PNL-10909
UC-605

Accounting for Uncertainty in Systematic Bias in Exposure Estimates Used in Relative Risk Regression

E. S. Gilbert

December 1995

The study was funded by Grant R01 OH03270 from the National Institute for Occupational Safety and Health of the Centers for Disease Control and Prevention

Pacific Northwest Laboratory
Operated for the U.S. Department of Energy
by Battelle Memorial Institute

PNL-10909

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC06-76RLO 1830

Printed in the United States of America

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161

PNL-10909
UC-605

Accounting for Uncertainty in Systematic Bias in Exposure Estimates Used in Relative Risk Regression

E. S. Gilbert

December 1995

The study was funded by Grant
R01 OH03270 from the
National Institute for Occupational
Safety and Health of the Centers
for Disease Control and Prevention

Pacific Northwest Laboratory
Richla 1, Washington 99352

SUMMARY

In many epidemiologic studies addressing exposure-response relationships, sources of error that lead to systematic bias in exposure measurements are known to be present, but there is uncertainty in the magnitude and nature of the bias. Two approaches that allow this uncertainty to be reflected in confidence limits and other statistical inferences were developed, and are applicable to both cohort and case-control studies. The first approach is based on a numerical approximation to the likelihood ratio statistic, and the second uses computer simulations based on the score statistic. These approaches were applied to data from a cohort study of workers at the Hanford site (1944-86) exposed occupationally to external radiation; to combined data on workers exposed at Hanford, Oak Ridge National Laboratory, and Rocky Flats Weapons plant; and to artificial data sets created to examine the effects of varying sample size and the magnitude of the risk estimate. For the worker data, sampling uncertainty dominated and accounting for uncertainty in systematic bias did not greatly modify confidence limits. However, with increased sample size, accounting for these uncertainties became more important, and is recommended when there is interest in comparing or combining results from different studies.

CONTENTS

TABLES

1.0 INTRODUCTION

It is well known that results of exposure-response analyses can be distorted in various ways if exposures are measured with error and these errors are not taken into account. Recent research on this problem has emphasized adjusting analyses for random measurement errors with little specific attention given to the generally simpler problem of adjustment for systematic errors (Armstrong 1990, Clayton 1992, Pierce *et al.* 1989, Richardson and Gilks 1993, Thomas *et al.* 1993). This paper addresses the situation in which potential for systematic bias exists, but the exact nature of the bias is not known with certainty.

The studies that motivated this paper were those of nuclear workers exposed occupationally to external radiation (Gilbert *et al.* 1993a, Gilbert *et al.* 1993b, Carpenter *et al.* 1994, Gribbin *et al.* 1993, IARC 1994). A major objective of these studies is to provide a direct assessment of the carcinogenic risk of exposure to ionizing radiation at low doses and dose rates. To accomplish this objective, risk estimates (expressed per unit of dose) and confidence limits are compared with risk estimates that serve as the basis for radiation protection standards, and which have been based on extrapolation from data on persons exposed at high doses and dose rates (National Academy of Sciences 1990, UNSCEAR 1988). In making these comparisons, it is important to adjust recorded doses for systematic biases and to reflect the uncertainty in these adjustments in confidence limits and other inferences. Dose estimates for workers were obtained from personal dosimeters, and are subject to several sources of error that lead to systematic bias of an uncertain nature.

2.0 MATERIALS AND METHODS

2.1 APPROACHES FOR ACCOUNTING FOR SYSTEMATIC MEASUREMENT ERROR

Methods developed in this paper are applicable to either cohort or case-control studies, although illustrations are based on cohort studies of nuclear workers. Following Moolgavkar and Venzon (1987), the relative risk associated with exposure variable x is denoted by $R(x,\beta)$; that is, $R(x,\beta)$ is the incidence rate ratio for exposure x relative to absence of exposure (x=0). Let i index cases of the disease of interest (i=1, 2, ... N) with case i occurring at time t_i. With a cohort design, let m_i denote the number of subjects in the risk set for case i, where the risk set consists of subjects at risk of disease at time t_i who are similar to case i with respect to specified stratification variables. With a case-control design, let m_i denote the number of matched controls for case i. For either design, let x_{i0} denote the dose of case i, and x_{ij} denote the dose of the j th person in risk set i (j th control). In the nuclear worker studies, x_i has usually been taken to be the dose accumulated by some specified lag period prior to the time the case occurs, and analyses have been stratified on calendar-year period, sex, and other variables thought to affect the baseline risk. Finally, let $\underline{v}_i$ designate various characteristics of the risk (or case-control) sets. For a cohort study, the partial likelihood is then given by

$$L_x\ (x,\beta) = \Pi_i\ \{R(x_{i0},\beta)/[\mathrm{Sum}_j R(x_{ij},\beta)]\} \qquad (1)$$

For a case-control study, this same expression results, but is referred to as a conditional likelihood.

