Occupational Cohorts:Confounder & Effect-modifier Models

Principal Investigator: Xiaonan Xue, Ph.D.
Division of Biostatistics
Department of Epidemiology & Population Health
Albert Einstein College of Medicine
Bronx, NY 10461
Tel. 718-430-2431
Fax. 718-430-8570
Email. Xiaonan.xue@einstein.yu.edu

Co-investigator: Roy E Shore, Ph.D.
Vice Chairman and Chief of Research
Radiation Effects Research Foundation
Hiroshima, Japan

Grant Number: 5 R01 OH008087

Project Starting Date: 9/1/2004

Project Ending Date: 12/31/2008

Abstract

Epidemiologic studies of cohorts exposed occupationally to carcinogens or other toxicants often have several features that complicate statistical analyses of the data: longitudinal data, repeated measures, cumulative exposures; certain covariates are substantially correlated with cumulative exposure; health endpoints of interest often have other potential confounding or effect-modifying exposures; and the effect of exposure measurement errors may be significant. Statistical issues related to several of these features have not been dealt with satisfactorily. In this proposal, several statistical methods were developed to improve upon existing methods for analyzing occupational cohort data by minimizing these limitations. Particularly, we proposed the use of ridge regression as an alternative method when multicollinearity exists among covariates; we examined how different types of measurement errors affect the estimation of risk associated with repeated occupational exposure to radiation, while controlling for possible confounders such as time since exposure. We have showed that random measurement error has a larger impact on the relative risk estimates compared to the error due to the Minimum Detection Level (MDL) of the dosimeter. We developed new statistical methods to design efficient cohort study and methods to estimate and test for effect modification when traditional methods failed. These methods were developed with a focus on the goal of reaching a non-statistical audience therefore can be easily adopted by epidemiologists and clinicians.

Table of Contents:

Section 1 Significant Findings and Their Impacts to Occupational Health Studies **4**

Section 2 Scientific Report and Publications **5**

Specific Aim 1 7

Specific Aim 2 11

Specific Aim 3 15

Specific Aim 4 28

References 31

Publications **33**

Inclusion Criterion **33**

Material available for other investigators **34**

Section 1. Significant Findings and Their Impacts to Occupational Health Studies

Epidemiologic studies of cohorts exposed occupationally to carcinogens or other toxicants often have several features that complicate statistical analyses of the data: longitudinal data, repeated measures, cumulative exposures, and often the exposure rates, change with time; certain covariates are substantially correlated with cumulative exposure; health endpoints of interest often have other potential confounding or effect-modifying exposures; and the effect of exposure measurement errors may be significant. Statistical issues related to several of these features have not been dealt with satisfactorily. The main goal of this proposal is to develop statistical methods that will improve upon existing methods for analyzing occupational cohort data by minimizing these limitations.

First, we examined how the estimation of risk associated with radiation exposure can be affected by potential confounders such as years of exposure, attained age, which are correlated with cumulative radiation exposure. We demonstrated that collinearity between variables can induce unstable estimates of covariate effects and may speciously mask or amplify the putative radiation effect in the analysis. We proposed the use of ridge regression as an alternative method when multicollinearity exists among covariates. The method was applied to the ORNL study to estimate the risk for all-cause mortality. We concluded that the previous finding that occupational radiation exposure is not related to risk of death remains valid after taking into account collinearity.

Second, we examined how different types of measurement errors affect the estimation of risk associated with repeated occupational exposure to radiation, while controlling for possible confounders such as time since exposure. We showed that random measurement error has a larger impact on the relative risk estimates compared to the error due to the Minimum Detection Level (MDL) of the dosimeter. Therefore, it is essential to assess the impact of random error and correct for it appropriately so as to obtain an unbiased estimator of risk. We applied the methods to the ORNL study and demonstrated that after proper adjustment of measurement error, there is no significant relationship between radiation exposure and mortality. However, we suspect that smoking can potentially intensify the effect of radiation. Since no information on smoking was obtained in the ORNL study, there is still some uncertainty with the risk estimation.

Third, since this is a methodological grant, we also developed new statistical methods needed for occupational radiation cohort studies, which are often studies with a long follow-up, repeated visits and rare outcomes such as cancers and deaths. Particularly, we developed method to design efficient cohort study and method to estimate and test for effect modification when traditional methods failed. These methods were developed with a focus on the goal of reaching a non-statistical audience therefore can be easily adopted by epidemiologists and clinicians.

Section 2. Scientific Report and Publications

Epidemiologic studies of cohorts exposed occupationally to carcinogens or other toxicants often have several features that complicate statistical analyses of the data: repeated visits often with a long follow-up, cumulative exposures over a long period time to carcinogens or other toxicants, and the exposure rates can change with time; certain covariates are substantially correlated with cumulative exposure; health endpoints of interest often have other potential confounding or effect-modifying exposures; and there are many sources of measurement errors are associated with the exposure and the effect of exposure measurement errors may be significant. Statistical issues related to several of these features have not been dealt with satisfactorily. The main goal of this proposal is to develop statistical methods that will improve upon existing methods for analyzing occupational cohort data by minimizing these limitations and also apply the developed methods to a cohort of Oak Ridge National Laboratory (ORNL) workers (Gilbert, 1993; Frome, et al, 1997) to re-examine the risk associated with occupational radiation exposure.

