



**Final Performance Report
For**

**Estimation of Highest Task Silica Exposures
June 28, 2001**

**Grant Number: R03 OH03775-02
National Institute for Occupational Safety and Health (NIOSH)**

**Principal Investigator
L. Faye Grimsley**

**Department of Environmental Health
University of Cincinnati
Cincinnati, Ohio, 45267-0056**

Table of Contents

List of Figures and Tables	ii
Abstract	iii
Significant Findings	iv
Usefulness of the Findings	v
1.0. Introduction.....	1
2.0. Background	2
3.0. Methods and Procedures	3
3.1. Method One.....	5
3.2. Method Two.....	6
3.3. Method Three	7
3.4. Estimation of Exposure Where No Measurement Data Are Available	9
3.5. Adjustment of Data for Free Silica	12
3.6. Determination of Task with Highest Silica Exposure	13
3.7. Exposure-Response Analysis.....	14
3.8. Categorization of Highest Task Exposure.....	14
3.9. Comparison of Highest Task and Cumulative Exposure Metrics	15
4.0. Results.....	17
4.1. Number of Subjects Assigned Imputed Exposure Estimates.....	18
4.2. Odds Ratios Comparison of Three Methods and Using Zero for Imputation	19
4.3. Imputation Methods Validation.....	23
5.0. Discussion	24
6.0. Conclusions.....	26
Acknowledgements	27
References	28
Anticipated Publications.....	30

List of Figures and Tables

Figures

- Figure 3.1:** Outline of Procedure for Missing Data Estimation
Figure 3.2: Method Number One Flow Diagram
Figure 3.3: Method Number Two Flow Diagram
Figure 3.4: Method Three—Expectation Maximization (EM) Algorithm, Flow Diagram
Figure 3.5: Task Exposure Ranking Worksheet
Figure 3.6: Highest Task Determination, Flow Diagram
Figure 4.1: Odds Ratios for No Imputation (Using Zero) Versus Imputation Methods
Figure 4.2: Odds Ratios, by Highest Task Silica Exposure From Using Zero and Imputation Methods

Tables

- Table 3.1:** Average Quartz Content of Settled Dust Samples, by Commodity & Work Location
Table 4.1: Number of Impinger Samples, Mean Silica Exposure, Standard Deviation and Range by Commodity
Table 4.2: Range and Mean Results in Million Particles Per Cubic Foot (mppcf) by Commodity for Each Method
Table 4.3: Number of Subjects Assigned Imputed Exposure Estimates
Table 4.4: Exposure Groups, and Odds Ratios 95% Confidence Intervals (CI) for No Imputation (Using Zero), Geometric Mean, Median Exposure and EM Algorithm Imputation Methods
Table 4.5: t-test Results for Odds Ratios Comparison for Methods 1,2,3 With Highest Task Odds Ratio Using Zero (Liu, 1997)
Table 4.6: Comparison of Number of Subjects by Exposure Group

Abstract

More than 1 million workers in the United States are exposed to silica and approximately 100,000 will develop lung diseases. Despite efforts to decrease silica exposure in the workplace, overexposures and deaths from silicosis continue in certain operations and industries. According to the National Institute for Occupational Safety and Health (NIOSH), each year more than 250 workers die from silicosis.

The objectives of the study were to investigate ways to estimate silica exposure in unsampled tasks and to evaluate the relationship between highest task silica exposure and the development of silicosis. Data from the North Carolina Dusty Trades, a unique resource for exposure-response studies of the effects of silica exposure is the basis for the study. Three methods of imputation were used. The methods included estimating exposures based on the geometric mean exposure at sampled commodity-specific tasks, the median exposure of commodity-specific tasks, and a maximum likelihood estimation statistical model. A case-control study design was used to evaluate the effect that estimated silica exposure from the highest task had on the development of silicosis.

The results indicate that the highest task silica exposure is related to the development of silicosis. Exposure estimates including values imputed using the geometric mean resulted in an overall odds ratio and (95 % CI) of 2.27 (1.49, 3.44), for the median imputation exposure method the overall odds ratio was 2.27 (1.49, 3.44); and overall odds ratio of 2.14 (1.47, 3.11) for the statistical model. These odds ratio results were not significantly different from overall odds ratio of 2.22 (1.52, 3.26) that was calculated using zero for missing data in the highest task.

These results show that the exposure-response relation based on the highest task exposure metric did not increase when non-zero values were imputed using three methods for tasks never sampled. The results are consistent, however, suggesting that peak exposures may be important in the development of silicosis. Further work is needed to better understand the relation between peak exposure and the development of silicosis.

Significant Findings

The specific aims of this investigation were to 1) Develop three imputation methodologies to estimate missing environmental exposure data, 2) Evaluate exposure-response for imputed highest task silica exposures and the development of chronic silicosis, and 3) Contrast any difference between highest task and cumulative silica exposures and the outcome of silicosis. The significant findings are listed below.

Three imputation methods were developed based on the geometric mean exposure at sampled commodity-specific tasks, the median exposure of commodity-specific tasks, and a maximum likelihood estimation statistical model. The methods were used to estimate approximately 23,000 values for tasks without measurements using 9,492 impinger dust count measurements. A total of 610 companies were included in the imputation procedures; for 282 of these companies, no sampling had ever been conducted. Results from the three imputation methods produced similar results. However, the number of subjects (n=888) assigned estimated silica exposure values for each method were different.

Fewer subjects (269) were assigned estimated exposure values when the statistical model method was used as the imputation procedure. Three hundred and thirty-one subjects' highest exposure was based on estimated values for the geometric mean method and 354 for the median exposure method. The difference is due to the larger concentrations imputed using the geometric mean and median methods, resulting in more tasks with estimated values being selected as the highest.

Conditional logistic regression was used to calculate odds ratios based on the highest task silica exposure for each subject and disease outcome. Exposure estimates including values imputed using the geometric mean resulted in an overall odds ratio (95 % CI) of 2.27 (1.49, 3.44), using the median imputation exposure method the overall odds ratio (95 % CI) was 2.27 (1.49, 3.44); and overall odds ratio (95 % CI) of 2.14 (1.47, 3.11) for the statistical model. These odds ratio results were not statistically significantly different from overall odd ratios (95 % CI) of 2.22 (1.52, 3.26) that were calculated using zero for missing data in the highest task from a previous study.

Limited comparisons were made between the odd ratios of the cumulative exposure and the highest task exposure. Further research is needed to investigate and determine a valid comparison methodology. However, the odd ratios using the highest task exposure were approximately 1.5 times higher than odd ratios (95% CI) of 1.39 (1.02, 1.89) that were observed when the cumulative exposure was used in a previous study using the North Carolina Dusty Trades data.

The results from this investigation suggest that the highest task exposure metric based on imputed data odds ratios did not improve the exposure-response relationship for silica and silicosis observed in a previous study using zero and when estimated by using the geometric mean, median or the statistical model. However, the results from the three methods are consistent and may suggest the importance of peak exposures and the development of silicosis.

Usefulness of the Findings

While the three methods used to estimate missing data in the North Carolina Dusty Trades data are comparable and similar results were obtained when using no imputation, these methods could perhaps be used to estimate missing data in other investigations where task exposure is the exposure metric.

Developing and testing the three methods not only permitted the evaluation of whether peak exposures increase the risk of developing silicosis, but also provided a foundation for future research studies that will advance our understanding of approaches for estimating missing exposure data.

The findings from this study can serve as a guide to possible changes in sampling methods, and targeted control techniques to reduce peak silica exposures that might be implemented to decrease disease in the workplace.

1.0. Introduction

Silica is one of the most abundant elements in the earth's crust. It can exist in several polymorphs and occurs naturally in crystalline and amorphous forms. The most common forms of crystalline silica are quartz, tridymite, and cristobalite, with quartz being the most abundant. Silica exposure has been known to exist in the workplace for centuries (Ramazzini, 1964; Corn, 1980). Overexposures continue to be documented by the Occupational Safety and Health Administration (OSHA) and the Mine Safety and Health Administration (MSHA) in the United States and may occur in industry locations such as foundries, mining operations, construction and agriculture (NIOSH, 1999). OSHA estimates approximately 100,000 of the one million workers exposed to silica will develop lung diseases (OSHA, 1996).

The adverse health effects of silica have been known for many years. According to Crofton & Douglas (1981), the lung disease silicosis is uniquely associated with exposures to crystalline silica. Silicosis is defined as a fibrotic pulmonary disease caused by long-term inhalation of respirable crystalline silica. Usually, exposure of 20 to 30 years is necessary before the disease becomes apparent (Davis, 1996).

