

STATISTICAL ISSUES IN RADIATION DOSE-RESPONSE ANALYSIS OF EMPLOYEES OF THE NUCLEAR INDUSTRY IN OAK RIDGE, TN

in ASA 1997 Proc Statistics in Epidemiology, pp 58-63

E. L. Frome, Oak Ridge National Laboratory
J. P. Watkins, Oak Ridge Associated Universities

Key Words: Poisson Regression; regression diagnostics; Comprehensive Epidemiologic Data Resource; relative risk; dose-response

Abstract:

Poisson regression methods are used to describe dose-response relations for cancer mortality for a subcohort of 28,347 white male radiation workers. Age specific baseline rates are described using both internal and external (U.S. white male) rates. Regression analyses are based on an analytic data structure (ADS) that consists of a table of observed deaths, "expected" deaths, and person-years at risk for each combination of levels of seven risk factors. The factors are socioeconomic status, length of employment, birth cohort, age at risk, facility, internal exposure, and external exposure. Each observation in the ADS consists of the index value of each of the stratifying factors, the observed deaths, the expected deaths, the person-years, and the ten year lagged average cumulative dose. Regression diagnostics show that a linear exponential relative risk model is not appropriate for these data. Results are presented using a main effects model for factors other than external radiation, and an excess relative risk term for cumulative external radiation dose.

Data sets and computer programs are available via the Internet at <www.epm.ornl.gov/~frome/>—see (Oak Ridge Mortality Study).

1. Introduction

This report considers statistical issues related to combining data from multiple facilities to evaluate the potential adverse health effects of low-level occupational exposure to ionizing radiation. A detailed description of the data collection and validation procedures, as well as tabular and graphical summaries of the resulting data base are presented in companion reports [13, 12, 14]. The analysis files (which are SAS data sets, see Figure 1) are the starting point for the data analysis process. After reviewing the radiation exposure data and the monitoring and recording procedures at each of the Oak Ridge plants, we decided to limit our dose-response analysis to the subcohort of white males who were ever employed

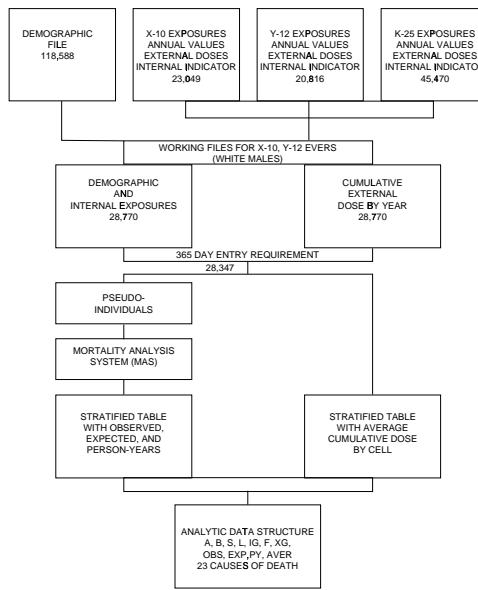


Figure 1: Procedure Used to Generate The Analytical Data Structure

at the X-10 or Y-12 plant. About thirty percent of these workers were employed at more than one Oak Ridge facility. Most of the workers in the subcohort wore personnel monitoring devices that recorded estimates of their external dose over their employment history. For internal radiation exposure the monitoring policies and procedures varied considerably among facilities and over time [13, 14]. In this combined analysis the internal radiation exposure IG was limited to the three categories (see Table 1).

2. Generating The Analytic Data Structure (ADS)

The three exposure analysis files at the top of Figure 1 contain the external dose estimate, and the internal exposure index for each year that a person worked at X-10, Y-12, or K-25. In addition to the usual *time related* variables (i.e. birth cohort, age-at-risk) there were three *time dependent* covariates:

- i) cumulative external dose in year t ;
- ii) sequential internal exposure category in year t ;
- iii) facility in year t (as described in [7])

The entry and exit dates for the time dependent cells also include a lag of two, ten, or twenty years. The person-epoch [3] approach with grouping on the time dependent covariates is used to generate the ADS shown at the bottom of Figure 1. The factors in Table 1 are used to illustrate the procedure that was followed to generate the ADS for our dose-response analysis using all cancer with a ten year lag as an example.