Commonly used forms for the relative risk are the log-linear relative risk model with $R(x,\beta) = \exp(\beta x)$, and the linear relative risk model $R(x,\beta) = 1 + \beta x$. Analyses of data on populations exposed to radiation, including worker studies, have been based primarily on the linear model (Gilbert 1989,

Gilbert *et al*. 1990). This model is also used for applications in this paper, and β is referred to as the excess relative risk or ERR.

Now let z designate an estimate of the true exposure x, with subscripting for z analogous to that for x. It is assumed that z is subject to systematic error such that $x_{ij} = h(z_{ij}, \underline{v}_i, \underline{\theta})$, and that $\underline{\theta}$, the vector of parameters indicating the magnitude and nature of the systematic error, is not known with certainty, but can be described by a density function $g(\underline{v}, \underline{\theta})$. When no confusion is likely to result, $h(z, \underline{\theta})$ is substituted for $h(z_{ij}, \underline{v}_i, \underline{\theta})$, and $g(\underline{\theta})$ is substituted for $g(\underline{v}, \underline{\theta})$. Most results in this paper are based on a simple model, $h(z, \theta) = \theta z$ with a lognormal density used for $g(\theta)$.

The partial likelihood function for β based on the estimated doses z can be written as

$$L_z(\beta, z) = \int g(\underline{\theta})\, L_z(\beta, z|\underline{\theta})\, d\underline{\theta} \qquad (2)$$

where $L_z(\beta, z|\underline{\theta})$ is the partial likelihood function for β conditional on $\underline{\theta}$. Because knowledge of both z and $\underline{\theta}$ is equivalent to knowledge of the true doses x, $L_z(\beta, Z|\underline{\theta})$ is obtained by substituting $x = h(z, \underline{\theta})$ in the likelihood given in equation 1. For the simple model with $x = \theta z$ and the linear relative risk model used for applications in this paper,

$$L_Z(\beta, z) = \int g(\underline{\theta}) \prod_i \{(1 + \beta\theta z_{i0})/(1 + \beta\theta M_i)\}\, d\underline{\theta} \qquad (3)$$

where M_i is the mean of the z_{ij}, $j=0, 1, \ldots, m_i$. Although closed form expressions for the likelihood in equation 2 are rarely available, inferences based on the likelihood ratio statistic can be obtained by approximating the integral in equation 2 by a sum of the form

$$L_z(\beta, z) = \sum_k g_k(\underline{\theta}_k)\, L_z(\beta, z|\underline{\theta}_k), \qquad (4)$$

where the parameter space for $\underline{\theta}$ is partitioned into K categories, and $\underline{\theta}_k$ is the value of $\underline{\theta}$ at a central point in category k. Because the distributions g have often been obtained subjectively, and are rarely known precisely, the use of a discrete approximation is unlikely to result in any serious loss of information. In some cases, g may be a discrete function to begin with and no approximation is necessary.

Calculations in this paper were based on the assumption that log θ follows a normal distribution with mean 0 and standard deviation σ, and intervals were chosen to cover the range from exp(-5σ) to exp(+5σ) with log θ_k^U - log θ_k^L = 0.5 σ, where θ_k^U and θ_k^L are the upper and lower boundaries of interval k; θ_k (see equation 4) was taken to be the geometric mean of these boundaries. Intervals with logarithmic length 0.5 σ gave nearly identical results to finer intervals of length 0.1 σ. Multivariate lognormal distributions were used for situations in which $\underline{\theta}$ was multidimensional.

The sum in equation 4 can be evaluated for several candidate values of $\hat{\beta}$, and in this way both the value $\hat{\beta}$ that maximizes $L_z(\beta, z)$, and confidence limits β^- that satisfy

$$2[\log L_z(\hat{\beta}, z) - \log L_z(\beta^-, z)] = \chi^2_\alpha, \qquad (5)$$

can be determined, where χ^2_α is the appropriate percentile of the single degree-of-freedom chi-square distribution. Although this is a trial and error approach, it was not unduly difficult to implement for most applications of interest in the nuclear worker studies. However, when $\underline{\theta}$ is multidimensional, the number of terms in equation 2 may become large, and one may prefer computer simulations in which $\underline{\theta}$ is repeatedly sampled from the density $g(\underline{\theta})$.

The score statistic based on equation 2 is not generally tractable, but one can develop a somewhat ad hoc procedure based on the "usual" score statistic that would be used under the assumption of no systematic bias in dose estimates. With uncertainty in systematic bias, the usual asymptotic properties of this score statistic are modified, but computer simulations can be

conducted to evaluate its distribution under the assumption that these uncertainties follow the density $g(\underline{\theta})$.

Previous analyses of worker data have used computer simulations to address the inadequacy of asymptotic methods (Gilbert *et al.* 1993a, Gilbert *et al.* 1993b, IARC 1994) and this approach has been described in detail by Gilbert (1989). For calculations in this paper, this approach has been extended by randomly selecting a value of $\underline{\theta}$ from $g(\underline{\theta})$ for each sample selected in the simulation process. Grouped exposure data were used for these simulations.