Many studies have been conducted to assess cumulative silica exposure and its association with silicosis. Several studies including Rice et al. (1985), Landrigan (1987), and Rosenman et al. (1996) support a relationship between cumulative silica exposure and silicosis. However, limited information is available regarding short-term task peak exposures to silica and the development of silicosis.

When considering substances that produce chronic diseases (e.g., silicosis), cumulative exposure received over some period of time is generally expected to be a relevant index of dose in epidemiological investigations rather than peak levels (Rappaport, 1991). Studies have shown that this may not be the best or only relevant exposure metric (Rice et al., 1988; Checkoway and Rice, 1992; Liu, 1997) and consideration should be given to other means of assessing silica exposure in studies of silicosis. One such way of examining the exposure-response relationship can be by studying peak exposure intensity and disease outcome. This investigation focused on estimating peak (highest) silica exposures resulting from tasks performed and the risk of developing chronic silicosis.

The objectives of this study were to investigate various ways to estimate missing environmental exposure data and to evaluate whether high intensity short-term occupational exposures to silica increase the risk of silicosis. A case-control study design is used to evaluate the effect that silica exposure from the highest task during the working lifetime has on the development of chronic silicosis.

Data from the North Carolina dusty trades surveillance program were used for the analysis. The data sets consist of environmental, work history data and chest x-ray readings that were collected from 1935 to 1980 (Rice, 1983). A total of 216 cases and 672 controls have been selected. Diagnosis of silicosis is based on serial chest x-ray readings and occupational history.

2.0. Background

Silica exposure levels are related to different occupational jobs. In the North Carolina Dusty trades study Rice et al. (1984), demonstrated a strong exposure-response relationship and that the time of the highest exposure job during the entire work history is an important determinant in the development of silicosis.

A follow-up to this investigation was conducted by Liu (1997) in partial fulfillment of a masters of science at the University of Cincinnati. The hypothesis tested if there was a difference in the odds of developing silicosis between those characterized by the highest exposure tasks and cumulative exposure. The potential link between short-term, task-related exposure and risk of developing silicosis was investigated in this study. The highest task metric was categorized and the odds ratios calculated. The results showed that the odds ratios calculated for the highest exposure task metric were approximately 1.5 times higher than those calculated for the cumulative metric.

Previous calculations of odds ratios used a value of 0 where task was not sampled. In a follow-up analysis by Rice et al. (1997) using imputed values to estimate the exposure at un-sampled jobs, and a recalculation of the cumulative metric showed that the strength of the exposure-response relationship increased. In the Liu analysis, 0 was used for any task not sampled. Therefore, it was of interest to impute task-related exposure and investigate the results.

The purpose of this study was to determine if there was a significant difference in the risk of developing silicosis between study subjects characterized by highest task imputed silica exposure and those characterized by using zero as the exposure for tasks without measurements.

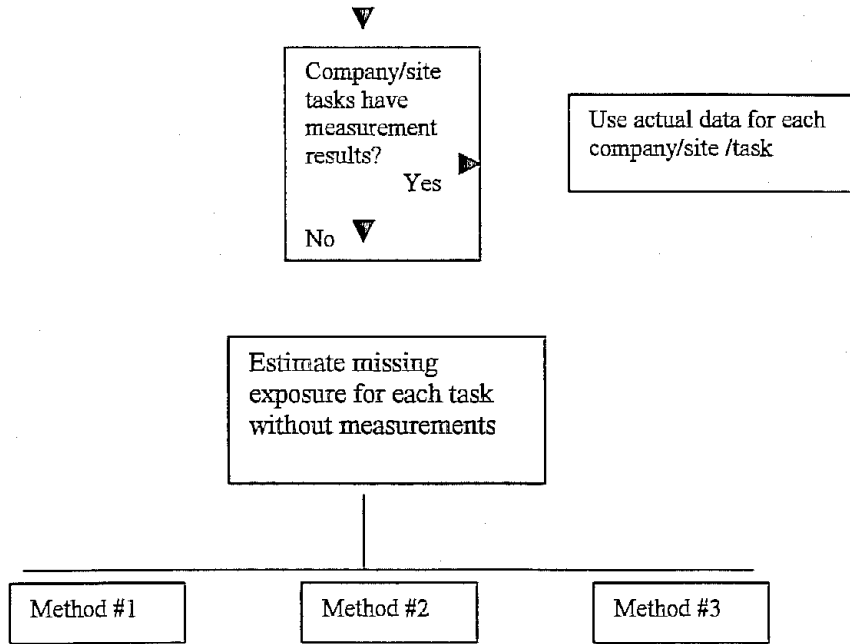
3.0. Methods and Procedures

In this investigation, a case-control study design was used to evaluate the effect that silica exposure from the highest task has on the development of silicosis. Data from the North Carolina dusty trades surveillance program were used for the analyses. The data sets contain more than 12,000 samples that were collected in 11 different commodities. Background information for the North Carolina Dusty Trades environmental data indicate that impinger sample results are not available for some tasks. To identify the unsampled tasks, a review of the environmental data was conducted. The tasks without measurements were identified for silica exposure estimation. Three imputation methods were used to estimate missing exposures. After estimation of missing silica exposure, the highest exposure task was identified for each case and referent and the exposure-response relationship was evaluated.

To estimate missing values for tasks, three imputation methods were used. (See Figure 3.1 for Outline) The first method estimated missing values based on the geometric mean of commodity-specific tasks. The second method involves using the median exposure for specific tasks for all company locations to estimate missing data points. The third method imputes a maximum likelihood estimate based on a statistical model that was developed to predict the exposure values for tasks without measurements from mean exposures and independent variable parameter estimates.

Figure 3.1 Outline of Procedure for Missing Data Estimation

Evaluate Measurement Data for each Commodity/Company/Site Task



Link Task Exposure Data with 888 Study Subjects

Identify Highest Exposure Task

Evaluate Exposure Response

Compare Results of Methods

Specific Aim Number One

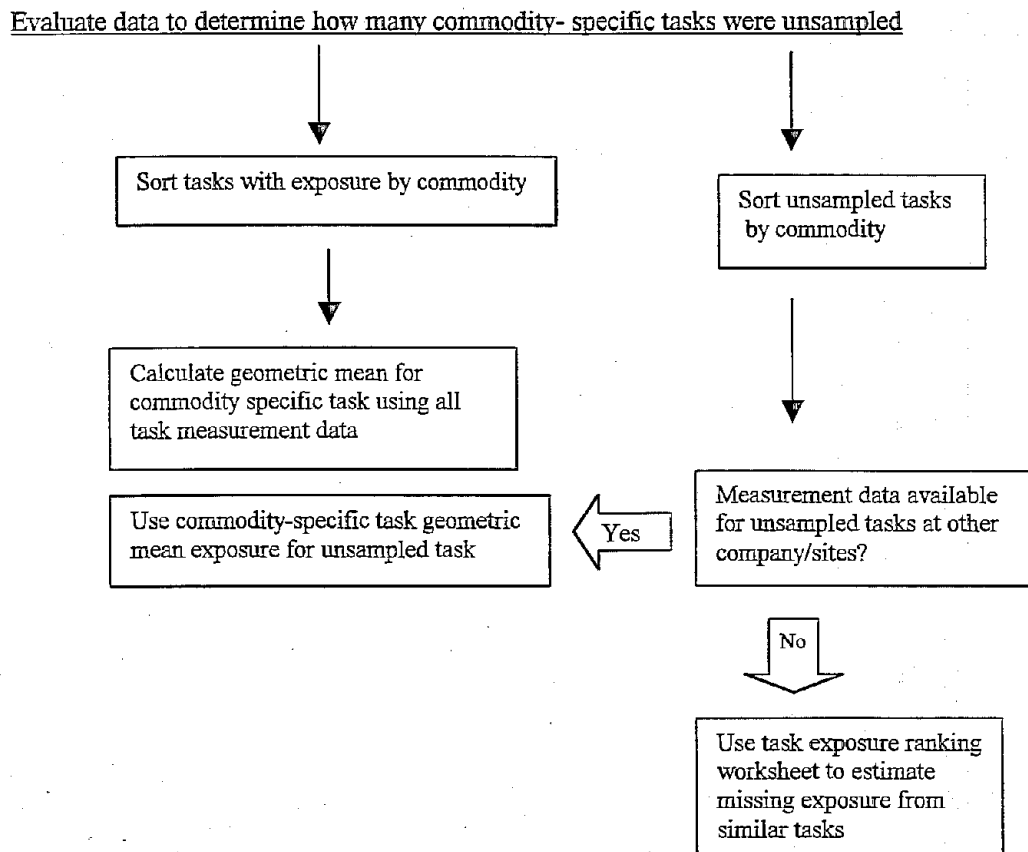
Develop three methodologies to impute highest task silica exposure for missing environmental exposure data.

