A. APPENDIX: A Mortality Study Of Employees of The Nuclear Industry in Oak Ridge, Tennessee

A major goals of this report was to consider various methodological and technical 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 [38, 37, 39]. The analysis files (which are SAS data sets) are the starting point, see Fig. A1, for this report and are available through [8] (URL http://cedr.lbl.gov/CEDRhomepage.html). The demographic analysis file contains an exhaustive roster of all persons hired at one or more of the Oak Ridge facilities before December 31, 1982. The facility comparison results presented in Tables II-V of this report were obtained using data in the demographic analysis file. Workers employed less than thirty days or whose data contain certain "critical errors" were excluded from the mortality analysis. The radiation dose-response results presented in Tables VI–XI were restricted to white males employed at either X-10 or Y-12 and use the yearly external and internal radiation exposure data contained in the three additional analysis files shown in Fig. A1. In both the facility comparison and the dose-response analyses two or more steps are required to generate the Analytic Data Structure (ADS) which provides stratified tables for Poisson regression. This Appendix gives a detailed outline of the steps in the data analysis process.

A.1. Facility Comparison Analyses.

All of the facility comparison analyses are based on data in the demographic analysis file. Diseases of the respiratory system (ICD 460–619) for white males—see line 5 in Table III is used as an example of how to read these tables (see Step 3 below). The Mortality Analysis System (MAS) software that is used in Step 2 is an "in house" system that was not developed for external distribution. Similar results can be obtained using the the **DATAB** program in **epicure** [34].

A.1.1. *FCP Step 1- Create Working File*.

The **first step** in the data reduction process was to create a working file that contained all of the variables that are needed as input to the Mortality Analysis System (MAS) software [28]. This intermediate working file contains data for individuals who were followed for at least one year and whose data were free of critical errors. Length of employment was defined as a two level factor, and person-time accrual began at the date that a person had been followed for one year. At that time, people were categorized as employed less than or equal to one year or greater than one year. Consequently, the only time dependent variable is facility. MAS was not designed to