ORNL is a US Department of Energy (DOE) research and development facility that began operation in Oak Ridge, Tenn, in 1943. It is one of several facilities included in a large follow-up study of the health and mortality of workers at DOE facilities (Gilbert, 1993b). This cohort contains 27,982 deaths among 106,020 persons employed at four Federal nuclear plants in Oak Ridge between 1943 and 1985 (Frome, et al, 1997). The main objective was to evaluate the health effects of occupational radiation exposure. For the reasons discussed in Frome, et al (1997), our analysis will be restricted to white males

ever employed at X-10 and Y-12 sites, two nuclear facilities in ORNL. The other facilities are not included. The data then reduces to 28,770 white males. The other race and gender groups at these facilities are not included because their numbers are too small for meaningful analysis; they would contribute little to the assessment of effects of low-level radiation and would unduly complicate the dose-response analysis. The following variables will be considered as potential confounders in the dose response analysis: socioeconomic status indicated by paycode (monthly or non-monthly pay period), birth year, age at employment, job type and facility as well as time-dependent variables such as length of employment, time since first exposure and attained age; among them, age at employment, job type, length of employment, time since first exposure and attained age will be considered as potential effect modifiers.

Four specific aims were proposed in the proposal. In below we describe works accomplished and findings for each specific aim.

Specific Aim 1. Develop methods to assess the degree of collinearity and develop methods to adjust for collinearity among confounders and exposures.

In linear and logistic regression models, the ridge regression estimator has been applied as an alternative to the maximum likelihood estimator in the presence of collinearity. The advantage of the ridge regression estimator over the usual maximum likelihood estimator is that the former often has a smaller total mean square error and is thus more precise. We generalized this approach for addressing collinearity to the Cox proportional hazards model. Specifically, we define the ridge estimator of the exposure effect to be the

estimator with minimum length among those that increase the weight sum of square error by a small amount. The ridge estimator can be calculated based on the maximum likelihood estimator and the estimated second derivative of the log partial likelihood function, which can be easily obtained using standard statistical software.

We used simulation studies to evaluate the performance of the ridge estimator. The simulations have indicated that the ridge estimator had smaller bias and mean square error and higher predictability compared to the maximum likelihood estimator (MLE).

We applied the proposed ridge estimator for the Cox model to the ORNL cohort. The main objective is to evaluate the health effects of occupational radiation exposure (Wing, et al., 1991). In this paper, we considered a subset of 2,134 workers who entered the factory after 1957 and had worked there for at least a year. The reason that only the workers who started after 1957 were considered is to ensure the quality of dose estimates: in earlier years the dose estimates were subject to heavy censoring due to the minimum detection limit of the dosimeter (Xue, et al. 2004). The radiation exposure varies largely between subjects as well as within subjects over time. Because of the non-constant exposure rate, cumulative exposure has been used to measure radiation exposure (Wing, et al., 1991, 1993; Frome, et al., 1997; Cardis, et al., 2005). It takes many years for the radiation exposure to manifest its carcinogenic effect. Previous analyses have shown that a 20-year lag for cumulative radiation dose fits the data better than shorter lag assumptions (Wing, et al, 1993; Frome, et al., 1997 and Cardis, et al, 2005). Therefore, a 20-year lag was used in this paper. A Cox model was used to evaluate the association of

radiation exposure with all-cause mortality, in which the cumulative radiation exposure was treated as a time-dependent variable. The following variables were considered as potential confounders in the dose response analysis: (1) age at exposure divided into three categories: below 30, 30-39 and above 40; (2) birth year divided into three categories: before 1930, 1930-1939 and 1940 or after; These two variables have been conventionally treated as categorical variables instead of continuous ones in the radiation epidemiological literature as from a biological perspective it is too restrictive to assume a linear association between these two variables and risk for mortality (Gilbert, 1989; Gilbert, et al., 1993a,b and Frome, et al., 1997) (3) socio-economic status for which paycode was used as a surrogate marker: monthly paid vs non-monthly paid with the monthly paid to be a higher socioeconomic category; (4)number of years exposed to radiation exposure as a continuous time-dependent variable; (5) active worker status (currently employed) was also included as a time-dependent binary variable to take into account the healthy worker effect because workers who continued employment, and consequently remained exposed, tend to be healthier. Data on smoking, chemical exposures, medical exposures to ionizing radiation, and cancer mortality were not available. A Cox model was applied to the data set and based on the maximum likelihood estimates for every additional 10mSV dose of radiation exposure increases the risk of death by 4%, however, the effect is not statistically significant. The distribution of radiation exposure is skewed with most of them being zero or very small doses and a few with large doses. The cumulative dose difference between the 90th percentile and the 10th percentile is about 50mSv (Wing, et al, 1991), which results in an estimated relative risk for mortality of 1.2. Age below 30 at initial exposure is associated with the highest

risk for death. The risk decreases dramatically if exposure occurred at a later age. Birth before 1930 is associated with the highest risk and birth after 1930 has a much lower risk. High social economic status (monthly paid) is a protective factor. Active worker (currently employed) is associated with less risk of death than nonactive workers with borderline significance. Risk increases 2% for every additional year of exposure given the same level of cumulative exposure. These findings are consistent with what has been seen in the literature (Wing, et al, 1993; Frome, et al., 1997 and Cardis, et al, 1995).

However, the potential collinearity issue in occupational radiation exposure studies has not been addressed previously. Years of exposure is likely to be positively correlated with cumulative exposure. Currently-employed status also tends to be correlated with cumulative dose and years of exposure. The sample correlation between variables was calculated for each year between 1977 to 1984 (since a lag of twenty years was used for the exposure variable) and averaged over these years. There is a strong correlation between current employment status and years of exposure (0.81), although the primary exposure variable, the cumulative radiation exposure, is not highly associated with any other covariates. There are a total of 103 deaths in this sample. With eight covariates in the model, collinearity can have a big impact on the estimators. We therefore applied the ridge regression showed that the ridge estimator for the relative risk associated with cumulative radiation exposure is very close to the MLE. As the ridge parameter increases, the relative risk estimate is reduced by up to 6% only. The high consistency in the estimates of the relative risk associated with radiation exposure between the ridge estimation and the MLE may result from the fact that the cumulative

radiation exposure is not highly associated with any of the other covariates. The ridge confidence intervals are narrower but all include 1. The predicability is very close to that of the MLE and is even slightly higher for larger ridge parameters, indicating that in this case the ridge estimation does not offer an improvement in predictive ability over the MLE. In summary, we conclude that the previous finding of no association between occupational radiation exposure and all-cause mortality remains valid after adjusting for collinearity between covariates.