Three methods were developed and used to estimate silica exposure for tasks with missing exposure data. Procedures for each method are outlined below.

3.1. Method One

The imputation procedure for method one involved a number of steps. First, the impinger data set (total dust counts) was evaluated to identify how many commodity-specific tasks at each company/site had missing values. The missing values were then estimated by calculating a mean of the natural logarithm dust counts from all companies within the commodity for a specific task. The geometric mean was calculated by taking the exponent of the natural logarithm mean dust counts. This geometric mean value was used as the exposure estimate for the same specific tasks without measurements within that commodity. (See Figure 3.2 for method one flow diagram) To estimate exposure for tasks where no data are available, a task exposure ranking worksheet was developed. (See section 3.4 for more details)

Figure 3.2 Method Number One Flow Diagram

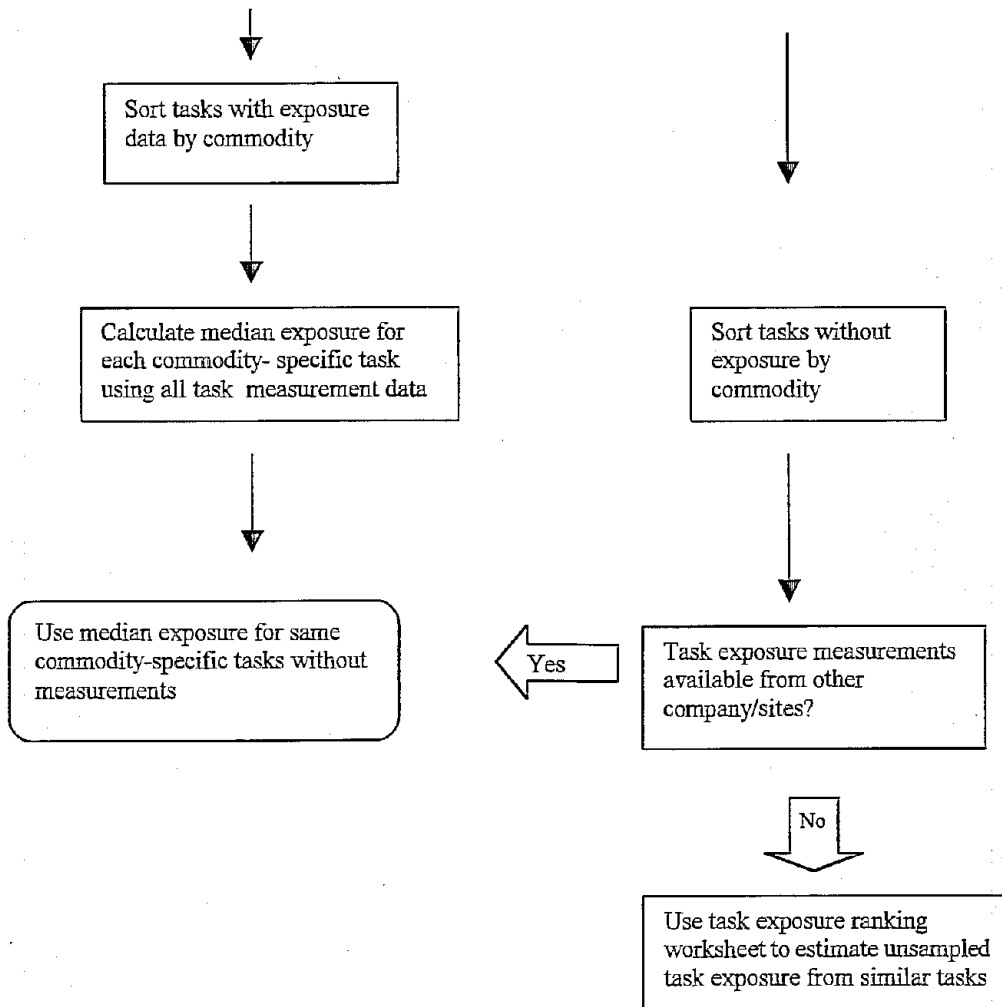


3.2. Method Two

The missing values for each commodity-specific task at every company/site were estimated by determining the median exposure from all measurements for a specific task from all company/sites within a commodity. This median exposure was used as the exposure estimate for the specific tasks without measurements within that commodity. (See Figure 3.3 for method two flow diagram) For tasks where no data are available, the task exposure worksheet was used.

Figure 3.3 – Method Number Two Flow Diagram

Evaluate data to determine number of commodity-specific tasks without measurements



3.3. Method Three

The statistical model developed in method three used all measurements to predict an exposure value for each task within a commodity without measurements. The predicted values generated for commodity-specific tasks were used as the exposure estimate for commodity-specific tasks without measurements.

The following describes how the statistical model for the third method was developed using multiple regression analysis. The full model included one dependent variable (natural log of dust count exposure) and four independent variables (commodity, company/site, task, and date of sample collection). All available variables from the data set were included to see if any had an effect on the outcome variable. A stepwise approach was used to determine the final model based on inclusion or omission of independent variables.

The stepwise procedure can be conducted as a forward selection or backward elimination process (Cody and Smith, 1997). In this investigation, backward elimination was used because all independent variables could have an effect on the dependent variable. The backward stepwise procedure starts with all independent variables entered into the model. Then, in successive steps, variables are deleted until all those remaining produce F-statistics significant at a specific level. Variables were deleted from the regression model based on a statistical criteria of $\alpha = 0.10$. The final model listed below was used to estimate the missing data points (predicted value of Y) utilizing the intercept constant, as well as parameters that were weighted by regression coefficients. The least squares criterion was used to determine best-fitting values for the intercept and the regression coefficients. The magnitude of the relationship between the dependent variable and the independent variables was addressed by evaluating the multiple correlation coefficient (R). The overall accuracy of the model was evaluated by squaring this coefficient. The R^2 represents the proportion of variance in the dependent variable accounted for by the independent variables. The final model for dust count exposure included only the independent variables of commodity and task.

Final Model:

$$Y' = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

where Y'	= predicted value for variable Y
α	= intercept constant
X_1 - X_2	= independent variables of commodity and task
β_1 - β_2	= regression coefficients for the independent variables
e	= error

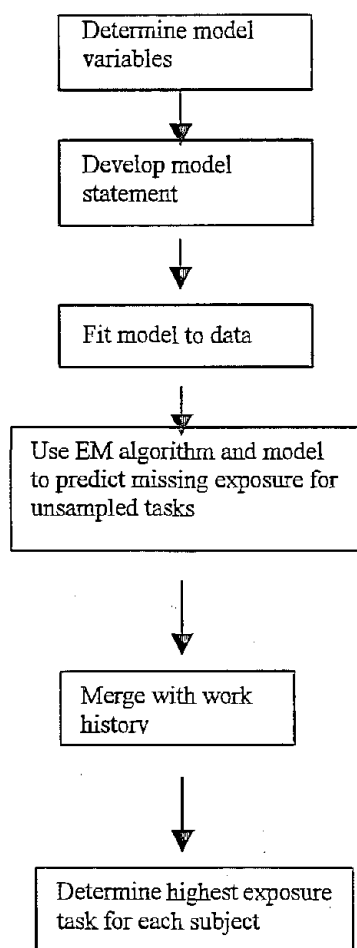
The final model R^2 was 0.59 which can be interpreted that almost 60% of the variability in task exposures were due to commodity and task.

Then an Expectation-Maximization (EM) algorithm in conjunction with the final model was used to estimate the missing data. The EM algorithm is an iterative procedure that is performed in two steps, an Expectation (E) step and a Maximization (M) step. In the E-step an expected value is calculated for each commodity-specific task, based on mean exposure. In the M-step the

expected values are substituted for the missing data from the E-step and then the multiple regression model was ran to obtain new parameter estimates as if no data were missing. The procedure iterates through these two steps until convergence is obtained (Little and Rubin, 1987). In this investigation, the mean of the natural logarithms for commodity-specific tasks were used to replace missing values in the expectation step and then new parameters were estimated based on these mean values. A predicted value was determined in the maximization step for each task without a measurement. After predicted values were determined, all were exponentiated.