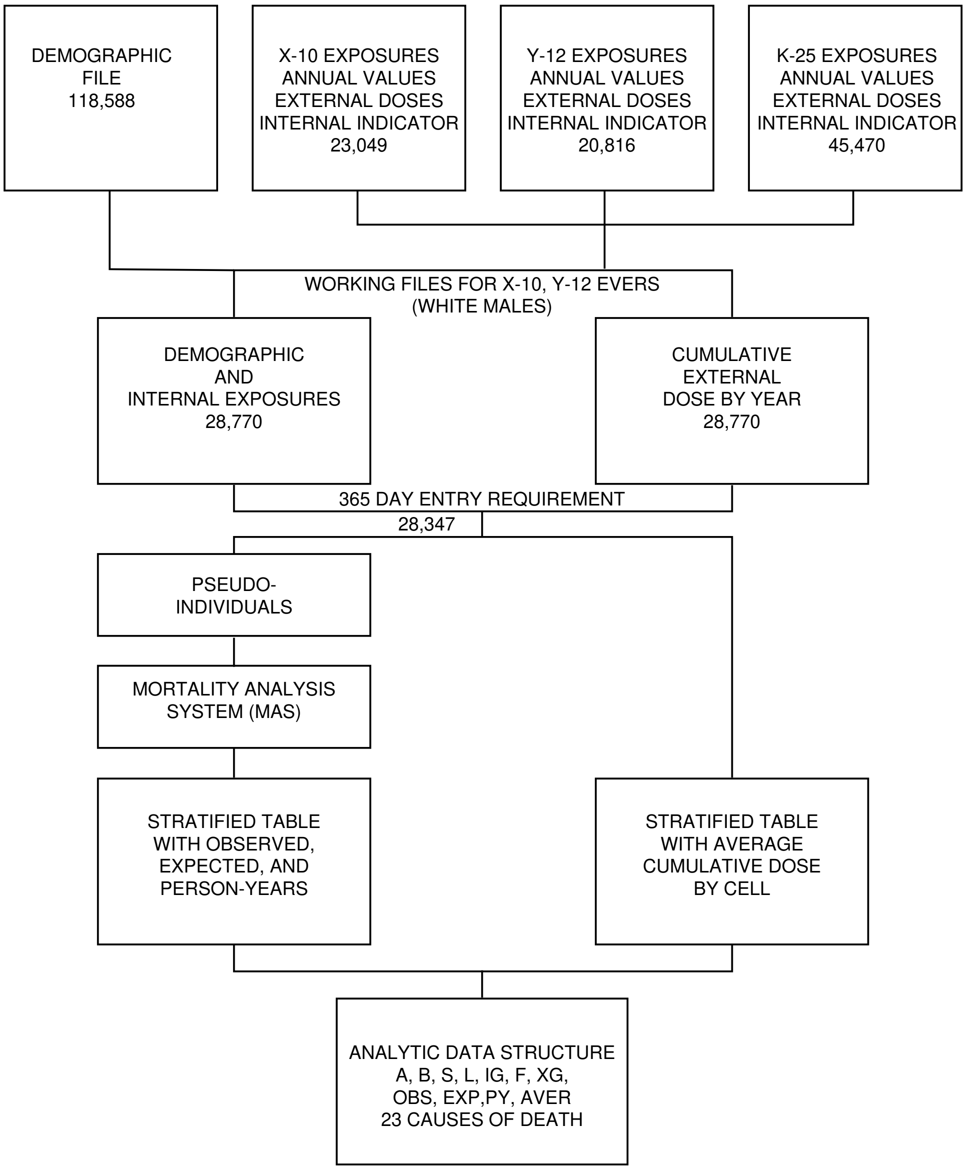


Figure AI: Analytic Data Structure

handle time dependent variables so two records were created in the working file for individuals that were employed at more than one facility. The first record contained entry and exit dates for the first facility of employment and the second record contained entry and exit dates for the multiple facility category. The variables included on each record were race, gender, facility, pay code, length of employment, vital status, birth date, and data needed o determine entry and exit dates, all of which are required by MAS to calculate person time.

A.1.2. *FCP Step 2- Define Factors***.**

The **second step** in the analysis was to define each of the factors that are included in the ADS and to specify certain options in MAS that determine how age and calendar year are represented. In the facility comparison analyses age-related effects were implicitly represented through the use of the external rates, and calendar period was included as a factor in the analysis. The resulting ADS (see Table AI) contained data for each of the maximum possible 180 cells where the person-years are greater than zero and a separate ADS was generated for each race and gender group. For example, for white males there were 175 cells and the observed deaths and SMRs for each cause are listed in Table III. The expected deaths were obtained using rates for the U.S. population, [30], i.e. the expected deaths in each cell are $\mu_j^* = \Sigma_i n_{ij}\lambda_i^*$, where n_{ij} is the person-years in the i^{th} age group of the j^{th} stratum and λ_i^* is the age-cause-specific death rate for the appropriate U.S. race and gender group.

A.1.3. *FCP Step 3- Fit Main Effects Model***.**

The **third step** was to fit a main effects model (see Eq. 2) for each race and gender group for each cause of death of interest—the results are given in Tables III–V. Consider, as an example of how to read these tables, diseases of the respiratory system (ICD 460–619) for white males—see line 5 in Table III. Table AII (see first five rows) shows the parameter estimates for each facility in log percentage (L%) units in column 2 and the corresponding estimates of the SMRs (adjusted for *S* and *L*) at the mid-point of the 1965–69 interval of follow up. These facility effect estimates estimates are listed in columns 4 through 8 or row 5 in Table III. The estimated relative risk (adjusted for *L* and *F*) for monthly compared to nonmonthly workers is exp (-78.6/100) = 0.46 with 95% confidence interval of (.35, .59) indicating a substantial 5% difference in respiratory system disease rates between monthly (white collar) and nonmonthly (blue collar) workers. The estimated relative risk (adjusted for *F* and *S*) for short term workers (< 1 year) relative to those who worked one year or longer is 1.30 with 95% CI = (1.17, 1.44). The period trend parameter estimate 0.97 (see the last line of Table AII) represents the yearly change in the log SMR in L% units over the forty years of follow-up. The parameter estimates for each cause of death in Tables III–V can be combined to estimate the SMR for each facility, at the mid-point of each follow-up interval for each level of *S* and *L*:

log (SMR) = log (facility SMR in 1965–69)
+ (effect due to *L*)
+ (effect due to *S*)
+ (Interval-Midpoint–1967.5) * 0.97.