Specific Aim 2. Develop an empirical Bayesian approach to evaluate effect modification. To evaluate effect modification for repeated measurements, often people use the generalized estimating equation (GEE) ((Liang & Zeger, 1986) because GEE takes into account possible correlations between visits from the same subject. However, when we examine risk for rare events such as certain type of cancer, the traditional GEE approach may either fail to converge or fail to estimate risk accurately. In this paper, we propose a Monte-Carlo approach to empirically estimate the significance of the exposure-disease association as well as its effect modification.

To estimate the risk disease association, first, we proposed to use a linear regression to model the risk. Specifically, $P(Y_{ij}=1)=\beta_0+\beta_1 x_{ij}$ where X is the exposure variable and Y is the binary outcome variable (0 or 1), and i represents subjects and j represents visits. Under this model, β_1 represents the exposure-disease association. The convergence of a linear regression model is much easier to achieve as compared to a logistic regression model because the parameter space for an attributable risk is within (-1, 1). When events are very rare, the estimate of the attributable risk is close to 0 and is not near the

boundary of its parameter space. Therefore, the linear regression is much easier to converge under rare events compared to the logistic regression. However, statistical inference for β_1 in the linear GEE model relies on the normal approximation for $\hat{\beta}_1$ which may be questionable. To avoid this problem, we propose a permutation approach to obtain an empirical p-value associated with H_0: $\beta_1 = 0$. The procedure of the permutation test is described in the following procedure:

1. Compute $\hat{\beta}_1$ in the above linear regression model;
2. Permute Y across subjects and visits;
3. Calculate β_1 based on the permutated data: $\hat{\tilde{\beta}}_1$;
4. Repeat steps 2-3 for 1000 times;
5. Empirical two-sided p-value=percent of times the magnitude of $\hat{\tilde{\beta}}_1$ is greater than or equal to the magnitude of $\hat{\beta}_1$.

To test effect modification, for simplicity, here we assume a binary effect modifier, we propose to use the linear regression model: $P(Y_{ij} = 1) = \beta_0 + \beta_1 x_{ij} + \beta_2 w_{ij} + \beta_3 x_{ij} w_{ij}$. The following Monte-Carlo procedure is developed:

1. Compute $\hat{\beta}_3$ from the above linear regression;
2. Permute the vector of (x, w) across subjects and visits;
3. Calculate β_3 based on the permutated data: $\hat{\tilde{\beta}}_3$;

4. Repeat steps 2-3 for 1000 times;
5. Empirical two-sided p-value=number of times the magnitude of $\hat{\tilde{\beta}}_3$ is greater than the magnitude of $\hat{\beta}_3$.

Simulation studies were used to evaluate the performance of these two procedures. The first set of simulations was conducted to evaluate the validity and power of various tests for the exposure effect when the binary outcomes are rare and correlated within subjects. The results, summarized in Table 2.1, show the proportion of times the null hypothesis was rejected with the binary GEE model, the linear GEE model and our proposed permutation test. For the GEE models, we report the proportion of null hypotheses rejected among the simulated data sets in which the model converged as well as the proportion of null hypotheses rejected among all the data sets, including those in which the model failed to converge. Therefore, in the latter case, we assume that the data sets in which the GEE model did not converge failed to demonstrate a significant allele effect. Under the null, the proportion of null hypotheses rejected reflects the validity of the test. Table 2.1 shows that all the tests are valid since the rates of rejecting the null hypothesis under the null are all well below 5% with the permutation test being the least conservative. To compare the power of these three tests, we examine the rate of rejecting the null under the alternative. Table 2.1 indicates that the permutation test has higher power than either the binary or linear GEE models. Between the two GEE models, the linear GEE model has a much better convergence rate than the binary GEE model but lower power among all simulated data sets. The discrepancy in power between the two GEE approaches is due to differences in the way the effect size is measured. The effect

size for the linear model which is measured by the difference between event rates (i.e., attributable risk). The effect size for the binary model is measured by an odds ratio. Therefore, the effect size for the logistic regression is often much "higher" than that for the linear regression when event rates are low, which resulted in a higher power. As expected, all these tests have higher power with larger effect sizes and smaller correlations among the repeated measures.

Table 2.1. Percent of the null exposure effect (X) rejected under various models when P_0=0.001, n=400 and ρ=0.2 or 0.5 (correlation between repeated visits) with 500 simulations

	Binary GEE			**Linear GEE**			**Permutation test**
	OR	Among converged (%converged)	Among total	*D*	Among converged (%converged)	Among total	
ρ=0.2	1.0	1.6 (39)	0.62	0.000	1.1 (88)	0.97	2.6
	4.0	22.6 (73)	16.5	0.003	4.9 (99)	4.9	36.6
	10.0	66.8 (80)	53.4	0.009	48.8 (100)	48.8	81.2
ρ=0.5	1.0	4.2 (23)	0.98	0.000	0.0 (71)	0.0	2.8
	4.0	15.3 (47)	7.2	0.003	0.2 (90)	0.2	22.0
	10.0	39.4 (57)	22.5	0.009	11.4 (98)	11.2	62.2

The objective of the second set of the simulations was to evaluate the validity and power of various tests for the exposure variable (X) and an effect modifier (W) interaction effects when the binary outcomes are rare and correlated within subjects. The results are summarized in Table 2.2. We first examined the validity of each test. Under the null, Table 2 shows that both the linear GEE model and the permutation test are valid but the linear GEE model is more conservative. Since the interaction was defined under an additive model, a null interaction effect in an additive model does not correspond to a null interaction effect in a multiplicative model. Therefore, we could not evaluate the significance level for the binary GEE model in this setting. However, since the binary

GEE model failed to converge in over 90% of the simulated data sets, it is not considered a feasible alternative for testing the interaction effect when events are rare. Compared to the linear GEE model, the permutation test has slightly higher power. Since the results in Table 2 were similar for ρ=0.2 and ρ=0.5, the simulations for the interaction effect were conducted for ρ=0.2 only.