Model building procedures in SAS (SAS, 1996), were used to develop the estimates. (See Figure 3.4 for method three flow diagram) For a commodity-specific task where no measurement data were available, the EM algorithm used the mean of natural log measurements from the relevant commodity.

Figure 3.4 Method Three – Statistical Model (Expectation-Maximization (EM) Algorithm), Flow Diagram



3.4. Estimation of Exposure Where No Measurement Data Are Available

For methods one and two, a different approach was used for the imputation for tasks where no data are available at any company in a commodity. This method entails using results from similar tasks within the same commodity. The missing values were estimated by using methods number one and number two in conjunction with a qualitative method. To use this approach, tasks were qualitatively classified into similar groups. The task exposure ranking worksheet seen in Figure 3.5 was developed to assist in classifying tasks into similar groups.

The task exposure ranking worksheet consists of 20 set criteria predicted a priori to have an effect on dust generation and exposure. The worksheet is designed to identify specific activities and operations that will lead to dustiness in the process area, thereby increasing the potential for exposure for a task performed in a specific area.

Each task was evaluated using the exposure ranking worksheet and process information and assigned an exposure rank of category 1, category 2, category 3, or category 4. Exposure ranks are qualitatively defined as follows: category 1 (minimal dust exposure), category 2 (low dust exposure), category 3 (moderate dust exposure), and category 4 (high dust exposure). The more criteria present for a particular task, the higher the potential rank (i.e., dustier). Cutpoints for each exposure rank were based on predetermined ranges. For example, category 1 (0-1) would have zero or one criteria checked, category 2 (2-4), category 3 (5-8), and category 4 (≥ 9). These ranges were based on similar categorizations that were used by other authors. For example, Hessel (1986) used 0 (non-dusty), 3 (low), 6 (moderate), and 12 (high) to assign qualitative values to describe potential silica exposures. Siemiatycki et al., (1988), used a three point categorization scale of 1 (low), 4 (medium) and 9 (high) to create an exposure based index for several chemicals. Checkoway (1993) assigned dustiness factors of 0 (unexposed), 1 (low), 3 (medium), and 6 (highly exposed) to categorize potential silica exposures for diatomaceous earth workers.

After classifying tasks, methods one and two were employed to estimate the missing data points. After ranking each task using the worksheet, exposure was estimated based on placing the tasks without measurements into categories with measured tasks from the same commodity. Based on the geometric mean (method one) and the median exposure (method two) for commodity-specific tasks, the tasks with measurements were grouped into four categories (1, 2, 3, and 4) based on quartiles. Category number 1 represents low exposure and categories 2, 3, and 4 represent increasing exposure.

To estimate missing exposures using this procedure, the task was ranked and placed into categories as outlined above and the geometric mean or median exposure was determined for each commodity category and used for relevant similar tasks with missing values. These values were then added to the impinger exposure data set.

Most of the tasks assigned exposure values using this procedure represented jobs that were considered less dusty (e.g., office or supervisors).

Figure 3.5 Task Exposure Ranking Worksheet

Missing Task Exposure Profile

Task Name/ID: _____

Sample Type: Personal or General

Commodity: _____

Location Where Task Performed:

___ Underground ___ Surface ___ Mill

Date: _____

Criteria		Yes	No	Comment
1.	All or part of task performed within a process area?			
2.	Task performed in proximity or near dusty operations?			
3.	Task isolated to this area or other dusty areas of process?			
4.	Minimal water is present in the process where task performed?			
5.	Task performed by using heavy equipment (crane, dozer, pan, euclid, etc.)?			
6.	Task involves using truck or mobile equipment (e.g., towmotors, power shovel)?			
7.	Any hand tools (e.g., shovel, fork, pick, hammer, chisel, etc.) used to perform task?			
8.	Task involves using power hand tools (e.g., saw, grinder, drill)?			
9.	Task involves using pneumatic hand tools (surfacers, polisher, buffer)?			
10.	Pneumatic equipment (jackhammer, drill, abrasive blaster) used to conduct task?			
11.	Is dust generated from mechanical action (e.g., screening, crushing, milling, mixing, shaking) where task is performed?			
12.	Task is performed near pressing, extruding, drying or baking operations?			

Figure 3.5 Task Exposure Ranking Worksheet

13.	Task involves manual labor such as loading, bagging, digging?			
14.	Task performed in shop (e.g., machining, welding, woodwork)?			
15.	Task performed may be associated with maintenance activity (cleaning, oiling, etc.)?			
16.	Housekeeping activity (sweeping or using compressed air) is needed to conduct task?			
17.	Task is performed inside enclosed space (boxcar, railcar, booth, or shed)?			
18.	Task involves finishing or handling (stacking, packaging, labeling) final product?			
19.	Respiratory protection used while performing task?			
20.	Engineering controls (LEV, water spray, enclosures) used to suppress dust?			

Total number of criteria present that effect dust generation and exposure* =

Exposure Rank =**

*The more criteria checked as yes the greater the potential for dust exposure and thus higher the rank.

**Exposure Rank should be classified as category 1(0-1), category 2 (2-4), category 3 (5-8), or category 4 (9 or >).

A Process area is defined as any location where there is a potential for dust generation from vehicular traffic, or material processing, transport and handling.

3.5. Adjustment of Data for Free Silica

Since the impinger sampling results represented total dustiness in the workplace, an adjustment to free silica was conducted by using the percent quartz in a sample of settled dust. Once all tasks within a commodity company/site had exposure estimates, the total dust counts were adjusted for free silica content by multiplying the average quartz content of settled dust samples, by commodity and sample location (e.g., mill, surface, underground). The average quartz contents are listed in Table 3.1.

**Table 3.1. - Average Quartz Content of Settled Dust Samples,
by Commodity & Work Location**

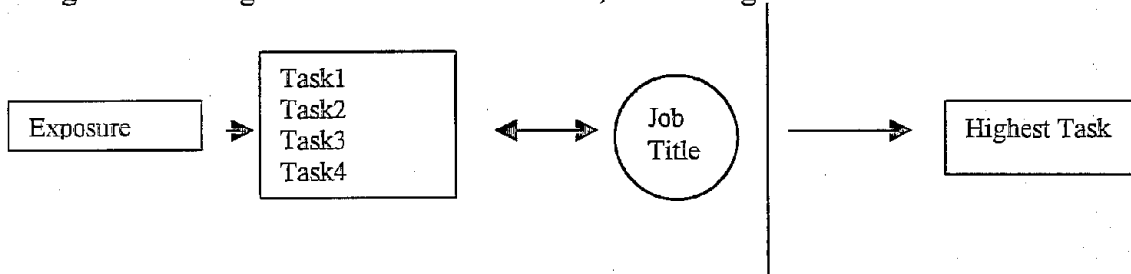
Mineral Commodity	Location	(%SiO ₂)
Mica/Spar	Plant	6.5
Foundry	Plant	32
Granite	Plant	25
	Surface	27
Kaolin	Plant	30
Lithium	Plant	35
Other	Plant	18
	Surface	12
Pyrophyllite	Plant	49
	Surface	54
	Underground	20
Hard Rock	Plant	13
	Underground	20
Crushed Stone	Plant	29
Talc	Plant	1.5

(From Rice, 1983)

3.6. Determination of Task with Highest Silica Exposure

The exposure estimates from the three methods were then linked to each subjects' job titles and associated tasks and the task with the highest silica exposure was identified for each subject. The subject's highest silica exposure was based on company/site-task mean exposures. The highest silica exposure was determined by calculating mean values for all company/site tasks using measurements and imputed values. This exposure was used in the exposure response evaluation described in section 3.7. The process used to determine the highest task is outlined in the flow diagram in Figure 3.6.

Figure 3.6 – Highest Task Determination, Flow Diagram



Specific Aim Number Two

Determine odds ratios based on highest task exposures.

3.7. Exposure-Response Analysis

In this study, conditional logistic regression analysis was used to investigate the relationship between the disease and associated exposure. A conditional logistic regression equation for a binary outcome predicts the natural log of the odds for a subject being in one category vs. the other. The categories for this study were coded as 0 for no disease (referent) and 1 for disease (case) and represent the dependent variable (Kleinbaum, 1994). The independent variable, the concentration of quartz, was calculated for the highest exposure task metric for each subject.