For example, for long term nonmonthly workers at K-25, during the last interval of follow-up (1980–85), with mid-point = 1982.5, we have
log(SMR) = 6.7 + 0 + 0 + (1982.5-1967.5) * 0.97 = 21.3L%,
so that the estimated SMR is exp(0.213)= 1.24. The lowest estimated SMR was exp(-1.41)= 0.24 and occurs for:

X-10	Monthly	Long-Term	1945–49
-43.0 +	-78.6 +	0.0 +	-19.4 = -141.0 L%.

The largest estimated SMR is exp(0.474)= 1.61 and occurs for:

K-25	Non-Monthly	Short-Term	1980–84
6.7 +	0.0 +	26.1 +	14.6 = 47.4 L%.

The likelihood ratio statistic (see column 9 of Table III) for the null hypothesis of no differences among facilities was 24.4 with 4 *df* ($p<.01$). Table AII indicates that the risk of respiratory disease was much higher for TEC, K-25 and multiple facility workers than for X-10 or Y-12 workers. For example, the estimated relative risk for K-25 with X-10 as the reference facility is $exp[(6.7-(-4.0))/100] = 1.64$. A large value for the LRT statistic (in column 9 of Tables III–V) indicates that there were differences among facilities in the death rates for that cause of death after adjusting for *S*, *L*, period of follow-up, and age through use of external rates.

B. Dose-Response Analyses.

Evaluation of the radiation exposure data and the monitoring and recording procedures at each of the Oak Ridge plants is described in detail in our companion report [39]. After reviewing these results we decided to limit our dose-response analysis to the sub-cohort of 28,770 white males who were ever employed at the X-10 or Y-12 plant (see Statistical Methods). 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 [38, 39]. In this combined analysis the internal radiation exposure was limited to a categorical variable with three levels (see Materials and Methods and Table AIII) .

Table AI: Analytic Data Structure (ADS) For Facility Comparison Analyses

F	S	L	P	obs_cause 1	exp_cause 1	…	obs_cause j	exp_cause j
1	1	1	40	0	0.31	…		
1	1	1	45	7	21.7	…		
⋮	⋮	⋮	⋮	⋮	⋮			
5	2	2	75	0	1.0	…		
5	2	2	80	0	1.5	…		

where the factors are defined as follows:

F	Facility	S	Socioeconomic Status	L	Length of Employment	P	Period of Follow-Up
1=	X-10	1=	Non-Monthly	1=	1 year + 1	1=	1940
2=	TEC	2=	Monthly	2=	$\leq$1 year 2	2=	1945
3=	Y-12					3=	1950
4=	K-25					⋮	⋮
5=	MULT					9=	1980

B.1. Generating The Analytic Data Structure (ADS)

The three exposure analysis files at the top of Fig. A1 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 encountered in occupational cohort studies (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, and

iii) facility in year t (as described at the beginning of the Appendix).

The entry and exit dates for the time dependent cells also include a lag of two, ten, or twenty years. The person-epoch [11] approach with grouping on the time dependent covariates is used to generate the ADS shown at the bottom of Fig. A1. The factors in Table AIII are used to illustrate the procedure that was followed to generate the ADS for the dose-response analysis.

B.1.1. *ADS for Dose-Response: Working Files.*

The **first step** in the data reduction process was to create the “working files” (see Fig. A1) for the sub-cohort of X-10/Y-12 white males that are free of critical errors. These files were needed to deal with combined data for individuals employed at more than one facility. The *Demographic and Internal Exposure* file contains demographic data, and a yearly value that indicates the facility of employment, with a special value being assigned when a person worked

Table AII: Oak Ridge Workers Facility Comparison Analysis Parameter Estimates for the Main Effects Model for White Males for all Diseases of the Respiratory System (ICD8 460-519)

Parameter	**Estimate (L%)**	**SE**	**SMR/RR**	**95%CI**
F-Facility				
X-10 vs US	-43.0	12.3	.65[a]	(.51, .83)
TEC vs US	1.7	5.2	1.02	(.92, 1.12)
Y-12 vs US	-29.4	12.9	0.75	(.58, .96)
K-25 vs US	6.7	5.4	1.07	(.96, 1.19)
MULT vs US	6.3	7.2	1.06	(.92, 1.23)
S-Socioeconomic				
Nonmonthly	0.0	.	1.0	
Monthly	-78.6	13.6	.46[b]	(.35, .59)
L-Length of Employment				
Greater than one year	0.0	.	1.0	
One year or less	26.1	5.3	1.30[b]	(1.17, 1.44)
Period Trend	0.97	0.28	0.97[c]	(.42, 1.53)