Table 2.2. Percent of the null interaction effect rejected with n=200 clusters and P(X=1) = 40%; P_{00} = 0.0005 and ρ=0.2 and 500 simulations

	Binary GEE			**Linear GEE**		**Permutation**
OR_1 / OR_0	Among converged (%converged)	Among total	D	Among converged (%converged)	Among total	
0.75	0.0(6)	0.0	0.000	1.2(96)	1.2	2.4
2.00	2.3(9)	0.2	0.003	4.2(100)	4.2	10.6
5.75	7.1(8)	0.6	0.010	48.4(100)	48.4	57.2
23.0	22.9(10)	2.3	0.019	82.6(100)	82.6	88.0

Specific Aim 3. Develop methods to evaluate and correct for various types of measurement errors.

Occupational exposures are subject to several types of measurement errors. We considered two of the most common types of measurement error associated with occupational exposures: the error due to below minimum detection level (MDL) doses and random measurement error. Doses are often recorded as zero when the exposure is below the minimum detection level (BMDL). BMDL values entered as zero lead to an underestimation of the true exposure and can result in either an overestimate or underestimate of risk associated with the exposure. Random measurement error leads to

an inefficient and attenuated estimate of risk associated with exposure. However, the levels of bias and inefficiency that can result from the simultaneous presence of both types of measurement error have not previously been studied. In addition, the impact of these measurement errors on the type I error and type II error for an exposure-response effect is unclear. Since the magnitude of the random error associated with cumulative exposure may vary with individuals and across time within an individual, traditional methods to correct for random measurement errors are not applicable here. Further, correcting errors for minimum detectable levels and random errors simultaneously is too complex for analytical solutions. Therefore, we used simulation studies to quantitatively evaluate the magnitude of the bias, inefficiency, type I and type II errors associated with them.

In the simulation, two scenarios were considered: (1) the exposure effect exists and $\beta = 0.02$ per 10 mSv; (2) no exposure effect, i.e., $\beta = 0$. The baseline hazard function is assumed to be dependent on the individual's attained age with a log of relative risk of 0.089 per year increase, as obtained from the ORNL study (Xue, et al, 2004). The age at entry is simulated from a uniform distribution ranging between 25 and 40, comparable to the range in the ORNL study. For simplicity, we assume each subject has 20 years of radiation exposure and there were no other covariates except the exposure variable. Then based on the true dose level and an assumption of a 5-y lag, we generate survival times based on a Cox model for each scenario. The number of subjects is chosen to be N=200. A small sample size is chosen here so that it will be computationally feasible to run many iterations. Because of the small size of the simulated data, every

subject is followed until death so that the study has 83.6% power to detect an effect of $\beta = 0.02$ (based on simulation results).

Within each scenario, three levels of lognormal random measurement errors were generated: small, medium and high. We used the uncertainty factor K to represent the magnitude of random errors and K is set to be 1.5, 2.0 or 4.0, respectively. The interpretation of K is the following: 95% of the observed doses measured with random errors, Z, will fall between Y/K to KY where Y is the true dose. For example, if Y=1.0 mSv, then 95% of the observed doses will fall between 0.67 and 1.5 mSv when K=1.5, between 0.5 and 2.0 mSv when K=2.0, or between 0.25 and 4.0 mSv when K=4.0. Therefore, the uncertainty levels considered here should cover a majority of plausible values that have a nonignorable impact on risk estimates. After adding the random errors to the true exposure, we then applied the MDL accordingly. Doses below the MDL were set to be zero. The observed doses were used to obtain the estimated β.

The simulation was repeated 500 times. We summarized the simulation results for scenario (1) by calculating the relative bias, defined as the mean of the estimated β relative to the true β, and the mean of the estimated standard error of β based on the observed dose relative to that based on the true dose. We also calculated the proportions of the nominal 95% confidence intervals of β not containing zero and containing the true β. The first proportion is interpreted as the power of the test of a null exposure effect and the second proportion is interpreted as the coverage probability of the nominal 95% confidence interval. For the second scenario when the exposure effect is null, we

calculated the average of the estimated β and the proportion of the 95% confidence intervals containing zero. This proportion is interpreted as the coverage probability of the nominal 95% confidence interval as well as one minus the type I error associated with the test of a null exposure effect.

1) With An Exposure Effect ($\beta > 0$)

The relative bias was plotted against the uncertainty factor K in Figure 1(a). The plot suggests that with the random measurement error and with or without the error due to MDL, the estimated β is biased towards null (the relative bias is less than 1). As the degree of random error goes up, the amount of bias goes up; the amount of bias also goes up as the error due to MDL increases. However, the random error has a stronger effect on the estimate of the relative risk compared to the MDL. The standard errors are also biased downward but at a less degree (results not shown).

To see how the errors affect the estimate of β when doses were measured weekly, the weekly doses were sampled from a lognormal distribution such that the annual dose approximately follows a lognormal distribution. A series of weekly readings was added to form the estimate of the annual dose. Weekly multiplicative random measurement errors were generated using the same values of K as the annual readings. The sum of the observed weekly doses is a better estimate of the true annual dose than the observed annual dose when the film badge was monitored annually. This occurs because the random errors from multiple measurements tend to cancel each other out.