A total of 216 cases and 672 controls have been selected from the North Carolina dusty trades database. Cases were selected from workers who were diagnosed with silicosis between 1935 and 1980. Diagnosis of silicosis is based on serial chest x-ray readings and occupational history. The controls were selected from a pool of disease-free workers by matching each case with up to 4 controls by race, year of birth within 5 years, year of hire within 5 years, and the duration of follow-up. All cases and controls selected were male, 660 were white and 228 were black.

Four exposure groups based on the highest tasks metric were developed as described in section 3.8. The lowest group is the referent, which was considered the non-exposed group; groups 1, 2, and 3 were groups of increasing quartz exposure. Conditional logistic regression was used to calculate the odds ratio for each "exposed" group compared with the referent.

3.8. Categorization of Highest Task Exposure

Categories for the highest task exposure metric for the cases and referents were determined from measured values and the imputed estimates from the three methods.

Various algorithms have been used to develop exposure groups. Lynch and Ayer, (1966) used the upper and lower bound of exposures for categorization. Dement, (1983), and Rice, (1983), grouped exposures based on this categorization as well. Usually, the upper bound of the exposures placed in each category is 3-4 times the lower bound. In this investigation, the upper bound is 3 times the lower bound. The lowest group (0) is the referent; groups 1, 2, and 3 represent categories of increasing quartz exposure. The values were grouped into four category groups numbered 0-3 representing a referent group and three others based on increasing exposure.

The average for each group was determined for all three methods. These categorizations led to various group exposures for each method. The cutpoints for all three methods were categorized as exposure with less than 2.99 million particles per cubic foot (mppcf) being considered the referent group (group 0), and the range of exposures for in groups 1,2, and 3 are 3.0-10.99, 11.0-29.99, and 30 or more mppcf, respectively. The exposure group means for method one were 2, 6, 19 and 75 mppcf, for method two 2, 7, 18, and 80 mppcf, and for method three 2, 6, 17, and 75 mppcf.

In developing these groups, it was suggested that the means for each group must be significantly different from adjacent groups (Lynch et al., 1966). In this study, the groups were tested for significance by using a two-way analysis of variance and were found to be significant ($\alpha = 0.05$).

In the previous study by Liu (1997), four exposure groups based on an estimate of the highest task exposure from measured values were developed. The lowest group is the referent, groups 1, 2, and 3 were groups of increasing quartz exposure based on exposure over the duration of the task. Group 1 exposure is roughly equal to a mean of 10 million particles per cubic foot and has a range of 6 – 13.99 mppcf. Group 2 represents a mean exposure of 27 million particles per cubic foot and has a range of 14 – 35.99 mppcf. The average exposure in group 3 is approximately 109 million particles per cubic foot and represents exposures > 36 mppcf.

Specific Aim Number Three

Contrast any difference between highest task and cumulative exposure metrics and the outcome of silicosis.

3.9. Comparison of Highest Task and Cumulative Exposure Metrics

One limitation of this investigation involved the inability to compare the results from this study with studies that used cumulative exposures. From a previous study, Rice (1983) calculated cumulative silica exposures for each subject. The cumulative exposures were grouped into four categories and then the exposure-response relationship was evaluated comparing the exposed groups to the referent group. The category groups were based on the following categorization: exposure less than 20 million particle years was considered the reference category, group 0. The range of exposures for groups 1, 2, and 3 were 20-59.9, 60-179.9, and 180 or more million particle years, respectively.

In the current investigation and the previous investigation by Liu (1997), the silica exposure from the highest task during an individuals working lifetime was used as the exposure variable. Therefore, comparisons of the highest task exposure from the Liu investigation (no imputation) and exposures values using the three imputation methods were conducted. The results are outlined below in section 4.2.

Perhaps a better comparison would be to see if there is a difference in the exposure-response relationship using imputed data from the three methods to calculate job title values and then calculating cumulative exposures. Further research is needed to determine a valid comparison methodology for cumulative and peak silica exposure metrics.

4.0. Results

The impinger data set contained 9,492 samples that were collected in 11 different commodities within 328 companies. The samples were collected from 1935 to 1980. The total number of samples that were available for each commodity and minimum and maximum silica dust counts are outlined in Table 4.1.

Most commodities in the impinger data set have some measurements for various tasks, with the exception of tasks that were considered less dusty, where there are no measurements available. In addition to the companies in the impinger data set, there were 282 companies where study subjects worked, but sampling was not performed. This was probably due to limited resources within the North Carolina Dusty Trades Program or short-duration jobs. Since 282 companies had no measurements and none of the 328 companies in the original impinger data set had measurements for all tasks, 610 companies were included in the imputation process.

Table 4.1 Number of Impinger Samples, Mean Silica Exposure, Standard Deviation and Range by Commodity

Commodity	Number of Impinger Samples	Mean (mppcf)*	Standard Deviation	Range (mppcf)
Mica/Feldspar	1471	6.0	23.7	0.01 - 432
Foundry	2971	4.0	17.8	0.03 - 656
Granite	1582	7.0	50.9	0.01 - 1833
Kaolin	267	14.0	39.1	0.06 - 354
Lithium	46	3.0	3.5	0.07 - 18.0
Other	122	3.0	5.0	0.02 - 36.0
Pyrophyllite	586	22.0	79.7	0.1 - 1105
Quartz	78	24.0	60.9	0.8 - 360
Hard Rock	255	7.0	29.4	0.07 - 380
Crushed Stone	1891	9.0	43.3	0.01 - 1057
Talc	223	0.8	1.9	0.01 - 14.0
Total	9492			

* - million particles per cubic foot (mppcf)

The results from the three imputation methods are described below. The mean and range results in million particle per cubic foot (mppcf) are similar for the three methods and are presented in Table 4.2. The highest mean values were observed for the kaolin and pyrophyllite commodities and the lowest mean value is in the talc commodity.

Table 4.2 Range and Mean Results in Million Particles Per Cubic Foot (mppcf) by Commodity for Each Method

Commodity	(Range) Mean mppcf		
	Method #1	Method #2	Method #3
Mica/feldspar	(0.10 – 354) 25	(0.10 – 354) 25	(0.10 – 354) 25
Foundry	(0.62 – 35) 10	(0.83 – 35) 10	(0.83 – 35) 8
Granite	(0.25 – 134) 19	(0.25 – 134) 19	(0.25 – 134) 19
Kaolin	(0.15 – 326) 37	(0.15 – 326) 49	(0.15 – 326) 34
Lithium	(0.46 – 4.0) 3	(0.67 – 4.0) 3	(0.33 – 4.0) 3
Other	(0.60 – 18) 4	(0.60 – 18) 4	(0.33 – 18) 3
Pyrophyllite	(16 – 74) 43	(16 – 74) 43	(16 – 74) 43
Quartz	(5.0 – 10) 9	(5.0 – 9.0) 8	(14 – 14) 14
Hard Rock	(3.0 – 127) 28	(3.0 – 127) 24	(3.0 – 127) 18
Crushed Stone	(0.59 – 177) 23	(0.57 – 177) 14	(0.30 – 177) 14
Talc	(0.04 – 3.0) 2	(0.05 – 3.0) 2	(0.06 – 3.0) 2

*Method #1- Geometric Mean, #2- Median Exposure, #3- Expectation Maximization

4.1. Number of Subjects Assigned Imputed Exposure Estimates

The number of subjects that were assigned exposures based on estimates from imputation rather than measurements is listed in Table 4.3. The number of subjects requiring imputed exposure estimates was the highest when the geometric mean and the median were used to estimate the exposures. Thirty seven percent (331/888) of the subjects were assigned estimated exposures when the geometric mean method was used and 40% (354/888) when the median method was used. When the EM algorithm was used as the imputation procedure, 30% (269/888) of the subjects were assigned estimated highest silica exposures. This lower number of estimated exposures used in the EM algorithm method is due to the number of actual measurements for another task that are higher than the imputed task value.