[a]Estimated SMR's for the Facilities.
[b]Estimated relative risk for S, L, and t.
[c]Yearly change in log SMR in L% units.

at more than one facility. Facility was lagged in all analyses and for years when a person had internal exposure data from more than one facility, the data were combined by retaining the highest value for that year. A special variable to identify workers employed at TEC before being employed at X-10 or Y-12 was created. We have assumed that these individuals have no external exposure but that many of them have internal exposure and will enter follow-up classified as internal exposed workers [38]. The *Cumulative External Dose* file contains yearly cumulative external dose (with a two, ten, or twenty year lag) data from all previous years of employment at all Oak Ridge facilities.

B.1.2. *ADS Pseudo-persons for Time Dependent Variables.*

In the **second step** the variables in these two working files were used to divide each individuals 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 MAS for exact computation of person-years. Table AIII shows that there were (3 x 3 x 10 = 90) possible time dependent exposure categories. Note that a person can only move from lower to higher dose categories. The entry and exit dates for the cumulative external dose groups were obtained by interpolation within years after the appropriate lag had been introduced. The entry and exit dates for internal exposure groups were based on the mid-point of the year in which a category change occurred. Workers employed at K-25 or TEC before they worked at X-10 or Y-12 entered follow-up in the "Multiple" facility category, whereas individuals first hired at X-10 or Y-12 contributed person-years to the facility of first hire until the end of the study or until the date they became a multiple facility worker for the appropriate lag. For the situation described in Table AIII, and using a ten year lag, the 28,347 (note that 423 workers did not have at least one year of follow-up and were excluded from the analyses) workers were partitioned into 61,597 pseudo-individuals for processing by MAS. A more detailed description of the exact person-year stratification method is given in [5] Appendix IV) and [26] who also discussed other software that can be used for these computations.

B.1.3. *Run MAS Program to Obtain ADS.*

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. In the example these factors are A, B, S, L, IG, F, and X (see Table AIII). Each of these matrices was then multiplied by the appropriate rate table for each cause of death category provided as input. Since we have stratified on birth cohort and age at risk the cause specific observed and expected deaths are added over the calendar year dimension (this is an option in MAS). In our example this results in a seven dimensional table that contains 4,230 cells with person

Table AIII: 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.

years greater than zero. Each cell in the ADS contains the index values for each of the factors, the person-years, the average dose, the observed and expected deaths for each cause of death category of interest—the cause of death categories were determined by the rate table provided by [30]. The resulting ADS (see Fig. A1) for the ten year lag analyses reported in Tables VI–VIII is a SAS data set with variables named OBS*j*, EXP*j* for each value of *j* (i.e. cause of death category) as well as the indices corresponding to the factors A, B, S, L, IG, F, X. Each stratum in the ADS also contains the total person-years at risk (PY) and the radiation dose variable D, which is the average dose of all person-years within each stratum (calculated by a separate SAS program). The ADS used to obtain the results in Tables VI—VIII is available through CEDR.

The all cancer subset used to obtain the results in this Appendix can be obtained on the World Wide Web (www)—see Oak Ridge Mortality Study;Data and Computer Programs— at URL `http://www.epm.ornl.gov/~frome/`.

B.1.4. *Table with Marginal Distribution Facility By Dose Group.*

Table AIV 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. Table AIV also shows the marginal average dose, i.e., the person-year weighted average of all the average doses from each cell in the full table. The person-year distributions in this table shows that the dose distribution at Y-12 was considerably different from that of X-10 and multiple workers. The person-years in the X-10 only category is greater than reported in previous studies [17, 16] since individuals who were initially employed at X-10 and then at another Oak Ridge facility contributed person-years to the appropriate dose groups in the X-10 only category until ten years (the lag) after the date at which they became a multiple facility worker. In most dose groups the marginal average dose was slightly less than the mid-point of the interval used to define the category. In dose-response analyses the average dose in each cell was used.