Another simple solution may be appropriate in some situations. With the simple model, $h(Z, \theta) = \theta Z$, it may be reasonable to assume that the distribution of $\hat{\beta}_\theta$, the maximum likelihood estimate of β conditional on θ, is lognormal with the standard deviation of log $\hat{\beta}_\theta$ given by σ_β, and that $g(\theta)$ is also lognormal with the standard deviation of log θ given by σ. In this case, $\hat{\beta}$, the unconditional estimate of β, follows a lognormal distribution such that the variance of log $\hat{\beta}$ is $\sigma_\beta^2 + \sigma^2$. However, for application to the worker data, where confidence limits for β frequently include negative values, the assumption of a lognormal distribution for $\hat{\beta}_\theta$ did not seem appropriate.

2.2 STUDIES OF NUCLEAR WORKERS EXPOSED TO EXTERNAL RADIATION

Data from a study of workers at the Hanford site and, to a lesser extent, data from a pooled analyses of workers at Hanford, Oak Ridge National Laboratory (ORNL), and Rocky Flats are used to illustrate the methods were described above. Except where noted otherwise, the statistical methods in this paper are the same as those in previous analyses, and details (such as the definition and choice of stratification variables) can be found in Gilbert *et al.* (1993a, 1993b). Tables 2.1 and 2.2 show basic characteristics of the Hanford worker data, and of the combined data including workers at Hanford, ORNL, and Rocky Flats. Hanford is by far the largest population both in terms of the number of workers and the total exposure.

TABLE 2.1. Doses of the Hanford Worker Study Population and of the Combined Study Population of Workers at Hanford, ORNL, and Rocky Flats[a]

	Hanford	Combined Hanford, ORNL, and Rocky Flats
Number of workers	31,763	44,086
Total recorded person-Sv	702	1,085
Average recorded dose (mSv)	22	25

(a) Excludes 857 Hanford workers who had confirmed internal depositions of radionuclides or who received a substantial portion of their dose from neutrons or from low-energy photons (≤100 keV) (Gilbert and Fix 1995).

TABLE 2.2. Number of Deaths in Hanford Worker and Combined Study Populations by Recorded Cumulative Dose[a]

Cumulative Dose (mSv)	All Cancer Except Leukemia	Leukemia Excluding CLL	All Cancer Except Leukemia	Leukemia Excluding CLL
0-	940	25	1,177	36
10-	319	14	412	23
50-	47	1	66	1
100-	37	2	56	4
200-	27	0	32	1
400+	4	1	7	1
Total	1,374	43	1,750	66

(a) Excludes 857 Hanford workers who had confirmed internal depositions of radionuclides or who received a substantial portion of their dose from neutrons or from low-energy photons (≤100 keV) (Gilbert and Fix 1995).

Dose estimates for workers at these facilities were obtained from personal dosimeters worn by workers, and are subject to both systematic and random uncertainty. Sources of systematic uncertainty, of concern for this paper, include bias resulting from the fact that dosimeters were limited in their ability to respond accurately to all radiation energies to which workers were exposed or to radiation coming from all directions. They also include bias resulting from the fact that modern dosimetry programs are usually designed to estimate "deep dose" (defined as the energy absorbed at a depth of 1 cm in tissue) rather than the organ doses needed for epidemiologic purposes, particularly for comparing risk estimates with those based on studies of populations exposed at high doses. In general, the magnitude of the bias depends strongly on both the radiation energy and the geometry (direction from which the radiation was received). Because the distributions of energies and geometries in worker exposure environments are not known with certainty, uncertainty in the magnitude of the bias results. Additional uncertainty results because historical documentation, especially for early periods of plant operation, is not always adequate to allow precise evaluation of the magnitude of bias from some sources.

Extensive efforts have been made to document the dosimetry systems used at Hanford (Wilson *et al.* 1989) and, very recently, Gilbert *et al.* (1996) made a specific effort to quantify biases and their associated uncertainties. For estimating lung dose, where bias would be expected to be similar to that in estimating doses to many other internal organs, overall bias factors, defined as the ratios of the recorded dose to the lung dose, were estimated to be 1.3, 1.1, 1.2, and 1.1 for the respective time periods 1944-56, 1957-71, 1972-83, and 1984-89 when different dosimetry practices were in use. For estimating bone marrow dose, relevant for analyses of leukemia mortality, the respective factors were estimated to be 1.8, 1.4, 1.6, and 1.4. Gilbert *et al.* also evaluated random errors, but these are not considered in the present paper.

Gilbert *et al.* (1994) use lognormal distributions to quantify uncertainties, which are expressed as 95% uncertainty factors K determined so that the

intervals obtained by respectively dividing and multiplying by them include 95% of all observations. Estimates of these uncertainty factors were based on judgments of the authors regarding how well bias factors from various sources were known. For both lung dose and bone marrow dose, the estimated factors reflecting uncertainty in the systematic bias factors were estimated to be about 1.5 for doses received 1957 and later, and about 1.7 for earlier doses. Although these uncertainty factors give a sense of how well the bias factors are known, they cannot be considered to provide a fully rigorous characterization of uncertainties. For this reason, in the adjustments made below, several uncertainty values are evaluated.