Thus, the impact of random measurement error is smaller comparing to the annual monitoring case. The MDLs for weekly doses were chosen so that the chances of missed doses were 0% and 90% and 95% on average, respectively. With $\beta = 0.02$, as shown in Figure 1 (b), with only random error, when the level of random error is low, there is little bias in the estimate of the log of the relative risk. When the level of random error becomes medium or high, the log of the relative risk is biased downward by 12% and 46%, respectively. When both the random measurement error and censoring due to minimum detection level exist, even with a very high proportion of BMDL censoring (from 90% to 95%), the bias remains smaller compared to that for the annual readings. With a small level of random error, as indicated in Figure 1 (b), the log of the relative risk is overestimated from 7% to 10% as the censoring proportion changes from 90% to 95%. With a medium or a high level of random error, β is underestimated because the underestimation from random error fortuitously offsets the overestimation from censoring due to BMDL.

We also used simulations to examine the impact of errors on the inferences of β. Table 3.1 summarizes the power when doses were read annually. As we mentioned earlier, without any errors, the study has 83.6% power to detect β of 0.02. With the inclusion of random error, the power decreases. The power also decreases as the censoring due to MDL increases. The impact of the random error on the power is larger than that of the error due to MDL. The coverage probability of the true β is less than 95%, indicating that the 95% confidence interval obtained from the regular Cox model is too narrow. Again, the coverage probability decreases as the level of random error

increases and the error due to MDL increases, but the random error has a larger impact on the coverage probability than the error due to MDL does.

Table 3.1. Properties of statistical inference on β when doses in annual readings are measured with multiplicative random measurement error and error due to minimum detection level (result obtained with N=200 subjects 500 simulations, β=0.02)

MDL (% of below MDL doses)	Uncertainty factor	% of 95% CI does not include 0 (power)	% of 95% CI include true beta
0 (No MDL)	1.5	79.0%	89.4%
	2.0	74.9%	79.1%
	4.0	61.5%	36.1%
33 (50%)	1.5	77.2%	82.2%
	2.0	74.1%	69.9%
	4.0	61.1%	31.3%
65 (75%)	1.5	75.9%	73.1%
	2.0	67.0%	69.4%
	4.0	25.7%	57.0%

2) With No Exposure Effect ($\beta = 0$)

We repeated the simulation with $\beta = 0$. When $\beta = 0$, Table 3.2 shows that neither the magnitude of the random error nor the level of censoring due to BMDL lead to a biased estimate of β: the mean of the estimated β is very close to 0. The errors in general lead to an underestimation of the standard error. The coverage probability of the 95% confidence interval is smaller but very close to its nominal level, indicating that the interval estimates are valid and the type I error of a null test is close to 5%.

Table 3.2. Properties of statistical inference on β when doses in annual readings are measured with multiplicative random measurement error and error due to minimum detection level (result obtained with N=200 subjects 500 simulations, β=0.00)

MDL (% of below MDL doses)	Uncertainty factor	Average of estimated beta	Ratio of estimated se with se based on true dose	% of 95% CI include true beta
0(No MDL)	1.5	0.0008	0.9612	94.2%
	2.0	0.0008	0.9305	93.8%
	4.0	0.0007	0.6869	94.6%
33 (50%)	1.5	0.0007	0.8481	94.4%
	2.0	0.0007	0.8251	94.0%
	4.0	0.0007	0.6360	94.2%
65 (75%)	1.5	0.0006	1.0035	94.4%
	2.0	0.0009	0.8978	94.4%
	4.0	0.0006	0.6268	94.8%

In summary, the simulations show that the impact of the random measurement error on the estimation and inference of the risk parameter is larger than that of the error due to a minimum detection level. Therefore, it is critical to correct for random measurement error if it exists. The impact of random measurement error on estimating the risk parameter is larger for annual readings than for quarterly and weekly readings because, for the latter, the random measurement errors tend to cancel out when the readings are added together and the sum of weekly readings therefore give a better estimate of the true annual dose than a single annual reading.

Censoring due to BMDL induces an underestimation of risk associated with exposure for annual readings, but it can lead to an overestimation of risk with weekly or quarterly readings. Similar findings were seen in Xue, et al (2004). The rationale is that

for weekly or quarterly measurements, the chance that the true exposure is below the MDL is so high that most of the subjects ended up with observed doses zero, regardless of whether the actual weekly or quarterly dose was close to (but below) the MDL or far below the MDL. Only a few high weekly or quarterly measurements ended up being observed. Therefore, the difference in observed cumulative doses between a person with a high exposure and a person with a low cumulative exposure was most likely artificially decreased so that the risk associated with the exposure was overestimated. On the other hand, for annual measurements, since the chance of BMDL doses is much lower compared to weekly or quarterly measurements, the observed doses were underestimated but not as much (i.e., the majority of them were still above zeroes) with the higher exposure being underestimated less and the low exposure being underestimated more. Therefore, the difference in observed doses between the high exposure and low exposure was artificially increased so that the risk associated with exposure was underestimated. The random measurement error on exposure in general leads to an underestimation of the risk estimator. With both errors acting simultaneously, the direction of the bias depends on which error plays a dominant role.

Further, even though the estimate of the risk effect is biased and the power of the null test is reduced by random error and error due to BMDL, the simulation shows that the type I error rate is not affected.

The simulation results were applied to a sample of historical occupational radiation exposure data from the Oak Ridge National Laboratory. In this analysis, we

considered a subset of 3,960 workers who entered the laboratory after 1945 and had worked for at least 3 years and were followed till 1984. The purpose of only considering workers who had worked for at least 3 years is to ensure adequate exposure for all the subjects so that we could evaluate the exposure measurement error models effectively. This results in 550 deaths in a total of 106,840 person years.