In previous reports for X-10 only workers [40, 41] different dose groups were defined and the interval mid-point was used for all groups except the highest group where the median value of all person-years in that interval was used. Table AIV also shows the marginal distribution of observed and "expected" deaths (using U.S. rates) by facility and dose groups for all cancer deaths, and the marginal total for each dose-group. These unadjusted SMRs indicate that mortality rates for all cancer causes are less than those for U.S. White males. To adjust for this selection effect we used the external standard rates and fit a multiplicative main effects model that includes all of the factors other than dose group. The "adjusted" expected deaths and SMRs are listed at the bottom of Table AIV. A GLIM4 program (g-tabaiv.txt) that can be used to obtain Table AIV is provided with the all cancer data set (gdata.txt) at URL `http://www.epm.ornl.gov/~frome/ORMS/index.html` (see **Data and Computer Programs** .

B.2. Fitting The Main Effects Model.

The general Poisson regression model 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 LRT statistics were obtained using widely available software [14, 34]. The baseline rate λ^0_{jk} 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 [40, 41] a parametric model was used to describe the baseline rates, i.e.

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

where A_{jk} = age at interval mid-point. In this report the external/internal model is used ([5], Chapter 4)

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

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 used to summarize the data for the seven dimensional table of observed and expected deaths for all cancers for the X-10/Y-12 subcohort.

Table AIV: Summary Results For White Males Ever Employed at X-10 or Y-12 By Dose Group Using Ten Year Lag

		Dose Category(mSv)										
Interval		0	.01-5	5-10	10-20	20-40	40-80	80-160	160-320	320-640	640+	
Midpoint		0	2.5	7.5	15	30	60	120	240	480	.	
Average		0	1.9	7.2	14.2	27.8	55.2	110.8	219.6	414.8	1773.8	
Facility		**Person Years**										**Total**
X-10		140426	57666	17220	18288	12255	6258	3271	1681	571	89	257726
Y-12		125574	35718	12504	8661	4699	1233	239	17	0	39	188684
Mult		89456	31798	13420	9199	6046	3997	1888	872	188	91	156955
Total		355457	125182	43144	36148	23000	11488	5398	2570	759	220	603365
		Observed and Expected [a] Deaths For All Cancer Causes										
X-10	Obs	110	100	42	54	30	22	14	10	3	1	386
	Exp	153.56	138.44	53.53	65.98	42.58	24.17	11.16	6.08	2.47	0.32	498.3
Y-12	Obs	153	69	42	30	14	4	0	0	0	1	313
	Exp	149.79	72.97	45.46	28.09	16.23	5.19	0.79	0.01	0	0.14	318.7
Mult	Obs	173	117	54	39	24	17	6	3	1	1	435
	Exp	200.3	116.08	66.93	41.89	26.3	18.37	7.54	3.11	0.73	0.58	481.8
Total	Obs	436	286	138	123	68	43	20	13	4	3	1134
	Exp	503.65	327.49	165.92	135.96	85.11	47.73	19.48	9.21	3.2	1.04	1299
SMR		86.6	87.3	83.2	90.5	79.9	90.1	102.7	141.2	125.1	287.6	87.3
		Adjusted [b] Expected Deaths For All Cancer Causes										
X-10		115.72	105.71	41.78	52.05	33.13	20.37	9.52	5.31	2.14	0.29	
Y-12		148.35	69.43	44.91	27.97	16.18	5.23	0.79	0.01	0	0.15	
Mult		182.39	104	59.99	37.82	23.6	16.35	6.75	2.85	0.68	0.59	
Total		446.47	279.12	146.69	117.83	72.9	41.96	17.06	8.17	2.82	1.03	
"SMR"		97.7	102.5	94.1	104.4	93.3	102.5	117.3	159.1	141.8	290.3	

[a] Expected deaths based on U.S. white male rates.

[b] Adjusted for all factors (except external dose) in the main effects model (see Table VI).