Table 4.3 – Number of Subjects Assigned Imputed Exposure Estimates

	Number of Subjects		
	Method One (Geometric Mean)	Method Two (Median Exposure)	Method Three (EM Algorithm)
Cases	56	68	45
Controls	275	286	224
Total	331	354	269

4.2. Odds Ratios Comparison of Three Methods and Using Zero for Imputation

Odds ratio results from the highest task exposure metric when zero was used for the missing data were compared to the imputation methods. The results indicate that the odds ratio (2.22) for the highest task exposure metric without imputation (Liu, 1997) was not significantly different from the odds ratios determined in this study (See Table 4.4). However, the three imputation methods used in this study to estimate missing data resulted in a slight increase in the overall odds ratios 2.27 for the geometric mean and median exposure imputation methods and a slight decrease in the odds ratio (2.14) for the EM algorithm. The results for the highest task metrics (using zero and imputation) presented in Table 4.4 and are graphed in Figure 4.1.

A comparison of the highest task exposure using zero for unsampled tasks and methods 1, 2, and 3 was conducted by testing for significant differences between their estimated odds ratios using an independent t-test (see Table 4.5). Table 4.6 shows the number of subjects in each exposure group from using zero and the three imputation methods.

Table 4.4 –Exposure Groups, and Odds Ratios 95% Confidence Intervals (CI) for No Imputation (Using Zero), Geometric Mean, Median Exposure and EM Algorithm Imputation Methods

Exposure Group	Odds Ratio (95% CI)			
	Method			
	No Imputation (Using Zero)	#1 (Geometric Mean)	#2 (Median Exposure)	#3 (EM Algorithm)
1,2,3 vs.0	2.22 (1.52, 3.26)	2.27 (1.49, 3.44)	2.27 (1.49, 3.44)	2.14 (1.47, 3.11)
1 vs.0	1.21 (0.77, 1.88)	1.57 (0.98, 2.50)	1.59 (0.99, 2.54)	1.48 (0.95, 2.30)
2 vs.0	3.02 (1.89, 4.84)	2.95 (1.80, 4.83)	2.89 (1.76, 4.74)	2.60 (1.63, 4.14)
3 vs.0	4.65 (2.89, 7.47)	3.61 (2.09, 6.24)	3.61 (2.09, 6.24)	3.16 (1.91, 5.22)

Figure 4.1 – Odds Ratios for No Imputation (using zero) Versus Imputation Methods

Odds Ratios for No Imputation and Imputation Methods

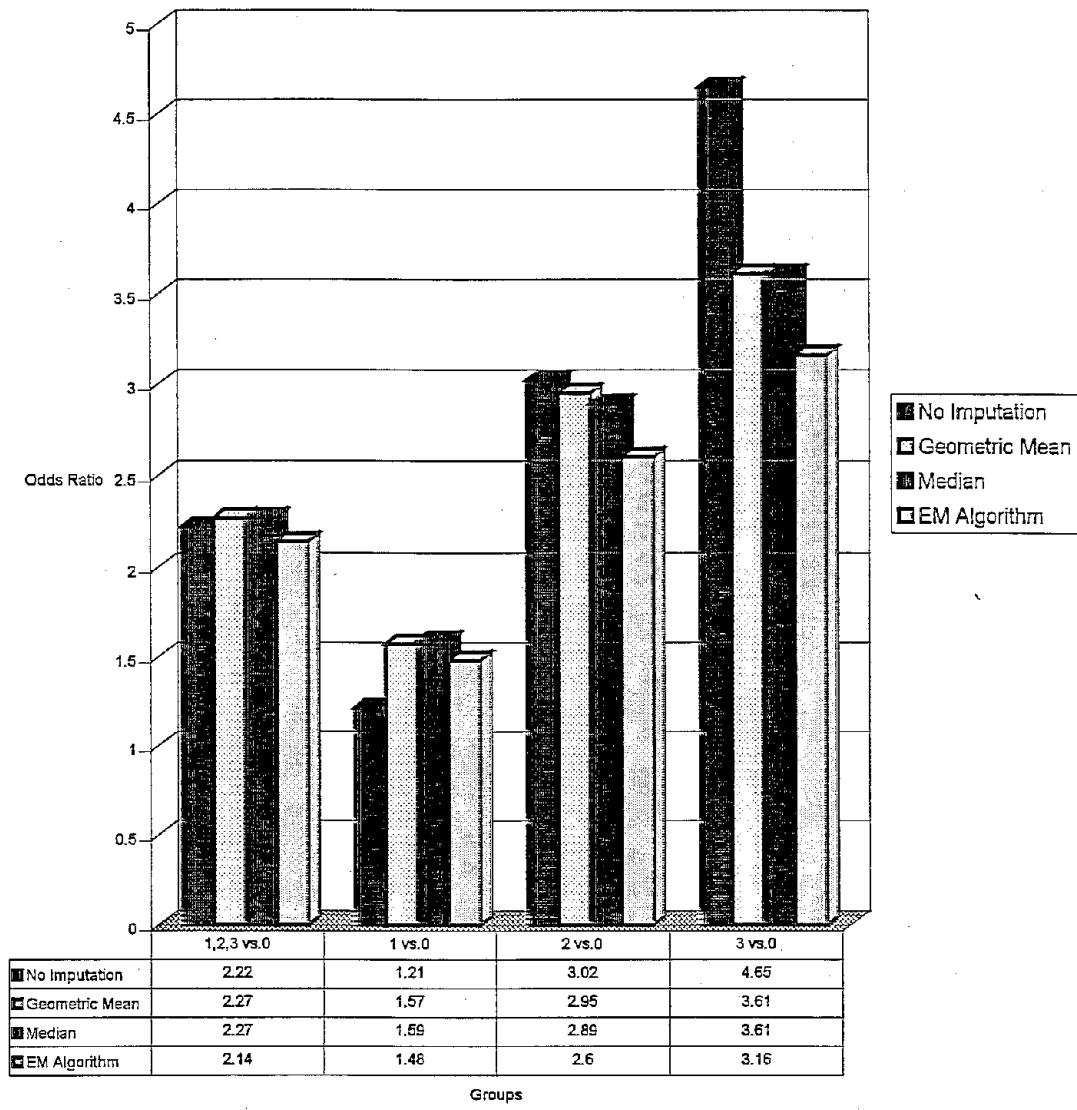


Table 4.5 – t-test Results For Odds Ratios Comparison For Methods 1, 2, 3 With Highest Task Odds Ratio Using Zero (Liu, 1997)

Method Compared	t- test	P-value
Method 1 vs. Using Zero	0.093	>0.50
Method 2 vs. Using Zero	0.093	>0.50
Method 3 vs. Using Zero	0.194	>0.50

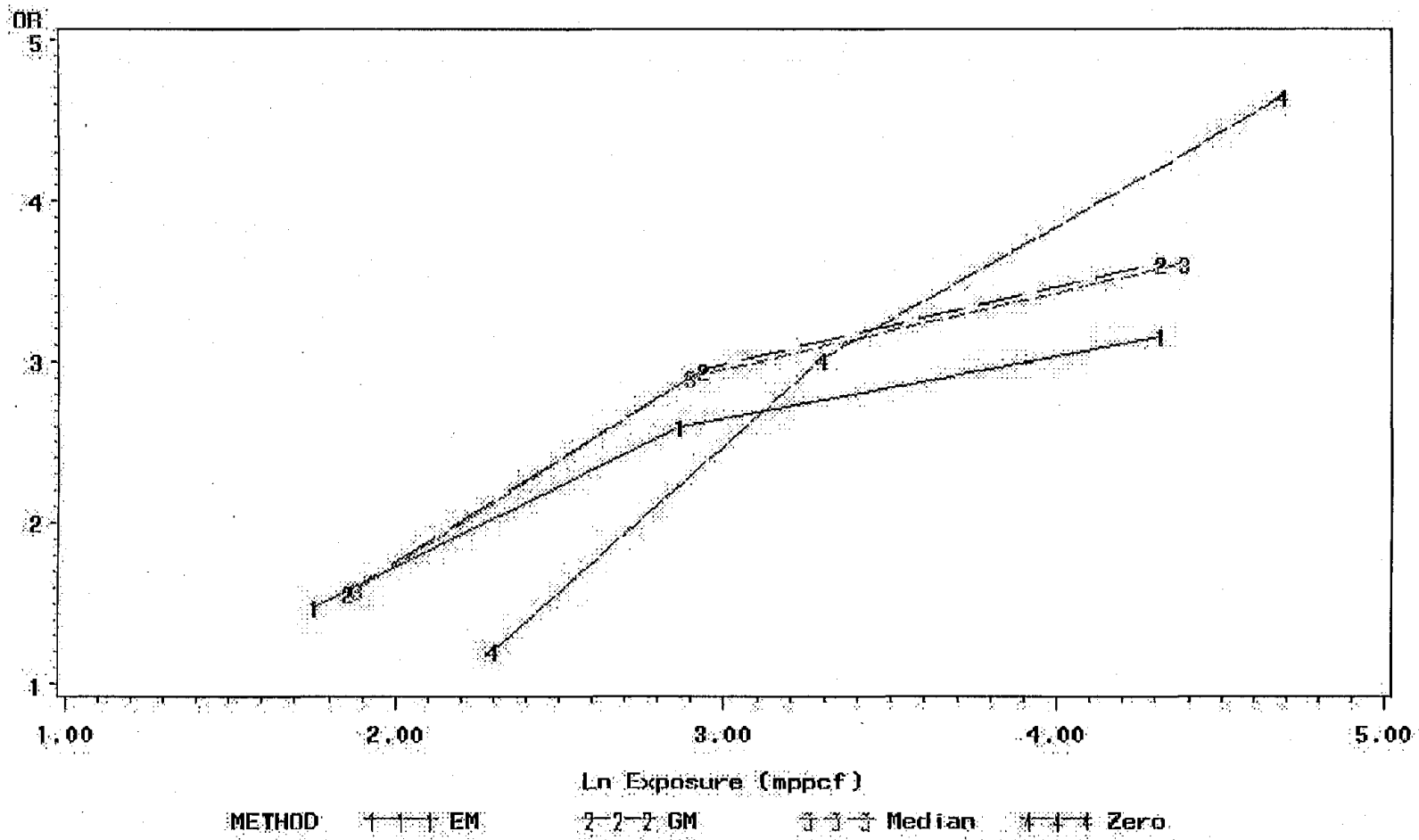
Table 4.6 – Comparison of Number of Subjects by Exposure Group