The evaluation of bias and uncertainty summarized above was intended to apply only to dose received from external exposure to high energy photons. For this reason, analyses in this paper have excluded 857 Hanford workers who had confirmed internal depositions of radionuclides or who received a substantial portion of their dose from neutrons or from low-energy photons ($\leq$100 keV). These workers were identified as described by Gilbert and Fix (1995).

An evaluation of bias and uncertainty in dose estimates of the type conducted for Hanford (Gilbert *et al.* 1996) has not been conducted for ORNL and Rocky Flats workers. For this reason, some analyses have allowed the uncertainty in systematic bias to be greater for ORNL and Rocky Flats workers than for Hanford workers. In addition, because many Rocky Flats workers were exposed to neutrons and to internal sources of radiation, some analyses have allowed for greater uncertainty for Rocky Flats than for ORNL. In analyses including these workers, correction factors similar to those for Hanford were used; specifically, the recorded doses of ORNL and Rocky Flats were divided by 1.1 and 1.5 to obtain respective estimates of lung and bone marrow doses.

3.0 RESULTS

Table 3.1 shows estimates of the ERR with both 90% and 95% confidence limits based on data on Hanford workers, and on the combined data. Results are shown for all cancer except leukemia, and for leukemia excluding chronic lymphatic leukemia (CLL), the two disease categories that have been of major interest in radiation risk assessment. For comparison, the Hanford estimates based on recorded doses without adjustment were -0.01 per Sv (90% CI = <0-1.0) for all cancer except leukemia, and -1.1 per Sv (90% CI = <0-2.1) for leukemia excluding CLL (Gilbert and Fix 1995). For the combined data, the comparable estimates were 0.0 per Sv (90% CI = <0-0.8) and -1.0 per Sv (90%CI = <0-2.2), respectively (Gilbert *et al.* 1993b).

Analyses of the combined data were based on the assumption that uncertainty was perfectly correlated for the three studies. Although this assumption is not likely to be realistic, it can be considered a "worst case" situation in that with less than perfect correlation, some "cancelling" out of bias may occur. Repeats of the leukemia simulations based on the assumption that uncertainties were independent in the three populations led to the same or slightly smaller upper confidence limits.

Results in Table 3.1 indicate that the absolute values of the maximum likelihood estimates decreased slightly as the uncertainty factor increased, but that the upper confidence limits did not change greatly with allowance for modest uncertainty (95% uncertainty factor of 2 or less). The effect of allowing for uncertainty was greater for leukemia excluding CLL than for all cancer except leukemia, and slightly greater for the combined analyses than for analyses based only on Hanford. Allowing greater uncertainty for ORNL and Rocky Flats than for Hanford did not greatly modify results over those obtained with a 1.5 uncertainty factor for all three studies; probably because the sample sizes for ORNL and Rocky Flats were much smaller than that for Hanford.

TABLE 3.1. Excess Relative Risk (ERR) Estimates (per Sv) with 90% and 95% Confidence Intervals (CI) for All Cancer Except Leukemia and for Leukemia Excluding CLL[a]: Hanford Worker Study and Combined Data on Workers at Hanford, ORNL, and Rocky Flats

95% Uncertainty Factor(s)	All Cancer Except Leukemia[b]			Leukemia Excluding CLL[b]			Leukemia Excluding CLL[c]	
	ERR	90% CI	95% CI	ERR	90% CI	95% CI	90% CI	95% CI
A. Hanford Workers								
1.0	0.20	<0-1.5	<0-1.8	-1.3	<0-4.3	<0-6.1	<0-3.6	<0-5.0
1.5	0.19	<0-1.5	<0-1.9	-1.2	<0-4.4	<0-6.3	<0-3.6	<0-5.0
2.0	0.17	<0-1.6	<0-2.0	-1.0	<0-4.5	<0-6.6	<0-3.7	<0-5.1
3.0	0.13	<0-1.8	<0-2.4	-0.7	<0-4.8	<0-7.3	<0-3.8	<0-5.6
B. Combined Data on Hanford, ORNL, and Rocky Flats Workers								
1.0	0.20	<0-1.1	<0-1.3	-1.2	<0-3.7	<0-5.1	<0-4.1	<0-5.5
1.5	0.19	<0-1.2	<0-1.4	-1.1	<0-3.7	<0-5.3	<0-4.2	<0-5.5
2.0	0.17	<0-1.2	<0-1.5	-0.9	<0-3.8	<0-5.5	<0-4.3	<0-5.9
3.0	0.13	<0-1.3	<0-1.6	-0.6	<0-4.1	<0-6.1	<0-4.6	<0-6.4
1.5, 2.0, 3.0[d]	0.16	<0-1.2	<0-1.5	-0.9	<0-3.7	<0-5.3	<0-4.2	<0-5.6
1.5, 3.0[e]	0.16	<0-1.2	<0-1.5	-0.9	<0-3.7	<0-5.3	<0-4.3	<0-5.7