The **first step** in the data reduction process was to create the “working files” (see Figure 1) for the subcohort of X-10/Y-12 white males. These files were needed to combine data for individuals employed at more than one facility. The *Cumulative External Dose* file contains yearly cumulative external dose data (with a ten year lag) from all previous years of employment at all Oak Ridge facilities.

In the **second step** the variables in these two working files were used to divide each individual into *pseudo-persons* according to changes in category membership for the time-dependent covariates. Each record in this file had an entry date and exit date for one of the time-dependent strata and all additional information that was needed by mortality analysis system (MAS) [9] for exact computation of person-years. In this example the 28,347 workers are partitioned into 61,597 pseudo-individuals for processing by MAS.

In the **third step** the MAS program was used to generate a person-years matrix for five year age and calendar year intervals for all possible combinations of factors of interest (see Table 1). Each of these matrices was then multiplied by the appropriate rate table for each cause of death category. This results in a seven dimensional table that contains 4,230 cells with person years greater than zero. Each cell in the ADS contains the index values for each of the factors, the person-years, the person-year weighted average dose, the observed and expected deaths for each cause of death category. The ADS used to obtain the results in Tables VI—VIII of Frome *et al.* [7] is available through CEDR [2]. The all cancer subset used in this report can be obtained at the URL in the Abstract. Table AIV in ORNL-6785 [7] shows the marginal distribution of person-years by facility and dose group for X-10/Y-12 workers (white males) with a ten year lag.

2.1 Fitting The Main Effects Model

The general Poisson regression model is

$$E(y_{jk}) = n_{jk}\lambda_{jk} = n_{jk}\lambda_{jk}^0 R(Z_j, X_k), \quad (1)$$

Table 1
Factors Used to Define ADS for Dose-Response
Analysis for White Males Employed at the X-10 or Y-12 Facility

Factor	Levels	Description
A	15	Attained age: Five-year intervals
B	5	Birth cohorts: Ten-year intervals
S	2	SES - Paycode: Monthly vs Non-Monthly
L	2	Length of employment: less than 1 year vs one year+
IG	3	Internal Exposure: EN - Eligible and Not monitored EM - Eligible and Monitored NE - Not Eligible
F	3	Facility: X-10, Y-12, or Other
X	10	External radiation dose group cut points (mSv): 0, 5, 10, 20, 40, 80, 160, 320, 640+.

where λ_{jk} represents the unknown mortality rate, y_{jk} is the number of deaths, and n_{jk} denotes the person-years at risk in the jk^{th} cell of an ADS. The j subscript indicates the dimensions of the ADS that correspond to factors of secondary interest (e.g., birth-cohort, SES) and the k subscript indicates exposure related covariates that are of primary interest, i.e. facility, internal exposure, and external dose.

Eq.(1) was used to describe the joint effects of each of the explanatory variables of interest on cause-specific mortality. Maximum likelihood estimates of the parameters and likelihood ratio test (LRT) statistics were obtained using Poisson regression [6]. The baseline rate λ_{jk}^0 represents the age-specific death rate for individuals at the reference level of each of the explanatory variables. In previous reports on the X-10 only subcohort [15, 16] a parametric model was used to describe the baseline rates, i.e.

$$\lambda_{jk}^0 = \exp [\alpha_k + \theta \log(A_{jk}/52.5)],$$

where A_{jk} = age at interval mid-point. In Frome *et al.* [7, 8] the external/internal model ([1], Chapter 4)

$$\lambda_{jk}^0 = \lambda_{jk}^* \exp (Z_j \alpha),$$

in which the baseline rates are assumed to be proportional to the known external standard rates λ_{jk}^* (U. S. white male rates) was used.

To contrast and further explain these two models both of them were applied to the data for the seven dimensional table of observed and expected deaths for all cancers for the X-10/Y-12 subcohort.

For the internal analyses the ERR main effects model is expressed as

$$r_{jk} = \exp [AL + B + S + L + IG + F] (1 + D),$$

where $r_{jk} = y_{jk}/n_{jk}$ is the observed cancer death rate per thousand person years at risk in the jk^{th} cell, and $AL = \log(A_{jk}/52.5)$. In this specification all of the terms are factors (one parameter for each level) except for AL and dose (D), and birth cohort is the “referent factor”.