B.2.1. *Main Effects– Internal Analysis.*

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

$$r_{jk} = exp\ \left[\log(A_{jk}/52.5) + B + S + L + IG + F\right]\,(1+D),$$

where $r_{jk} = y_{jk}/n_{jk}$ is the observed cancer death rate per thousand person years at risk in the jk^{th} cell. In this specification all of the terms are factors (one parameter for each level) except for age and dose, and birth cohort is the "referent factor". The baseline rates are described by the first six estimates in column 2 of Table AV, e.g. for the 1915 birth cohort the estimated baseline rates are log (rate) = 0.728 + 5.20 * log (age/52.5). Consequently, the estimates for the factor B in column two (lines 2 through 6) of Table AV 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), with the first level of each factor as the referent category. The next to last estimate in column two is the estimated dose-response coefficient for external radiation dose and represents the change in the ERR per Sv, which is equivalent to percent per 10 mSv (rem). Estimates of the standard errors of each of the parameter estimates are given in column 3 of Table AV, 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** [34] (see af-tabv.txt at `http://www.epm.ornl.gov/~frome/ORMS/data.html`)

B.2.2. *Main Effects– External/Internal Analysis.*

For the "external/internal" analysis we have

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

where $r_{jk} = y_{jk}/n_{jk}\lambda^*_{jk}$ is the SMR for the jk^{th} stratum and A = (age - 52.5)/100. The main differences between the internal and external 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 AV represent the SMR for each birth cohort in L% units, i.e. exp(-7.7/100) = 0.919 is the estimated SMR for the 1915 L% birth cohort at the reference level of each of the other factors. It is easy to interpret the levels of the referent factor. These estimates (lines 2 through 6 in column 4) show, for example, that the all cancer mortality rate for 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 deficit 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 in column 2. The estimate for the age term

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

		Internal[b]		External/Internal[c]	
	Term	Estimate	SE	Estimate	SE
Age		5.20	0.22	-0.27	0.38
	<1900	0.538	0.132	-9.2	13.4
	1900–09	0.731	0.102	-4.3	10.3
B -	1910–19	0.728	0.093	-10.1	9.3
	1920–29	0.565	0.103	-25.0	10.3
	1930+	0.072	0.181	-66.2	17.8
S[d]	Nonmonthly	0	.	0	.
	Monthly vs Nonmonthly	-41.0	8.3	-41.3	8.3
L -	Worked at Least 1 Year	0	.	0	.
	Worked less than 1 Year	11.8	9.1	11.3	9.1
IG -	Eligible Not Monitored	0	.	0	.
	Eligible and Monitored	5.8	7.5	4.3	7.5
	Not Eligible	2.6	8.7	2.9	8.7
F -	X-10 Only	0	.	0	.
	Y-12 Only vs X-10 Only	15.6	8.6	15.3	8.6
	Multiple vs X-10 Only	6.1	7.8	5.9	7.8
D[e] -	External Dose (Sv)	1.52	0.82	1.45	0.81
	95% CI	(0.18, 3.59)		(0.15, 3.48)	

[a]ERR main effects model: rate = exp [Age Term + B + S + L + IG + F] (1 + D)

[b]For internal analysis baseline rates are estimated using log (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.

[c]For External/Internal model baseline rates are estimated using log (λ_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 (age - 52.5)/100, was included to reflected any systematic age related departures from the external rates.

[d]The 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.

[e]ERR per Sv with 95% likelihood based confidence intervals.

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.

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 1915 birth cohort, who were long-term Y-12 only workers and were monitored for internal radiation exposure. Further, suppose that the first two were nonmonthly, and that the second two were monthly, and that two of them received cumulative external radiation doses of 0.01 Sv (1 rem). The parameter estimates from column 4 of Table AV are used to obtain:

B		S		L		IG		F		log (1 + ERR) L% 100 log (1 +1.5D)		log (SMR) L%	$\hat{SMR}$
-10.1	+	0	+	0	+	4.3	+	15.3	+	0	=	9.5	1.10
-10.1	+	0	+	0	+	4.3	+	15.3	+	1.5	=	11.0	1.12
-10.1	+	-41,3	+	0	+	4.3	+	15.3	+	0	=	-31.8	0.73
-8.5	+	-41.3	+	0	+	4.3	+	15,3	+	1.5	=	-30.3	0.74

This illustrates how to read the summary tables given in the Results Section of this report. Note in particular that the coefficients for the birth cohort factor represent a comparison (an SMR in L% units) of the internal control group with the external referent group. The coefficients for the potential confounding variables (S, L) and the exposure variables of (IG and F)are relative risk estimates (in L% units). The coefficient for D is the estimated ERR per Sv which is equivalent to percent per 10 mSv (rem).

B.3. Standard Analysis Procedure.