(a) Based on estimated lung dose for all cancers except leukemia and on estimated bone marrow dose for leukemia excluding CLL.
(b) Confidence intervals based on approximation to the likelihood ratio statistic using ungrouped doses.
(c) Confidence intervals based on computer simulations using the score statistic and grouped doses.
(d) The 95% uncertainty factor was 1.5 for Hanford, 2.0 for the ORNL, and 3.0 for Rocky Flats with uncertainties assumed to be perfectly correlated for the three studies.
(e) The 95% uncertainty factor was 1.5 for Hanford, 3.0 for ORNL and Rocky Flats with uncertainties assumed to be perfectly correlated for the three studies.

Analyses based on the simulated score statistic yielded different upper limits than the likelihood ratio procedure, and this may be partially due to inadequacies with the asymptotic approximation since the number of leukemia deaths was small and the dose distributions highly skewed. Also, the simulated results were based on grouped exposure data while the results using the likelihood ratio approach were based on ungrouped exposure data.

TABLE 3.2. Excess Relative Risk (ERR) Estimates (per Sv) with 90% and 95% Confidence Intervals (CI) Based on Several Modifications to the Hanford Worker Study Data[a]

95% Uncertainty Factor	All Cancer Except Leukemia			Leukemia Excluding CLL		
	ERR	90% CI	95% CI	ERR	90% CI	95% CI
	A. Estimated ERR Equal to Linear Estimates From High Dose Data[b]					
1.0	0.24	<0-1.6	<0-1.9	3.7	<0-18	<0-24
1.5	0.23	<0-1.6	<0-1.9	3.6	<0-19	<0-25
2.0	0.20	<0-1.7	<0-2.1	3.4	<0-20	<0-27
3.0	0.16	<0-1.8	<0-2.4	3.0	<0-23	<0-32
	B. Estimated ERR Equal to Ten Times Linear Estimate From High Dose Data.					
1.0	2.4	0.9-4.3	0.7-4.7	37	6.6-170	4.3-224
1.5	2.3	0.8-4.7	0.6-5.2	37	6.3-176	4.1-234
2.0	2.2	0.7-5.3	0.5-6.2	37	5.8-188	3.8-254
3.0	2.1	0.5-6.8	0.4-8.4	36	4.0-257	2.6-349
	C. As in Table 3.1A With Number of Deaths Increased by Factor of Ten.					
1.0	0.20	<0-0.58	<0-0.66	-1.3	<0,-0.09	<0-0.19
1.5	0.19	<0-0.60	<0-0.70	-1.2	<0,-0.04	<0-0.23
2.0	0.17	<0-0.65	<0-0.78	-1.1	<0-0.02	<0-0.28
3.0	0.16	<0-0.74	<0-0.96	-1.0	<0-0.09	<0-0.35
	D. As in A (Above) With Number of Deaths Increased by Factor of Ten.					
1.0	0.24	<0-0.62	<0-0.70	3.7	1.6-6.7	1.3-7.4
1.5	0.23	<0-0.65	<0-0.75	3.6	1.4-7.3	1.1-8.3
2.0	0.21	<0-0.71	<0-0.85	3.5	1.2-8.4	1.0-9.8
3.0	0.17	<0-0.83	<0-1.05	3.4	0.9-10.7	0.7-13.3
	E. As in B (Above) With Number of Deaths Increased by Factor of Ten.					
1.0	2.4	1.9-3.0	1.8-3.1	37	22-61	20-67
1.5	2.4	1.6-3.6	1.5-3.9	37	20-68	18-76
2.0	2.4	1.3-4.4	1.1-5.0	37	17-80	15-92
3.0	2.2	0.9-6.0	0.8-7.4	37	13-106	10-130

(a) Confidence intervals based on approximation to the likelihood ratio statistic using ungrouped doses.

(b) These estimates were 0.24 per Sv for all cancers except leukemia, and 3.7 per Sv for leukemia excluding CLL.

To investigate the effect of uncertainty in other situations that might occur, the Hanford data were modified in various ways with results shown in Table 3.2. For 3.2A, the Hanford data were adjusted to achieve an ERR estimate equal to the linear estimate from analyses of data on male atomic bomb survivors exposed as adults (UNSCEAR 1988); these estimates were 0.24 per Sv for all cancers except leukemia, and 3.7 per Sv for leukemia excluding CLL. For 3.2B, the data were adjusted to achieve ERR estimates that were ten times the atomic bomb survivor linear estimates. These adjustments were made by multiplying the doses of workers dying of the cause of interest by an appropriate factor. In general, the effect of accounting for systematic uncertainty became greater as the magnitude of the estimate increased.