For convenience in describing results the convention—[4] Chapter 22—of dropping Greek letters (that represent the unknown parameters) and listing the explanatory variables that define the relative risk function is used. This corresponds to standard GLIM notation [5] for a log-linear model with factors B, S, L, IG, and F (see Table 2).

The baseline rates are described by the first six estimates in column 2 of Table 2, e.g. for the 1910-19 birth cohort the estimated baseline rates are $\log(\text{rate}) = 0.728 + 5.20 * \log(\text{age}/52.5)$. Consequently, the estimates for the factor B in column two (lines 2 through 6) of Table 2 represent the log of the all cancer mortality rate for each of the five birth cohorts at the reference age 52.5 (i.e. they are the intercept parameters). The estimates for each level of the factors S, L, IG, and F are relative risks in L% units (see Törnqvist *et al.* [11]), with the first level of each factor as the referent category. The next to last value in column two is the estimated coefficient for external radiation dose and represents the change in the ERR per Sv. Estimates of the standard errors of each of the parameter estimates are given in column 3 of Table 2, and likelihood ratio based 95% CI are given for the ERR in the last row. The 95 percent confidence intervals were obtained using the *bounds* command in **epicure** [10] (at the URL in the abstract see af-tabv.txt at “Data and Computer Programs”).

For the “external/internal” analysis

$$r_{jk} = \exp[A + B + S + L + IG + F + D](1 + D),$$

where $r_{jk} = y_{jk}/n_{jk}\lambda_{jk}^*$ is the SMR for the jk^{th} stratum and $A = (\text{age} - 52.5)/100$. The main differences between this and the internal analysis are in birth cohort and age terms. The estimates for referent factor B (birth cohort) in column 4 lines 2 through 6 of Table 2 represent the SMR for each birth cohort in L% units, i.e. $\exp(-10.1/100) = 0.919$ is the estimated SMR for the 1910-19 birth cohort at the reference level of each of the other factors. Interpretation of the levels of the referent factor is straightforward. These estimates (lines 2 through 6 in column 4 of Table 2) show, that the all cancer mortality rates for the referent group (X-10 only nonmonthly long term workers that were eligible but not monitored for internal radiation exposure) are less than the U. S. white male rates, and that the decrease is larger for younger workers, i.e. the more recent birth cohorts. The parameter estimates for the factors S, L, IG, and F (in column 4) have the same interpretation as in the internal analysis, and are almost identical in numerical value to the corresponding estimates from the internal analysis (see column 2). The estimate for the age term in column 4 describes (and adjusts for) any systematic age-related difference in the external rates and the study cohort in percent per year units.

Table 2
Comparison of Parameter Estimates For ERR Main Effects Model All Cancer with Ten Year Lag
(N=28,347) White Males Ever Employed at X-10 or Y-12 Between 1943 and 1984

Age	Term	Internal ^a		External/Internal ^b	
		Estimate	SE	Estimate	SE
B	<1900	0.538	0.132	-9.2	13.4
	1900-09	0.731	0.102	-4.3	10.3
	1910-19	0.728	0.093	-10.1	9.3
	1920-29	0.565	0.103	-23.0	10.3
S ^c	1930+	0.072	0.181	-66.2	17.8
	NonM	0	.	0	.
L	M vs NonM	-41.0	8.3	-41.3	8.3
	Worked 1 Yr +	0	.	0	.
IG	Worked 1 Yr	11.8	9.1	11.3	9.1
	Elig Not M	0	.	0	.
	Elig M	5.8	7.5	4.3	7.5
F	Not Elig	2.6	8.7	2.9	8.7
	X-10	0	.	0	.
	Y-12 vs X-10	15.6	8.6	15.3	8.6
D ^e	Mult vs X-10	6.1	7.8	5.9	7.8
	Ext Dose (Sv)	1.52	0.82	1.45	0.81
95% CI		(0.18, 3.59)		(0.15, 3.48)	

^aFor internal analysis baseline rates are estimated using $\log(\text{age}/52.5)$ for the age term, so that for B=1910-19, $\exp(.743) = 2.11$ cancer deaths per 1000. Note that for U.S. white males born in 1915 the cancer mortality rate at age 52.5 is 2.24/1000.