A standard analysis procedure was used to evaluate the relative importance of the effect of each of the factors B, S, L, IG, F, and the dose variable D for each cause-of-death of interest. The multiplicative main effects model (see Eq. 3) was used for this screening procedure since in many situations the ERR model does not yield estimates of the dose related parameters.

1. Fit the main effects model Eq. 3 using external rates and compute the deviance and df (see Table AVI line 2).

2. Fit each of the models described in the first column of Table AVI, and compute the deviance, AIC, and LRT statistic **with the main effects model as the referent model**. The first line in Table AVI gives the deviance and other summary statistics for the "minimal model", i.e., the model that has only one parameter which is the log of the SMR. For

lines 3-8 in Table AVI the LRT test shows the effect of deleting terms not included in (see column one to identify the term that is deleted) the main effects model. For this example birth cohort (B) and paycode (S) are strong predictors of cancer risk, and length of employment (L), internal exposure (IG) and facility (F) are not. This is consistent with the results (estimates and standard errors) given in Table AV. The last seven lines show the effect of adding additional parameters to the main effects model. Lines 9–13 evaluate possible interaction of each of the factors with radiation dose (i.e. effect modification). The likelihood ratio test statistics give no indication that the dose response estimate changes with levels of any of the other factors. The last line in the Table AVI contains a relative risk parameter for each dose group, and provides a "lack-of-fit" test (FM) for the multiplicative dose-response. The relative risk estimates are shown graphically in Fig. 2A. Note that while the likelihood ratio test on line 14 of Table AVI does not indicate lack-of-fit of the linear dose-response model, the results in Table X and Fig. 2A suggest that the linear excess relative risk provides a "better" description of the dose-response relation (see Results for further discussion of this point).

3. Fit a model which includes the main effects and all possible interaction term for the potential confounding variables A, B, S, and L. This is usually done by stratifying on these factors so that parameter estimates are not computed. The resulting estimates for the exposure variables (IG, F, and D) are then compared with those obtained from the main effects model to verify that there is no residual confounding associated with the interaction terms of the stratification variables. The results of this stratified analysis for the example are given in Table AVII.

B.4. Regression Diagnostics For The Exponential Relative Risk Function For All Cancer.

In previous reports and in the screening procedure used in this report 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 this is to plot estimates of the relative risk versus the average dose in each of the dose categories (e.g. see Fig. 2A). Another approach is to use regression diagnostic techniques that describe the influence of individual data points on parameter estimates [15, 5]. 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 Fig. A2 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 $\hat{\beta}$, Eq. 3 was fit to each table obtained by deleting the j^{th} cell to obtain $\hat{\beta}_{(j)}$. The