One of the reasons that allowing for systematic uncertainty in dose estimates did not greatly affect results discussed thus far was that statistical uncertainty resulting from sampling error was very large and dominated results. To investigate how results might be modified if sampling uncertainty were smaller, analyses shown in Table 3.1A, 3.2A, and 3.2B were repeated with the numbers of deaths increased by a factor of ten; results are shown in Table 3.2C, 3.2D, and 3.2E. As expected, the effect of accounting for uncertainty in systematic bias became much more important with the increased sample size. Also, in one case (leukemia analyses 3.2C), the assumption of uncertainty factors of 2 or 3 yielded confidence limits that included zero, whereas the analyses not accounting for uncertainty had excluded zero.

With the Hanford data, Gilbert *et al.* (1996) judged that uncertainty in doses received before 1957 was larger than that for doses received later. For this reason, the analyses shown in Tables 3.1A and 3.2C were repeated using larger uncertainty factors for the earlier period. Specifically, analyses were conducted analyses with a 95% uncertainty factor of 1.5 for the period 1957 and later, and a 95% uncertainty factor of 3 for the period 1944-56. Uncertainties for the two time periods were assumed to be independent or, alternatively, to be perfectly correlated. Because most of the dose was

received in later years, this approach modified results very little over those in which the 95% uncertainty factor was assumed to be 1.5 for all the data.

To further investigate the use of different uncertainties for different parts of the data, analyses of a combined leukemia data set comprised of the actual Hanford data and of Hanford data adjusted to yield an ERR estimate that was twice the linear estimate obtained from male A-bomb survivors exposed in adulthood (7.4 per Sv) were conducted. Results of these analyses are shown in Table 3.3. With the actual sample size (3.3A), the assumption of different uncertainty factors did not greatly modify results over those obtained with the assumption of the same uncertainty factor. However, when the sample size was increased by a factor of ten, both the ERR estimates and the confidence limits depended on which part of the data was assumed to be more uncertain (last two lines of Table 3.3B). Of course, in a real situation, the appropriateness of combining disparate estimates from different segments of the data would need to be considered.

TABLE 3.3. Excess Relative Risk (ERR) Estimates (per SV) with 90% and 95% Confidence Intervals[a] (CI) for Leukemia Excluding CLL-Based on Combined Data Set Consisting of I) the Actual Hanford Worker Data the Hanford Data Modified so that the Estimated ERR was Equal to Twice the Linear Estimate From High Dose Data (7.4 per Sv)

90% Uncertainty Factors[b]		Uncertainty From Two Parts of Data, Independent			Uncertainty From Two Parts of Data, Perfectly Correlated		
I	II	ERR	90% CI	95% CI	ERR	90% CI	95% CI
		A. Actual Sample Size (86 Deaths)					
1.0,	1.0	1.9	<0-8.0	<0-9.8	1.9	<0-8.0	<0-9.8
1.5,	1.5	1.9	<0-8.2	<0-10	1.8	<0-8.3	<0-10
3.0,	3.0	1.7	<0-9.2	<0-12	1.4	<0-9.9	<0-13
1.5,	3.0	1.7	<0-7.7	<0-9.6	1.6	<0-7.9	<0-9.9
3.0,	1.5	1.9	<0-9.9	<0-13	1.8	<0-10	<0-13
		B. Number of Deaths Increased by a Factor of Ten					
1.0,	1.0	1.9	0.7-3.4	0.5-3.7	1.9	0.7-3.4	0.5-3.7
1.5,	1.5	2.2	0.9-4.0	0.8-4.4	1.8	0.7-3.6	0.5-4.1
3.0,	3.0	2.1	0.8-5.0	0.7-5.9	1.7	0.4-5.3	0.3-6.5
1.5,	3.0	1.4	0.6-2.8	0.5-3.2	1.1	0.5-2.4	0.4-2.7
3.0,	1.5	3.8	1.7-7.1	1.4-7.7	4.4	1.5-10	1.1-12

(a) Confidence intervals based on approximation to the likelihood ratio statistic using ungrouped doses.

(b) The first factor was applied to the component consisting of the actual Hanford data (I); the second factor was applied to the component consisting of the modified Hanford data (II).

4.0 DISCUSSION

An approach for accounting for uncertainty in systematic bias has been described and applied to data on workers exposed occupationally to external radiation. For Hanford workers, an extensive evaluation of bias and uncertainties in dose estimates had recently been completed, and the analyses presented in this paper were the first to apply adjustment factors to recorded doses (to obtain estimates of lung and bone marrow dose), and to allow for uncertainty in the magnitude of the adjustment factors. For all cancer except leukemia, where the estimated adjustment factors were only slightly greater than one, these modifications did not alter results greatly. For leukemia excluding CLL, the application of the adjustment factors increased the absolute values of estimates and confidence limits by a factor of about 1.5, a difference that is important in comparing worker-based estimates and confidence limits with those obtained from data on populations exposed at high doses, which serve as the basis for current radiation protection standards. For both all cancer except leukemia and leukemia excluding CLL, sampling uncertainty was large, and, thus, allowing for modest uncertainty in the adjustment factor did not greatly modify results over those with no such allowance.