^bFor External/Internal model baseline rates are estimated using $\log(\lambda_j^*)$ as an “offset” as part of the age term (λ_j^* are known rates from vital statistics for U.S. white males). Consequently, for B=1910-19, $\exp(-8.5/100) = .907$, is the estimated SMR for the 1915 birth cohort at the reference level of the other terms. The age variable, defined as $(\text{age} - 52.5)/100$, was included to reflect any systematic age related departures from the external rates.

^cThe coefficients for the factors S, L, IG, and F are relative risk estimates in L% units, i.e. the cancer relative risk estimate for monthly vs non-monthly workers is $\exp(-41.9/100) = 0.66$.

^dERR per Sv with 95% likelihood based confidence intervals.

Table 3

B	S	L	IG	F
-10.1	+	0	+	0
-10.1	+	0	+	0
-10.1	+	-41.3	+	0
-10.1	+	-41.3	+	0

	log (1 + ERR) L%	log (SMR) L%	SMR
+	0	=	9.5
+	1.5	=	11.0
+	0	=	-31.8
+	1.5	=	-30.3

The relative risk estimates for the exposure variables IG, F, and D are adjusted for effects of the confounding variables age, B, S, and L. Their values are readily combined to obtain an estimate of the SMR for any combination of factor levels and dose. Consider four individuals in the 1910-19 birth cohort, who were long-term Y-12 only workers and were monitored for internal radiation exposure. Further, suppose that the first two were non-monthly, and that the second two were monthly, and that two of them received cumulative external radiation doses of 0.01 Sv. Using the parameter estimates from column 4 of Table 2 we obtain the results in Table 3.

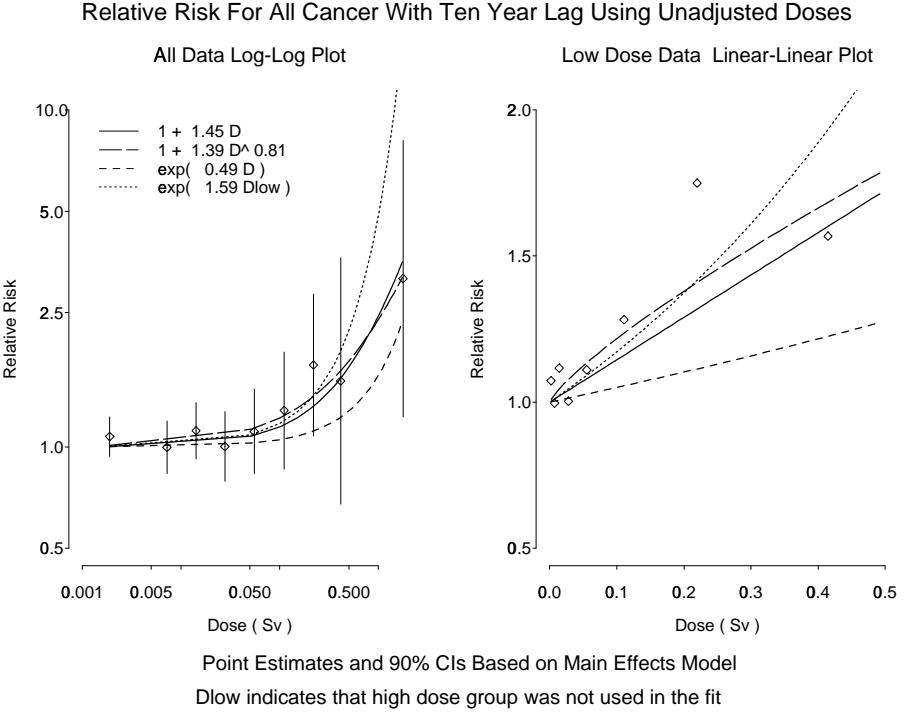


Figure 2: Fitted dose-response functions.