An individual's recorded radiation dose at ORNL was based on film badges from 1944 to 1975 and thermoluminescent dosimeters from 1975 on. Film badges were evaluated weekly from July 1944 to July 1956, when quarterly monitoring was initiated. The minimum detection level was from 0.10 mSv to 0.30mSv. Annual monitoring was initiated in 1975 using thermoluminescent dosimeters (MDL=0.20 mSv). The doses below the minimum detection levels were set to zero. However, not all the observed zero doses are BMDL doses. Some are legitimate zero doses since office workers tended to have no occupational external radiation exposure (Watkins, et al, 1997). Among those positive doses, it was estimated that about 98% of the weekly doses were BMDL doses, about 75% of the annual doses were censored due to BMDL (Xue and Shore, 2003) and about 80% of the quarterly doses were BMDL doses. Without any adjustment of the missed doses, the average person year cumulative dose is 4.58 mSv with an assumption of a 20-y lag.

In Xue, et al (2004), the data set described above was used to estimate the relative risk for all-cause mortality, with the exposure adjusted for missed doses due to BMDL using a monte carlo method. A log linear Cox proportional hazards model was used. Data

on smoking, chemical exposures, medical exposures to ionizing radiation were not available. The Cox model controlled for the sociodemographic variables age, birth year, pay code (monthly/nonmonthly) and active/inactive worker status. Pay code was used as an indicator of socioeconomic status, with monthly paid workers being at a higher socioeconomic level. Active worker status was considered because workers who continued employment, and consequently exposure, tend to be healthier. A lag of 20 years was used. Under a Cox model controlling for sociodemographic factors, the relative risk was estimated to be 1.018 (95% CI=(0.996, 1.040)) per 10 mSv increase in radiation exposure using the observed doses compared to 1.017 (95% CI=(0.997, 1.037)) using the doses adjusted for censoring due to BMDL. This provides no clear indication of a dose-response effect with respect to all-cause mortality. However, another possibility is that there is a dose effect that was masked by exposure measurement errors.

Several papers studied or addressed the issue of missing doses due to BMDL for the ORNL study (Kerr, 1993; Wing, et al., 1993; Watkins, et al., 1997; Frome, et al., 1997; Mitchell, et al., 1997). However, the magnitude of random measurement error and other types of systematic measurement error have not been examined for the ORNL study. In this paper we evaluate the dose response relationship for the ORNL study taking into account concomitantly both error caused by BMDL and random measurement error.

For the reasons stated earlier, we used simulations to evaluate the effect of random error in addition to the error due to BMDL on risk estimates. First, we generated weekly doses for 1944 to 1956, quarterly doses for 1957 to 1974 and then annual doses

for 1975 to 1984, using the dose distribution parameters described in the above simulation study. These parameters were chosen so that the simulated dose would have the same distribution as the annual exposure in the ORNL study. To ensure that the simulated subjects would have the same employment history, age distribution and censoring time distribution as the ORNL data set, we bootstrapped (sample with replacement) individuals from the ORNL subset using their hire year, entry age, job termination year and censoring time to generate our simulated data set. Since some of the annual doses in the ORNL study were legitimate zero doses because they worked in a job with no potential for exposure, we also obtained the indication of the true zero doses from the bootstrap sample and set those doses to be zero in our simulated dose data. We used the hire year and the job termination year for each individual in the bootstrap sample to define the cumulative exposure for this subject and his entry age and attained age over time to define his baseline hazard rate. His survival time was generated based on the hazard function estimated from Xue, et al, 2004 and X is the attained age. Since the average entry age was 30 years, we set the expected survival time since the entry into the study for a subject without any radiation exposure to be 40 years. If a subject's survival time was more than his censoring time obtained from the bootstrap sample, then his survival time was censored at his censoring time. Several levels of the random measurement errors described in the previous section were generated and added only to the true positive doses. The MDL levels were then applied to the doses measured with measurement error and doses below the MDL were treated as zeroes. The relative risk was estimated based on the doses with random measurement error and error due to MDL. The simulation was repeated 100 times. Due to computer intensiveness, we set the size

of the simulated data to be 1,320, a third of the size of the ORNL subset. The results indicate that when the magnitude of the random errors is small or medium, the estimate of β is very close to being unbiased. As the magnitude of random errors increases to a high level, the relative bias increases to 44%. There is a large uncertainty association with the bias factors. If the true relative risk is 1.018, then 95% of the time we will get a relative risk estimate in (1.013, 1.040) when the random error is small; (1.011, 1.039) with a moderate random error and (1.004, 1.027) with a large random error. Nevertheless, the confidence interval associated with the estimated log of the relative risk has 100% chance to exclude 0, indicating that the study without correcting for exposure measurement errors has 100% power to detect a significant β. This is an important finding. Even though measurement error in general reduces power, but with the size of the study and the magnitude of the effect, the levels of random error and error due to BMDL we considered here do not affect the power. To see if the result differs when the true β is smaller, we set $\beta = 0.009$ as well as $\beta = 0.003$, a half and a sixth of the size of the previous β, respectively, and repeated the simulations. The conclusions generally remain the same therefore the results are not presented here.

We also used the simulation study to evaluate the type I error of the null test: $\beta = 0$ by setting $\beta = 0$ and re-running the simulation. Consistent with the findings in the previous section, 95% of the estimated β are close to 0 and the estimated 95% confidence interval is close to its nominal value. This suggests that without correcting for the two measurement errors, the study maintains its type I error.

In summary, the hypothesis testing is minimally affected by the levels of errors considered in the study. If the true β is of a magnitude of at least 0.003 per 10 mSV (a minimal level of clinical importance), with the same study population and a size of at least 1,320 subjects, without correcting for measurement errors, our simulation studies show that we would still have almost 100% chance to detect a significant β. The fact that the naive estimate of β was not significant reinforces our conclusion that radiation exposure is not significantly associated with total mortality in this study. Since the uncertainty factor in most of the occupational radiation studies in U.S. rarely exceeded the levels we considered in the simulations, it is then concluded that random measurement error is not likely to be a major concern in the ORNL study.