Results of analyses of combined data from Hanford, ORNL, and Rocky Flats were also presented, but must be regarded as preliminary since a detailed evaluation of bias in dose estimates used in these studies has not been conducted. However, even with a 95% uncertainty factor of 3.0 for these studies (but smaller uncertainty for Hanford), results were not greatly different from those with no allowance for uncertainty.

In addition to the U.S. nuclear worker studies evaluated in this paper, studies have also been carried out in the United Kingdom (Carpenter *et al.* 1994) and Canada (Gribbin *et al.* 1993) and combined analyses of data from the three countries conducted (IARC 1994). Also, a collaborative study including workers in several additional countries is currently underway. As indicated

in some of the hypothetical results shown in Tables 3.2 and 3.3, uncertainty in systematic bias becomes more important as sample size increases. For this reason, it is important to allow for this uncertainty in future international combined analyses. It is hoped that further work to evaluate bias and uncertainty in dose estimates used in nuclear worker studies throughout the world will be undertaken.

Ideally dose-response analyses should take account of all errors in dose estimates including both random and systematic components. Several papers have addressed methods for a more general error structure with particular attention to random errors, but tractable solutions are not always readily available (Armstrong 1990, Clayton 1992, Pierce *et al.* 1989, Richardson and Gilks 1993, Thomas *et al.* 1993). Although it is hoped that methods for adjusting for all types of exposure measurement errors will continue to be explored, the approach described and illustrated in this paper provides a relatively simple procedure for accounting for one potentially important source of uncertainty.

The need to account for uncertainty in systematic bias is not of course limited to studies of nuclear workers. Attention to such bias is likely to be especially important in studies where sampling uncertainty is relatively small, and where some subjects are clearly subject to greater uncertainties than others. For example, in the atomic bomb survivor studies, the sampling uncertainty is much smaller than in the worker studies (National Academy of Sciences 1990, UNSCEAR 1988, Shimizu *et al.* 1990). In addition, the estimated yield of the Hiroshima bomb is estimated to be much less certain than that for the Nagasaki bomb (Kaul 1984), and thus the potential for systematic bias in dose estimates of Hiroshima subjects is much greater than for Nagasaki subjects. Another example is the large number of lung cancer case-control studies of the effects of exposure to residential radon, where sampling uncertainty may become small if analyses of combined data including several thousand cases are eventually conducted (Samet *et al.* 1991). Like the worker studies, risk estimates and confidence limits from the residential studies are

compared with those obtained from studies at higher exposure levels (in this case, from studies of underground miners) (Lubin 1994).

To apply the methods of this paper, systematic bias and its uncertainty must be specified. Although in some cases, relevant data may be available to provide an objective basis for determining the uncertainty distributions $g(\underline{\theta})$, in most cases, subjective judgments of experts are likely to be required. In these cases, it is desirable to perform analyses based on more than one set of assumptions regarding the magnitude of the bias, as has been done with the nuclear workers. With the workers, even allowing fairly large uncertainty did not greatly affect results, and these calculations thus allow one to be more confident that allowance for systematic uncertainty does not greatly modify conclusions. For the worker studies, the assumption that $g(\underline{\theta})$ was lognormal was judged appropriate. However, the general approach could be applied to other distributions in other situations.

The general approach described in this paper could also be applied to uncertainty in systematic bias resulting from factors other than exposure measurement, for example, bias resulting from confounding. To do this, it would be necessary to specify distributions describing the uncertainty, and this would also involve subjective judgments.

5.0 REFERENCES

Armstrong, B. 1990. "The effects of measurement errors on relative risk regressions." Am. J. Epidemiol. 132:1176-1184.

Carpenter, L., C. Higgins, A. Douglas, P. Fraser, V. Beral, and P. Smith. 1994. "Combined analysis of mortality in three United Kingdom nuclear industry workforces, 1946-1988." Radiat. Res. 138:224-238.

Cardis, E., E. Gilbert, L. Carpenter, G. Howe, I. Kato, B. Armstrong, V. Beral, G. Cowper, A. Douglas, J. Fix, S. Fry, J. Kaldor, C. Lave, L. Salmon, P. Smith, G. Voelz, and L. Wiggs. 1995. "Effects of low doses and low dose rates of external ionizing radiation: cancer mortality among nuclear industry workers in three countries." Radiat. Res. 142:117-132.

Clayton, D. G. 1992. "Models for the analysis of cohort and case-control studies with inaccurately measured exposures." In: Statistical models for longitudinal studies of health, J.H. Dwyer, P. Lippert, M. Feinleib, and H. Hoffmeiser, eds., pp. 301-331. New York: Oxford University Press.

Gilbert, E. S. 1989. "Issues in analyzing the effects of occupational exposure to low levels of radiation." Statist. in Med. 8:173-187.

Gilbert, E. S., S. A. Fry, L. D. Wiggs, G. L. Voelz, D. L. Cragle, and G. R. Petersen. 1990. "Methods for analyzing combined data from studies of workers exposed to low doses of radiation." Am. J. Epidemiol. 131:917-927.