2.2 Regression Diagnostics For The Exponential Relative Risk Function For All Cancer

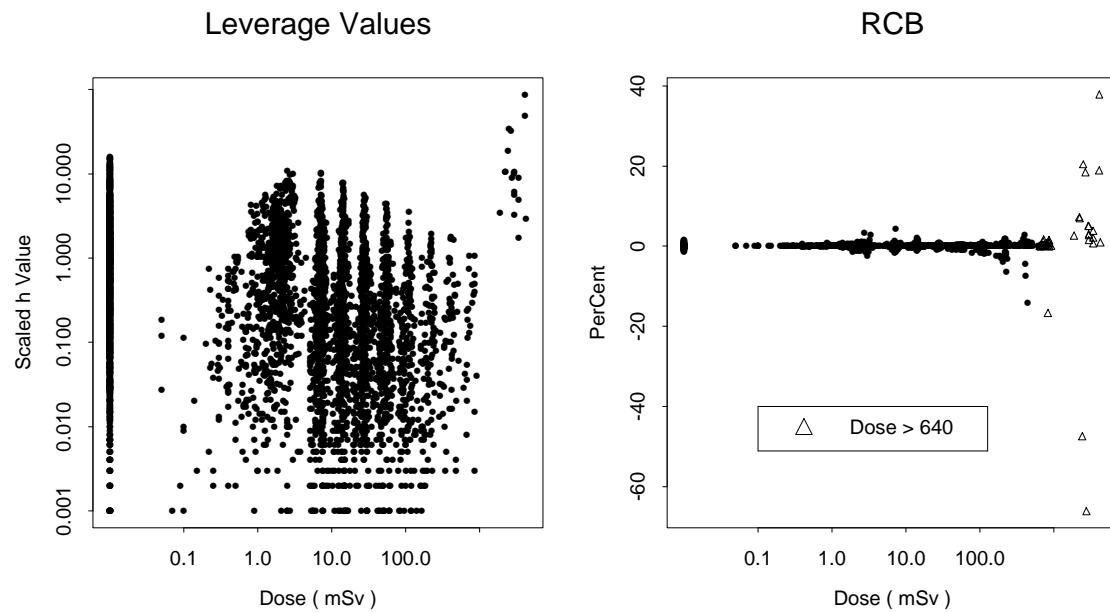
In previous reports [15, 16], the exponential relative risk function, $\exp(\beta D)$, was used to describe the effect of cumulative radiation dose on cancer risk. One way to check the validity of the exponential relative risk function is to plot estimates of the relative risk versus the average dose in each of the dose categories (see Figure 2). Another approach is to use regression diagnostic techniques that describe the influence of individual data points on parameter estimates [1, 6]. The diagonal term from the hat matrix h_j provides a measure of the influence of each cell in the ADS. The left hand panel of Figure 3 shows the h_j values—scaled so that average (h) = 2—plotted against the dose value for each of 4160 cells that were used in the example. This plot indicates that there are a small number of cells with high relative influence in the highest dose group. To directly evaluate the influence of each cell in the ADS on the dose-response coefficient β , the main effects model with exponential relative risk was fit to each table obtained by deleting the j^{th} cell to obtain $\hat{\beta}_{(j)}$. The relative percent change in $\hat{\beta}$ (RCB) due to the

j^{th} cell is then

$$RCB = 100 \left(\frac{\hat{\beta}_{(j)} - \hat{\beta}}{SE\hat{\beta}} \right)$$

The right hand panel of Figure 3 shows the RCB values plotted versus dose for this example. This plot shows that there are a small number of cells with high leverage values for the linear exponential dose-response coefficient. This suggests that the exponential relative risk dose-response relation may not be appropriate for this data. Table XI in [7] provides a more detailed evaluation of alternative dose-response functions, and Figure 2 further demonstrate this point.

Diagnostic Plots For Dose-Response Coefficient
 Exponential Relative Risk For All Cancer- Ten Year Lag Using Unadjusted Doses
 The Seven Dimensional ADS Contains 4160 Cells Zero Dose Points Plotted at 0.001



$RCB = 100 * [B(j) - 0.489] / 0.205$
 Where $B(j)$ Denotes the ML Estimate with the j th Cell Omitted From The Fit
 ORNL-6785 ORMS: X-10/Y-12 White Male Subcohort

Figure 3: Regression Diagnostics

3. Acknowledgments.

Funding was provided by the U.S. Department of Energy (DOE) with management and scientific oversight from the National Institute for Occupational Safety and Health (NIOSH). We thank the staff of ORAU's Center for Epidemiologic Research and Lockheed-Martin Energy Systems for their help in data collection, and Stelmo Poteet for assistance in manuscript preparation. The authors acknowledge the vital statistics offices of the individual states as the source of death record data and appreciate the offices' technical support of this research. The authors are solely responsible for the data analysis and the interpretation of the results.