Exposure Group	No Imputation	Method*		
		#1	#2	#3
0	260	253	253	330
1	325	323	324	262
2	161	190	189	174
3	142	122	122	122
Total	888	888	888	888

* - Method #1- Geometric Mean, #2- Median Exposure, #3- Expectation Maximization

Figure 4.2 shows an exposure-response curve that was plotted using the odds ratios (y-axis) from no imputation and the three imputation methods, and the exposure group means (natural log silica exposure) in mppcf on the x-axis. The slopes of the lines are similar and indicate the consistency in the imputation methods. Note that the method numbers on the plot are now transposed (e.g., Expectation Maximization (EM) algorithm was assigned to method 1, Geometric Mean (GM) is now method 2, and Median exposure is method three). Method four represents the data from the Liu, 1997 study.

Figure 4.2 – Odds Ratios, by Highest Task Silica Exposure From Using Zero and Imputation Methods



4.3. Imputation Methods Validation

As with any estimation method, there could be over, and under estimation of the exposure. An attempt was made to determine which method(s) was better at estimating exposure for tasks without measurements and each method's validity.

Several comparisons of the measured and imputed values were conducted. Comparisons were made using qualitative methods, which included comparing results from each method as well as comparing imputed values with measured relevant similar tasks.

An analysis of variance (ANOVA) was used to test for a significant difference ($\alpha=.05$) in the estimated silica exposures from the current study and measurements from the North Carolina Dusty Trades Program. The results from the ANOVA indicated that the results from the current study were not significantly different from the measurements. The mean silica exposure from the measurement data set is 7.6 mppcf ($n=9,492$), and mean silica exposure estimates from the geometric mean, median exposure, and EM algorithm imputation methods are 8.1 ($n=331$), 9.8 ($n=354$), and 5.4 (269) mppcf respectively. However, the results were significantly different from the previous study by Liu (1997). The mean silica exposure from the Liu study was 27.4 mppcf ($n=888$).

5.0. Discussion

In this investigation, an attempt was made to determine practical and meaningful procedures to estimate missing exposures in the North Carolina Dusty Trades data, which were used to test the hypothesis that there is a significant difference ($\alpha = 0.05$) in the risk of developing silicosis between those characterized by highest task imputed silica exposure and those characterized by using zero as the exposure for tasks without measurements. This undertaking is unique in that cumulative exposure is the usual choice of exposure metric in determining the exposure-response relationship for a chronic disease such as silicosis.

Choosing the appropriate way to estimate missing data is not an easy task. Some investigators use a measure of central tendency, such as the arithmetic mean or median. Investigators commonly calculate an exposure category mean and then apply this value to each individual with a missing value. In this study, several approaches were considered and investigated.

The original research protocol included development of three methods to estimate missing data for the North Carolina Dusty Trades exposure data sets. The imputation methods outlined in the original research plan involved using a mean exposure, the fifth percentile of the exposure distribution and a statistical model to estimate silica exposure for unsampled tasks.

The first method using the geometric mean exposure for imputing commodity-specific tasks provided a central tendency estimate for the missing data and represented a balanced approach to estimating exposure from various tasks that included a range of potential low to high exposures. The second method, the fifth percentile of the exposure distribution, was not a feasible option since unsampled tasks for very low exposures were limited. Based on this observation, it was highly unlikely that these estimated exposures for unsampled tasks would be used in the exposure response determination.

After investigating alternative methods of imputation, a decision was made to replace method two. As an alternative, the median exposure for all company/site exposures for commodity-specific tasks rather than the fifth percentile was used. The median was chosen because it would not be affected by extreme values that could bias the exposure estimates.

The third method encompassed calculating maximum likelihood estimates using an expectation maximization (EM) algorithm to estimate the missing data points. Estimation of the missing data points was conducted in two steps. A program was written to identify tasks without measurements for imputation and then the mean values of the natural logarithm were calculated for each individual commodity-specific task with measurements. The mean values were included in a multiple regression model as the dependent variable and regressed against the independent variables of commodity and task.

The calculated expected value for each commodity-specific task from this model was used to estimate exposure for the unsampled tasks. A total of 15 iterations of the EM algorithm were needed before there was convergence of the expected values in all commodities.

Other approaches considered included taking the 75th and 95th percentile; however, since the data distribution includes data that can be considered representative of worst-case sampling, using these values to estimate missing data points might bias the results.

Other alternatives for estimation include using qualitative procedures where exposure is given a categorical value such as low, medium or high. In the current study, a qualitative method was employed to estimate exposure for tasks where there were no measurements. The method involved using an instrument to group tasks where measurement data were available into categories based on the potential for dust exposure. However, this approach provided unrealistic exposures for some tasks and was revised to include measurement data for similar relevant tasks to facilitate the categorization of the tasks without measurements. Unrealistic exposures were determined based on input from individuals who have dusty trades work experience.

Another approach that was considered included using company measurements to estimate exposures for company/site-specific tasks. However, this approach would not have been feasible for all companies, since 46% (282/610) of companies had no measurements. To conduct a comprehensive characterization and categorization of every company without measurements, which would allow an accurate estimation of exposure, would have been impractical during this investigation. Future research could involve developing an exposure profile for each company based on commodity, location (e.g., county in which company was located), dates of operation or years that workers were employed, jobs and tasks performed, engineering controls, and use of personal protective equipment. Then, companies without measurements could be placed in categories with similar companies that have measurements.

Of course, one important question is how well did the methods estimate the missing values. A comparison of the results from the three methods to determine the accuracy of the exposure estimates was performed by using highest task exposure results from the previous study by Liu (1997), and measurements from the North Carolina Dusty Trades data sets. The significant difference in the results from the current study and the previous study by Liu could be due to the approaches used to estimate the actual means. Other ways to test the validity of the methods would be to estimate missing data from a similar data set to see if the estimated values are within the range of expected measurements. Since most of the estimates imputed in the current study were for tasks that would be expected to have low exposures, consideration should be given to using the methods on other data sets that would need exposure estimation for a wide range of tasks.

When a comparison of the odds ratios from each method was conducted, only small differences were observed between the outcomes for each method. This might be due to the fact that imputed values from the three data sets are very similar. The results from a t-test performed to determine if the methods were significantly different indicate there are no significant differences. Comparison of the odds ratios indicates no significant difference for the imputed values and measurement data. The comparable number of subjects in the referent group for each method could explain the similarities for the odds ratios for this study as well as those shown by Liu (1997).

6.0. Conclusions

In conclusion, the findings from the present study appear to suggest that imputed silica exposures for highest exposure tasks did not improve the exposure-response relationship. However, the results from this study show an increase in the risk of developing silicosis with an increase silica exposure. This finding is consistent with the previous investigation. The consistency of results shown in this study with the previous study where no imputation procedure was performed is most likely due to the number of subjects per exposure group. Some subject's were assigned to a different category. However, it is difficult to determine if subjects were assigned to a different category because of imputed values or simply because different category ranges were used to group the results from the imputation methods and when zero was used to estimate exposure.