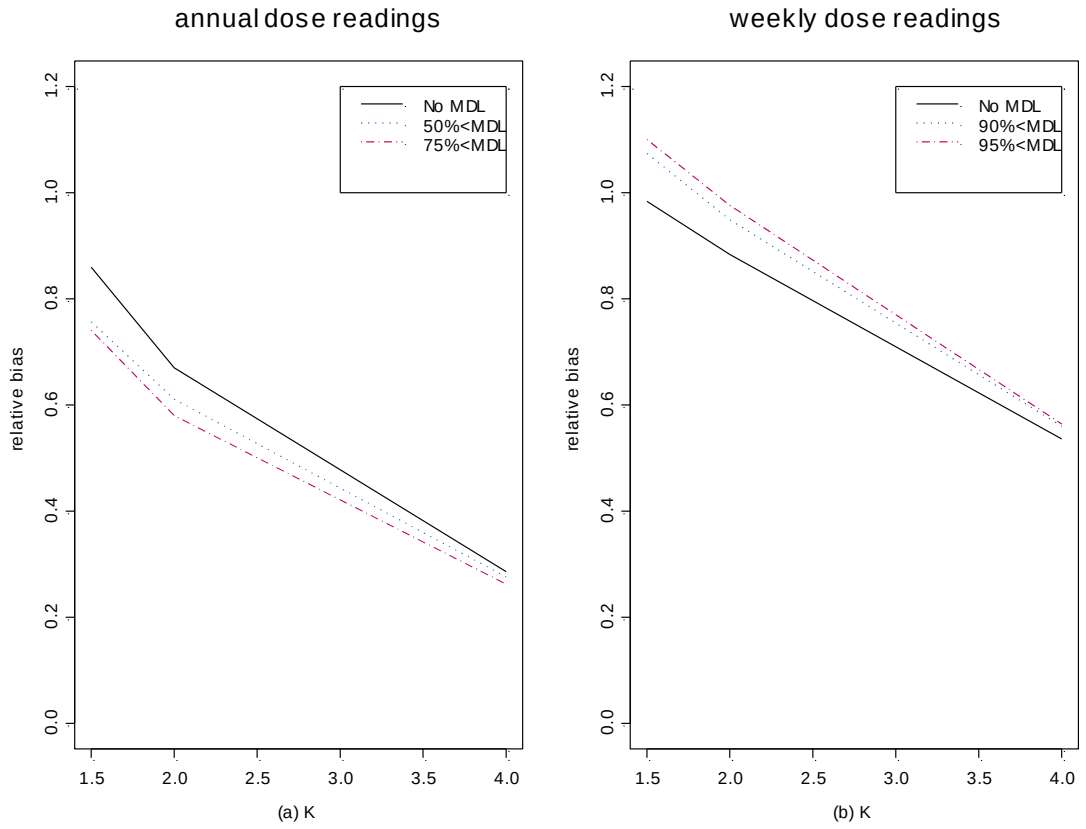

Figure 1. Relative bias in estimates of β when the exposure is subject to measurement error and error due to MDL.

Specific Aim 4. Develop a method to evaluate and control for time-dependent effect modification.

Stratification is often used to control for time-dependent effect modifiers. But for time-dependent strata, it is not always so clear whether time since enrollment or time since the entry of the current time-dependent stratum is a better time scale. We proposed first to use the time from enrollment as the time scale for each stratum and then use a linear splines model to allow the effect of radiation exposure to change within strata. However,

our simulations showed that the spline model we developed does not always perform well. More complicated spline models worked better but are difficult to apply. Therefore, instead of presenting complicated spline models which are not user-friendly consequently with limited applicability, we worked on other topic which is more relevant to occupational health studies.

Ideally when a cohort is available, it is best to use all individuals from the cohort for analysis. However, sometimes when expensive laboratory tests and/or limited samples are involved, this is not feasible and strategic subsets of cases and controls must be chosen for more detailed analysis. A case-cohort study for failure time data consists of a subcohort of randomly selected individuals from the underlying cohort, and either all additional cases not in the subcohort or a subset of the remaining cases. Covariate information is collected on this sample rather than the entire cohort, which greatly reduces the cost and effort in collecting covariates yet with limited loss of efficiency as compared to a large-scale cohort study with a low outcome rate.

However, it is often complicated to design a case cohort study. Therefore, we compared power for case cohort studies and case control studies. Particularly, let P_D represents the proportion of failure in the full cohort and p_j (j=1,2) is the proportion of the population in the jth exposure group and $\tilde{n}$ is the total number of subjects in the subcohort and q is the sampling fraction of the subcohort, the comparison between powers from case cohort studies and case cohort studies with the same number of cases and controls are summarized in the following table:

Table 4.1. Comparison of power computed using method of Cai and Zeng (2004) and case-control approach