References

- [1] N. E. Breslow and N. E. Day. *Statistical Methods in Cancer Research, Volume II: The Design and Analysis of Cohort Studies*. Number Scientific Publication 82. International Agency for Research on Cancer, Lyon, 1987.
- [2] CEDR. *Comprehensive Epidemiologic Data Resource*. U. S. Department of Energy, Office of Environment Safety and Health, Washington, D. C., May 1995. URL <http://cedr.lbl.gov/CEDRhomepage.html>.
- [3] D. Clayton. The analysis of event history data: a review of progress and outstanding problems. *Stat. in Med.*, 7:819–841, 1988.
- [4] D. Clayton and M. Hills. *Statistical Models in Epidemiology*. Oxford University Press, 1993.
- [5] B. Francis, M. Green, and C. Payne. Glim 4 the statistical system for generalized linear interactive modelling. Release 4, Royal Statistical Society, Oxford University Press, 1993.
- [6] E. L. Frome. The analysis of rates using Poisson regression models. *Biometrics*, (39):665–674, 1983.
- [7] E. L. Frome, D. L. Cragle, J. P. Watkins, S. Wing, C. Shy, W. M. Tankersley, and C. M. West. A mortality study of employees of the nuclear industry in Oak Ridge, Tennessee. Technical Report 6785, ORNL, 1997.
- [8] E. L. Frome, D. L. Cragle, J. P. Watkins, S. Wing, C. Shy, W. M. Tankersley, and C. M. West. A mortality study of employees of the nuclear industry in Oak Ridge, Tennessee. *Radiat. Res.*, (148):64–80, 1997.
- [9] R. W. McLain and E. L. Frome. A system for the analysis of cohort mortality data. In *SAS Users Group International Conference Proceedings*, SUGI II, pages 874–876, Cary, NC, 1986. SAS Institute.
- [10] D. L. Preston, J. H. Lubin, D. A. Pierce, and M.E. McConney. *epicure* Users's Guide. Technical report, Hirosoft International Corporation, Seattle, WA, 1993.
- [11] L. Törnqvist, P. Vartia, and Y. O. Vartia. How should relative changes be measured? *Am. Statist.*, (39):43–46, 1985.
- [12] J. P. Watkins, D.L. Cragle, E.L. Frome, C.M. West, D.J. Crawford-Brown, and W.G. Tankersley. Adjusting external doses from the ORNL and Y-12 facilities for the Oak Ridge nuclear facilities mortality study. Technical Report G-34, ORISE, 1994.
- [13] J. P. Watkins, J. L. Reagan, D.L. Cragle, E.L. Frome, C.M. West, D. Crawford-Brown, and W.G. Tankersley. Collection, validation, and description of data for the Oak Ridge nuclear facilities mortality study. Technical Report J42, ORISE, 1993.
- [14] J. P. Watkins, J. L. Reagan, D.L. Cragle, E.L. Frome, C.M. West, D. Crawford-Brown, and W.G. Tankersley. Collection, validation, and treatment of data for a mortality study of nuclear industry workers. *Appl. Occup. Environ. Hyg.*, (12):195–205, 1997.
- [15] S. Wing, C. M. Shy, J. L. Wood, S. Wolf, D. L. Cragle, and E. L. Frome. Mortality among workers at Oak Ridge National Laboratory: Evidence of radiation effects in follow-up through 1984. *J. Am. Med. Assoc.*, (265):397–402, 1991.
- [16] S. Wing, C. M. Shy, J. L. Wood, S. Wolf, D. L. Cragle, W. Tankersley, and E. L. Frome. Job factors, radiation, and cancer mortality at Oak Ridge National Laboratory: Follow-up through 1994. *Am. J. Indust. Med.*, (23):265–279, 1993.