Despite the fact that the imputed silica exposures for the highest exposure tasks did not improve the exposure-response relationship, the results are consistent. These results may indicate that peak exposures may be important in the development of silicosis. Further research is needed to fully understand peak silica exposures and the development of chronic silicosis.

Acknowledgements

I would like to thank my dissertation committee members for sharing their knowledge and guidance during the course of this research.

This research was funded by the National Institute for Occupational Safety and Health (NIOSH) under Grant Number R03 OH03775.

References

- Checkoway H, Rice CH: Time-Weighted Averages, Peaks, and Other Indices of Exposure in Occupational Epidemiology. *American Journal of Industrial Medicine* 21: 25-33, 1992
- Checkoway H, Heyer NJ, Demers PA, Breslow NE: Mortality Among Workers in the Diatomaceous Earth Industry. *British Journal of Industrial Medicine* 50: 586-597, 1993
- Cody R P, Smith J K: *Applied Statistics and the SAS Programming Language*. Fourth Edition, Prentice Hall, 1997
- Corn J K: Historical Aspects of Industrial Hygiene II. Silicosis. *American Industrial Hygiene Association Journal* 41: 26-33, 1980
- Crofton SJ, Douglas A: *Respiratory Diseases*. Third Edition. Blackwell Scientific Publications, 1981
- Davis GS: Silica. In *Occupational and Environmental Respiratory Disease*, (eds. P. Harber, et al.), Mosby, pp 373-399, 1996
- Dement JM, Harris RL, Symons MJ, Shy CM: Exposures and Mortality Among Chrysotile Asbestos Workers. Part I: Exposure Estimates. *American Journal of Industrial Medicine* 4: 399-419, 1983
- Hessel PA, Sluis-Cremer GK, Hnizdo E: Case-Control Study of Silicosis, Silica Exposure, and Lung Cancer in White South African Gold Miners. *American Journal of Industrial Medicine* 10:57-62, 1986
- Kleinbaum DG: *Logistic Regression: A Self Learning Text*. Springer-Verlag, 1994
- Landrigan PJ: Silicosis. *Occupational Medicine* 2:319-26, 1987
- Little RJA, Rubin DB: *Statistical Analysis with Missing Data*. John Wiley and Sons, 1987
- Liu G: *Assessment of the Association Between Development of Silicosis and the Highest Silica Exposure Tasks*, Masters Thesis, University of Cincinnati, 1997
- Lynch JR, Ayer HE: Measurement of Dust Exposure in the Asbestos Textile Industry. *American Industrial Hygiene Association Journal* 27: 431-437, 1966
- National Institute for Occupational Safety and Health: *Work-Related Lung Disease Surveillance Report*. DHHS (NIOSH) Pub. No. 2000-105, NIOSH, 1999.
- OSHA: News Release-Labor Secretary Calls for an End to Silicosis. U. S. Department of Labor, Office of Public Affairs, October 31, 1996

Ramazzini B: Diseases of Workers. trans. By William Cave Wright, 1713, New York Hafner Publishing Company, 1964

Rappaport SM: Selection of the Measures of Exposure for Epidemiology Studies. Applied Occupational Environmental Hygiene 6: 448-457, 1991

Rice CH: Exposure Reconstruction and Study of Silicosis in North Carolina, Doctoral Dissertation, University of North Carolina at Chapel Hill, 1983

Rice C, Harris R, Lumsden J, Symons M: Reconstruction of Silica Exposure in the North Carolina Dusty Trades. American Industrial Hygiene Association Journal 45:689-696, 1984

Rice C, Harris RL, Checkoway H, Symons MJ: Dose Response Relationships for Silicosis from a Case-Control Study of North Carolina Dusty Trades Workers. In: Silica, Silicosis and Cancer: Controversy in Occupational Medicine, (eds. Goldsmith DF, Winn DM, Shy CM), Philadelphia, 1985

Rice C, Harris RL, Symons MJ, Checkoway H: Factors Affecting Latency of Silicosis in North Carolina. Annals of Occupational Hygiene 32: 673-680, 1988

Rice CH, Checkoway H, Dosemeci M, Stewart P, Blair A: Effects of Exposure Estimation Procedures on the Evaluation of Exposure-Response Relationships for Silicosis. Annals of Occupational Hygiene 41: 485-490, 1997

Rosenman KD, Reilly MJ, Rice C, Hertzberg V, Tseng CY, Anderson HA: Silicosis Among Foundry Workers Implication for the Need to Revise the OSHA Standard. American Journal of Epidemiology 144: 890-900, 1996

SAS: SAS User's Guide: Statistics, SAS Institute Inc, 1996

Siemiatycki J, Wacholder S, Dewar R, Wald L, Begin D, Richardson L, Rosenman K, Gerin M: Smoking and Degree of Occupational Exposure: Are Internal Analyses in Cohort Studies Likely to Be Confounded by Smoking Status? American Journal of Industrial Medicine 13: 59-69, 1988

Anticipated Publications

- 1. Evaluation of the Exposure-Response Relationship for the Estimated Highest Task Silica Exposure and Silicosis.**
- 2. A Comparison of Imputation Methods Used to Estimate the Highest Task Silica Exposure for the North Carolina Dusty Trades.**

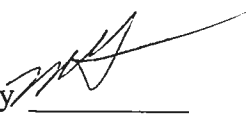


DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service
Centers for Disease Control
and Prevention (CDC)

Memorandum

Date: December 12, 2001

From: Michael J. Galvin, Jr., Ph.D., Lead Program Activity 
Office of Extramural Programs, NIOSH, D30

Subject: Final Report Submitted for Entry into NTIS for Grant 5 R03 OH003775-02.

To: William D. Bennett
Data Systems Team, Information Resources Branch, EID, NIOSH, P03/C18

The attached final report has been received from the principal investigator on the subject NIOSH grant. If this document is forwarded to the National Technical Information Service, please let us know when a document number is known so that we can inform anyone who inquires about this final report.

Any publications that are included with this report are highlighted on the list below.

Attachment

cc: Sherri Diana, EID, P03/C13

List of Publications

NIOSH Extramural Award Final Report Summary

Title: Estimation of Highest Task Silica Exposures
Investigator: L. Faye Grimsley
Affiliation: University of Cincinnati
City & State: Cincinnati, OH
Telephone: (513) 558-1843
Award Number: 5 R03 OH003775-02
Start & End Date: 9/30/1998–3/31/2001
Total Project Cost: \$71,907
Program Area: Asthma & Chronic Obstructive Pulmonary Disease
Key Words:

Abstract:

More than one million workers in the United States are exposed to silica and approximately 100,000 will develop lung diseases. Despite efforts to decrease silica exposure in the workplace, overexposures and deaths from silicosis continue in certain operations and industries. According to the National Institute for Occupational Safety and Health (NIOSH), each year more than 250 workers die from silicosis.

The objectives of the study were to investigate ways to estimate silica exposure in unsampled tasks and to evaluate the relationship between highest task silica exposure and the development of silicosis. Data from the North Carolina Dusty Trades, a unique resource for exposure-response studies of the effects of silica exposure is the basis for the study. Three methods of imputation were used. The methods included estimating exposures based on the geometric mean exposure at sampled commodity-specific tasks, the median exposure of commodity-specific tasks, and a maximum likelihood estimation statistical model. A case-control study design was used to evaluate the effect that estimated silica exposure from the highest task had on the development of silicosis.

The results indicate that the highest task silica exposure is related to the development of silicosis. Exposure estimates including values imputed using the geometric mean resulted in an overall odds ratio and (95 % CI) of 2.27 (1.49, 3.44), for the median imputation exposure method the overall odds ratio was 2.27 (1.49, 3.44); and overall odds ratio of 2.14 (1.47, 3.11) for the statistical model. These odds ratio results were not significantly different from overall odds ratio of 2.22 (1.52, 3.26) that was calculated using zero for missing data in the highest task.

These results show that the exposure-response relation based on the highest task exposure metric did not increase when non-zero values were imputed using three methods for tasks never sampled. The results are consistent, however, suggesting that peak exposures may be important in the development of silicosis. Further work is needed to better understand the relation between peak exposure and the development of silicosis.

Publications

No publications to date.