivated by limitations of current Bayesian analysis methods used for occupational exposure assessment. Bayesian Decision Analysis, a popular method conveniently implemented through a commercial software package, is inherently limited by its focus on the 95^{th} percentile of the exposure distribution, which is primarily relevant for compliance assessment. Other methods being introduced have sought to improve specification of prior distributions, and have become very complex, boutique approaches accessible only to a subset of academic researchers. Conjugate priors are more accessible because the marginal posterior distributions for the parameters of the exposure distribution have analytic expressions that can be readily sampled from. In our manuscript we include sample code for the R Project for Statistical Computing and for one method in Microsoft Excel.

The first conjugate prior method (Method I) we described includes an inverse-χ^2 prior distribution for the variance of the logarithm of the exposure data, and a normal distribution for the prior distribution of the mean of the logarithm of the exposure data, conditioned on the variance. The posterior distribution for the variance of the logarithm of the exposure distribution is inverse-χ^2, and the posterior distribution for the mean of the logarithm of the exposure distribution, conditioned on the variance, is normally distributed. Parameters of the inverse-χ^2 and normal marginal posterior distributions are calculated from the parameters of the prior and the data used to define the likelihood function.

The second conjugate prior method (Method II) we described includes an inverse-Γ prior distribution for the variance of the logarithm of the exposure data, and a normal distribution for the prior distribution of the mean of the logarithm of the exposure data. The posterior distribution for the variance of the logarithm of the exposure distribution is inverse-Γ, and the posterior distribution for the mean of the logarithm of the exposure distribution is normally distributed. Importantly, in Method II, the prior and posterior distributions are independent, which means that two different sources of information can be used to define the prior distributions. Specifically, we propose the use of a standard prior distribution for variance based on previously published studies of exposure variance.

The third conjugate prior method (Method III) uses non-informative prior distributions, which yield an inverse-χ^2 posterior distribution for variance of the logarithm of the exposure distribution and a normal posterior distribution for the mean of the logarithm of the exposure distribution.

The three methods were demonstrated using lead exposure data collected in a foundry by OSHA investigators. Methods I and II were found to yield similar mean values for the posterior distributions of the mean, geometric mean, geometric standard deviation, and 95th percentile, but Method II yielded substantially less variance in the posterior distributions for the mean and geometric mean than Method I. Method III had high variance in the posterior parameters of interest for the exposure distribution, and indicated a higher skewness (e.g., higher mean for the posterior of the geometric standard deviation distribution).

The publication describing this work was selected as “Editor’s Choice.”

Publications

Jones RM, Burstyn I: [2017] Bayesian analysis of occupational exposure data with conjugate priors. Annals of Workplace Exposures and Health 61: 504-514.

Jones RM, Burstyn I: [2016] Cross-classified occupational exposure data. Journal of Occupational and Environmental Hygiene 13: 668-674.

Human Subjects Reporting

This study was determined to not involve human subjects.

Materials Available for Other Investigators

Software code to implement the Bayesian methods in Study 2 is included in the online supplementary materials accompanying the article.

SCHOOL OF PUBLIC HEALTH
UIC

June 30, 2017

Rose Mosley, Grant Management Specialist
Centers for Disease Control and Prevention
Procurement and Grant Office
2960 Brandywine Rd 2097.01
Atlanta, GA 30341

RE: Final Closeout Reports for Grant 5 K01 OH 010537-03

Dear Colleagues:

Enclosed are the Final Performance report, Final Invention Statement and Certification (Form HHS 568), Equipment Inventory Listing, and Tangible Personal Property Report (SF-428-b) for our CDC/NIOSH grant, "Statistical Analysis of Occupational Exposure Data" (Grant 5 K01 OH010537-03).

The Federal Financial Report (FFR) is submitted through eRA Commons.

If you require any additional information, please contact me at the above address or e-mail us at rjones25@uic.edu or David J Pustek at dpustek@uic.edu.

Sincerely,

Rachael M Jones, Ph.D
Principal Investigator

Mitra Dutta, Ph.D
Vice Chancellor for Research

Environmental and Occupational Health Sciences
The University of Illinois at Chicago
2121 W. Taylor St. (MC 922)
Chicago, Illinois 60612

Phone (312) 996-1960
Fax (312) 413-9898
E-mail rjones25@uic.edu