Gilbert, E. S., E. Omohundro, and J. A. Buchanan. 1993a. "Mortality of workers at the Hanford site: 1945-1986." Health Phys. 64:577-590.

Gilbert, E. S., D. L. Cragle, and L. D. Wiggs. 1993b. "Updated Analyses of Combined mortality data on workers at the Hanford site, Oak Ridge National Laboratory, and Rocky Flats Weapons Plant." Radiat. Res. 136:408-421.

Gilbert, E. S. and J. J. Fix. 1995. "Accounting for bias in dose estimates in analyses of data from nuclear worker studies." Health Phys. 68:650-660.

Gilbert, E. S., J. J. Fix, and W. V. Baumgartner. 1996. "An assessment of bias and uncertainty in recorded dose from external sources of radiation for workers at the Hanford site." Health Phys., in press.

Gribbin, M. S., J. L. Weeks, and G. R. Howe. 1993. "Cancer mortality (1956-1985) amongst male employees of Atomic Energy of Canada Limited with respect to occupational exposure to external low linear energy transfer ionizing radiation." Radiat. Res. 133:375-380.

IARC Study Group on cancer risk among nuclear industry workers. 1994. "Direct estimates of cancer mortality due to low doses of ionizing radiation: an international study." Lancet 344:1039-1043.

Kaul, D. C. 1984. "Review of yield estimates for the Hiroshima and Nagasaki bombs." In: Second US-Japan joint workshop for reassessment of atomic bomb radiation dosimetry in Hiroshima and Nagasaki. Hiroshima, Japan: Radiation Effects Research Foundation.

Lubin, J. H. 1994. "Invited commentary: Lung cancer and exposure to residential radon." Am. J. Epidemiol. 4:323-332.

Moolgavkar, S. H. and D. J. Venzon. 1987. "General relative risk regression models for epidemiologic studies." Am. J. Epidemiol. 126:949-961.

National Academy of Sciences. 1990. Health effects of exposure to low levels of ionizing radiation, BEIR V. Report of the Committee on the Biological Effects of Ionizing Radiations, National Research Council. Washington, DC: National Academy of Sciences.

Pierce, D. A., D. O. Stram, and M. Vaeth. 1989. "Allowing for random errors in radiation exposure estimates for the atomic bomb survivor data." Radiat. Res. 123:275-284.

Richardson, S. and W. R. Gilks. 1993. "A Bayesian approach to measurement error problems in epidemiology using conditional independence models." Am. J. Epidemiol. 138:430-442.

Samet, J. M., J. Stolwijk, and S. L. Rose. 1991. "Summary: International workshop on residential radon epidemiology." Health Phys. 60:223-227.

Shimizu, Y., H. Kato, and W. J. Schull. 1990. "Studies of mortality of A-bomb survivors. 9. Mortality, 1950-1985: Part 2. Cancer mortality based on the recently revised doses (DS86)." Radiat. Res. 121:120-141.

Thomas, D. C., D. Stram, and J. Dwyer. 1993. "Exposure measurement error: Influence on exposure-disease relationships and methods of correction." Annu. Rev. Publ. Health 14:69-93.

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). 1988. Sources, Effects, and Risks of Ionizing Radiation, 1988 Report to the General Assembly, with Annexes. New York, NY: United Nations.

Wilson, R. H., J. J. Fix, W. W. Baumgartner, and L. L. Nichols. Description and Evaluation of the Hanford Personnel Dosimeter Program From 1944 Through 1989. PNL-7447. Richland, WA: Pacific Northwest Laboratory.

PNL-10909
UC-605

DISTRIBUTION

No. of Copies

OFFSITE

D. Cragle, Director
Center for Epidemiologic Research
Oak Ridge Institute of Science and Education
P.O. Box 117
Oak Ridge, TN 37831-0117

R. Hornung
National Institute for Occupational Safety and Health
4676 Columbia Parkway, MS R-44
Cincinnati, OH 45226

W. E. Murray
National Institute for Occupational Safety and Health
4676 Columbia Parkway, MS R-44
Cincinnati, OH 45226

FOREIGN

E. Cardis, Head
Programme on Radiation and Cancer
International Agency for Research on Cancer
150, Cour Albert-Thomas
69372 Lyon Cedex 08
FRANCE

No. of Copies

Lucy Carpenter, Ph.D.
Department of Public Health and Primary Care
Gibson Bldg.
Radcliffe Infirmary
Oxford OX2 6HE
UNITED KINGDOM

Colin R. Muirhead, Ph.D.
National Radiological Protection Board
Chilton, Didcot, Oxon
OX11 ORQ
UNITED KINGDOM

L. Salmon
8 Upton Close
Abingdon, Oxon14 2AI
UNITED KINGDOM

G. Cowper
4 Cartier Circle
Deep River, Ontario KOJ1PO
CANADA

ONSITE

(19) Pacific Northwest National Laboratory

J. A. Buchanan	K3-54
J. J. Fix	P7-02
E. S. Gilbert (15)	K3-54
PNNL Information Release (7)	