Table AVI: Summary Statistics For All Cancer with Ten Year Lag

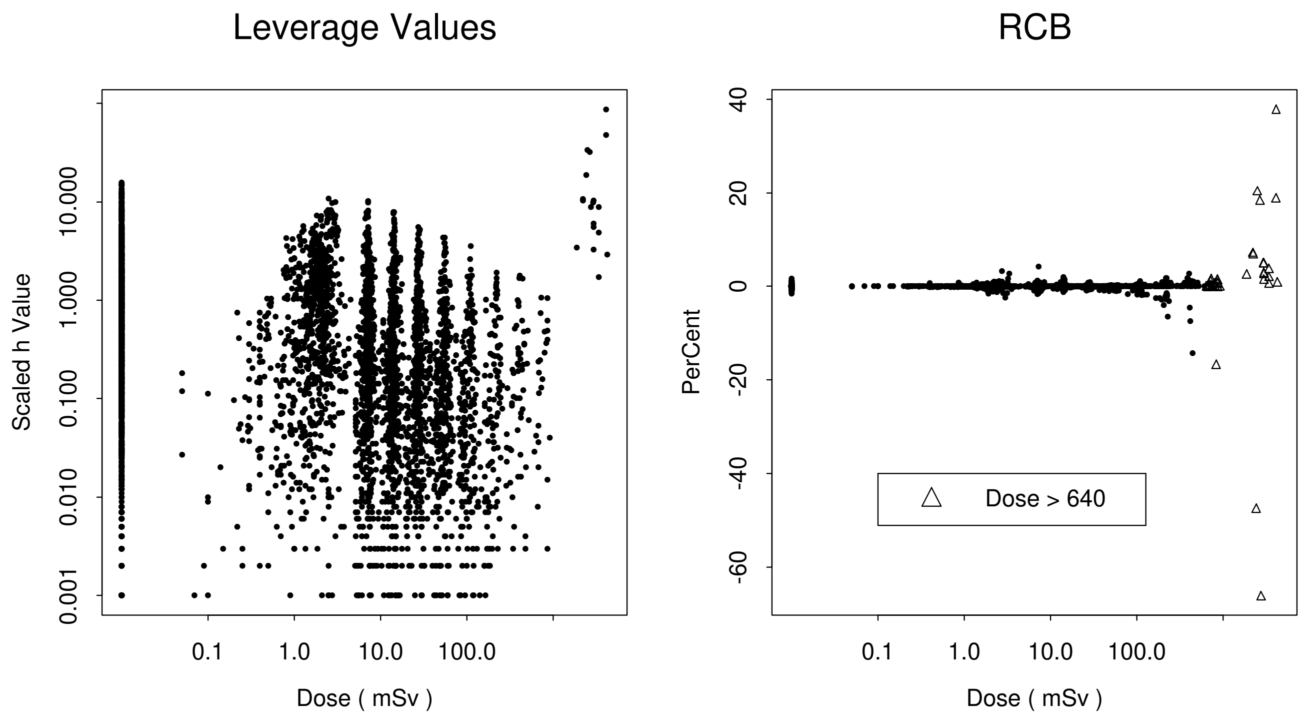
Model	**AIC**[a]	**m**[b]	**LR**[c]	**df**[d]	**p**[e]	**Dev**	**df**
Minimal	2080.77	1	62.27	12	0	2078.77	4229
Age+B+S+L+IG+F+D	2042.5	13	.	.	.	2016.5	4217
Age+B+S+L+IG+F	2044.05	12	3.55	1	0.0596	2020.05	4218
Age +S+L+IG+F+D	2047.23	9	12.73	4	0.0127	2029.23	4221
Age+B +L+IG+F+D	2067.45	12	26.93	1	0	2043.43	4218
Age+B+S +IG+F+D	2041.89	12	1.39	1	0.2385	2017.89	4218
Age+B+S+L +F+D	2039.18	11	0.68	2	0.7102	2017.18	4219
Age+B+S+L+IG +D	2041.05	11	2.55	2	0.2788	2019.05	4219
Age+B+S+L+IG+F+D.B	2043.4	17	7.1	4	0.1292	2009.4	4213
Age+B+S+L+IG+F+D.IG	2046.41	15	0.09	2	0.9571	2016.41	4215
Age+B+S+L+IG+F+D.S	2044.04	14	0.46	1	0.4966	2016.04	4216
Age+B+S+L+IG+F+D.L	2044.41	14	0.09	1	0.7692	2016.41	4216
Age+B+S+L+IG+F+D.F	2041.31	15	5.19	2	0.0746	2011.31	4215
Age+B+S+L+IG+F+XG	2054.08	21	4.42	8	0.8187	2012.08	4209

[a] AIC = devaince + 2m is is the Akaike information criterion.
[b] m = the number of parameters for the specified model
[c] LR = likelihood ratio test statistic.
[d] df = degrees of freedom for LR
[e] p = p-value for likelihood ratio test.

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 jth Cell Omitted From The Fit

ORNL-6785 ORMS: X-10/Y-12 White Male Subcohort

Figure AII: Regression Diagnostics

Table AVII: Parameter Estimates for Stratified Analysis[a] Using External Rates ALL Cancer with Ten Year Lag

	Term	Estimate(L%)	St Dev
IG-	Eligible Not Monitored	0	.
	Eligible and Monitored	4.3	7.7
	Not Eligible	-0.5	8.9
F-	X-10 only	0	.
	Y-12 only vs X-10 only	10.8	8.8
	Multiple vs X-10 only	3.6	7.8
D-	External Dose[b]	1.41	0.80
	95% CI	(0.13, 3.42)	.

[a]This corresponds to the model rate = exp(A*B*S*L+IG+F)(1 + D)
[b]ERR per Sv with likelihood based confidence interval

relative percent change $\hat{\beta}$ (RCB) due to the j^{th} cell is then

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

The right hand panel of Fig. AII 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 this dose-response relation may not be appropriate for these data. In the Results Section a more detailed evaluation of alternative dose-response functions is provided, and visual summaries which further demonstrate this point.