Full cohort size n	Failure prop. p_D (%)	p_1	θ	q (%)	$P_{SP}=P_V^*$	P_{CC}^*	$P_{SP} - P_{CC}$
1000	10	0.3	0.5	10	0.507	0.496	0.011
				20	0.615	0.610	0.004
			1.0	10	0.953	0.958	-0.005
				20	0.987	0.989	-0.002
		0.5	0.5	10	0.567	0.522	0.045
				20	0.680	0.630	0.050
			1.0	10	0.976	0.953	0.024
				20	0.995	0.988	0.008
	5	0.3	0.5	10	0.382	0.397	-0.014
				20	0.434	0.457	-0.023
			1.0	10	0.852	0.877	-0.025
				20	0.905	0.927	-0.022
		0.5	0.5	10	0.430	0.404	0.025
				20	0.488	0.458	0.029
			1.0	10	0.902	0.865	0.037
				20	0.943	0.918	0.025
5000	5	0.3	0.5	1	0.437	0.428	0.008
				2	0.620	0.623	-0.003
			1.0	1	0.907	0.929	-0.021
				2	0.988	0.993	-0.005
		0.5	0.5	1	0.490	0.475	0.015
				2	0.686	0.662	0.024
			1.0	1	0.945	0.923	0.021
				2	0.995	0.991	0.005
	1	0.3	0.5	1	0.310	0.322	-0.012
				2	0.375	0.401	-0.025
			1.0	1	0.743	0.784	-0.041
				2	0.844	0.881	-0.038
		0.5	0.5	1	0.348	0.338	0.009
				2	0.422	0.408	0.014
			1.0	1	0.805	0.772	0.033
				2	0.895	0.869	0.026

* P_{sp} = power based on SP_n; $P_{V=}$ power based on V_n in Cai and Zeng (2004); P_{CC} = power based on case-control approach

We demonstrated that for rare events, calculating the power for a case control study with the same number of cases and controls as a case-control study is a good approximation for the power of the case cohort study when event rates are low. This suggests that

existing sample size software for case control studies can be used for the calculation of sample sizes for case cohort studies. This finding will greatly simply the calculations of power and sample size for case cohort studies consequently significantly promote the use of case cohort studies in occupational health studies and in general epidemiological studies.

References:

1. Cai, J. and Zeng, D. (2004) Sample size/power calculation for case-cohort studies. Biometrics 60, 1015-1024.
2. E. Cardis, E. Gilbert, L. Carpenter, et al. Effects of low doses and low dose rates of external ionizing radiation: cancer mortality among nuclear industry workers in three countries", Radiation Research 142, 117-132, 1995.
3. Frome, E.L., Cragle, D.L., Watkins, J.P., Wing, S., Shy, C.M., Tankersley, W.G. and West, C.M. 1997. A mortality study of Employees of the Nuclear industry in Oak Ridge, Tennessee. Radiat. Res. 148, 64-80.
4. E.S. Gilbert, ``Issues in analyzing the effects of occupational exposure to low levels of radiation", Statistics in Medicine 8, 173-187, 1989.
5. E.S. Gilbert, E. Omohundro, J. Buchanan and N. Holter. Mortality of workers at the Hanford site: 1945-1986. Health Physics 64,577-590, 1993a.
6. Gilbert E. S., Cragle D. Wiggs L. 1993b. Updated analyses of combined mortality data for workers at the Hanford site, Oak Ridge National Laboratory, and Rocky Flats Weapons Plant. Radiat Res 136, 408-421.

7. Kerr GD. Missing dose from mortality studies of radiation effects among workers at Oak Ridge National Laboratory. Health Phys 66:206-208; 1994.
8. Liang, KY and Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* 1986; 73: 13 - 22.
9. Mitchell TJ, Ostrouchov G, Frome, EL, Kerr GD. A method for estimating occupational radiation dose to individuals, using weekly dosimetry data. Radiat Res 147:195-207; 1997.
10. Xue X, Shore RE. Methods for estimating occupational radiation doses subject to minimum detection levels. Health Phys 84:61-71; 2003.
11. X. Xue, R.E. Shore, X. Ye and M.Y. Kim. Estimating the dose response relationship for occupational radiation exposure measured with minimum detection level. Health Physics 87,397-404, 2004.
12. Watkins JP, Cragle DL, Frome EL, West CM, Crawford-Brown DJ, Tankersley WG. Collection, validation, and treatment of data for a mortality study of nuclear industry workers. Appl Occup Environ Hyg 12:195-205; 1997.
13. Wing, S., Shy, C.M., Wood, J.L., Wolf, S., Cragle, D.L. and Frome, E.L. 1991. Mortality among workers at Oak Ridge National Laboratory, Evidence of radiation effects in follow-up through 1984. J. Am. Med. Assoc. 265, 1397-1402
14. Wing, S., Shy, C.M., Wood, J.L., Wolf, S., Cragle, D.L., Tankersley, W. and Frome, E.L. 1993. Job factors, radiation and cancer mortality at Oak Ridge National Laboratory: follow-up through 1984. Am. J. of Ind. Med. 23, 265-279.

Publications:

Journal Articles:

1. Xue, X., Kim, M.Y. and Shore, R.E. Estimation of health risks associated with occupational radiation exposure: addressing measurement error and minimum exposure level detectable. *Health Physics*. 2006. 91(6): 582-591.
2. Kim, M.Y., Xue, X. and Du, Y. Approaches for calculating power for case cohort studies. *Biometrics.* 2006; 62, 929-922.
3. Xue, X., Kim, M.Y. and Shore, R.E., Cox Regression Analysis in Presence of Collinearity: An application to Assessment of Health Risks Associated with Occupational Radiation Exposure. *Lifetime Data Analysis* 2007 13:133-350.
4. Xue, X., Kim, M.Y. and Strickler, H. Assessing effect modification in longitudinal studies with rare outcome. *Submitted for review.*

Book Chapters:

1. **Xue, X.** and Hoover D.R. Statistical Methods in Cancer Epidemiological Studies. *Cancer Epidemiology Methods Mol Biol.* 2009;471:239-72

Inclusion of gender and minority study subjects: Since over 90% of the total recorded external dose was received by white males, the study is restricted to white males. The other race and gender groups are not included because they contain too few subjects to provide useful estimates of low-level radiation effects and would unduly complicate the dose-response analysis.

Inclusion of children: Since this is a study of the risk from occupational exposure, no children were involved in the study.

Materials available for other investigators: As we indicated in all of our publications, all the statistical programs used to implement our proposed statistical methods are available upon request.