HEALTH EFFECTS OF LOW-LEVEL IONIZING RADIATION

Final Report

July 1979

Vincent Stanford

Jorge C. Rios, M.D.

The George Washington University

Prepared for the Nuclear Regulatory Commission

Table of Contents

ABSTRACT

The investigators analyzed mortality data provided by The Nuclear Regulatory Commission on almost four thousand former employees at the Hanford works, one of the largest nuclear processing plants in the United States.

Fifty-nine percent of these employees were exposed to low-level ionizing radiation during the course of their occupational activities: the other forty-one percent were not. The purpose of our analysis was to investigate any relationship between occupational exposure to low-level ionizing radiation and subsequent death by cancer.

The analysis revealed several important findings. The statistical procedures employed show:

- No hazard for the aggregate male and female populations using basic bivariate procedures.
- The Mantel-Haenzel procedure reveals significant heterogeniety across age groups for males with regard to the degree of association between simple exposure and subsequent cancer death.
- In the 45-54 bracket for age at death, men who were not exposed died of cancer 11.8 percent of the time while exposed men died of cancer in 20.7 percent of cases.
- An age stratified multivariate analysis shows significant association of radiological exposure variables and subsequent cancer death for both the 45-54 age bracket and the 75-84 age bracket for males.
- Multivariate analysis shows a result of borderline significance for the women of the cohort. Further investigation as more data for women becomes available is recommended.

Background and Objective of the Study

There is an intense interest on the part of the general public as well as governmental agencies in the accurate determination of the long-term health effects from exposure to low level ionizing radiation.

Therefore, from the standpoint of the public health role of the NRC, it is important to analyze those data on human exposure to low level exposures to ionizing radiation which are available.

The NRC provided a tape which contains data on occupational radiation exposure and other relevant information. The objective of this study is the analysis of the data provided. More specifically, this report undertakes to:

- Examine the relationship between exposure of individuals to low level ionizing radiation and subsequent death by cancer.

- Describe the method employed to deal with the statistical variability of the data as it impacts the performance of the above two tasks.

The relationship between exposure and subsequent cancer death is discussed in the section entitled "Basic Statistical Tests". The relationship is further examined in the section entitled "Combined Impact," which deals with multivariate composites of both radiological and demographic variables as they apply to the prediction of death by cancer.

Specific Questions and Analysis Performed

The analysis of t. Nuclear Regulatory Commission's Hanford mortality data is comprised of four major phases:

. Univariate summary which allows for basic familiarization with the character of the data.

. The search for covariates which will provide for the detection of variables impacting the dependent variable--namely, death by cancer--but which are unrelated to low level ionizing radiation.

. Basic statistical analysis of the impact of independent variables which describe the low level ionizing radiation exposure of the study population on death by cancer. This will, of course, reflect important covariates uncovered by preliminary analysis.

. Multivariate analysis to assess the combined impact of the "risk factors" of low level ionizing radiation on death by cancer. This will be done in a manner analogous to that first used in assessing risk factors in coronary heart disease in the Framingham Study.

This analysis was performed with two goals in mind. These were to:

. Resolve, as best possible, the questions motivating the analysis which are set forth presently.

. Provide a general reference document from which other investigators can answer related questions with a minimum of computer work.

We compiled a list of specific questions which could be reasonably investigated using conventional statistical techniques and the variables at hand. They are:

. Is the probability of death by cancer significantly different for the population exposed to low level ionizing radiation from

that for the unexposed population?

- Is the rate of dosage per year related to the rate of death by cancer?
- Is the total lifetime dose related to the probability of death by cancer?
- What is the combined impact of the risk factors based on low-level ionizing radiation? That is to say, to what extent can we predict who will die of cancer knowing who was exposed and the characteristics of their exposure?
- Does age at death differ from exposed versus non-exposed Hanford workers?

Cause of Death Classification

The major orders of the ICDA classification consist of:

1. Infective and Parasitic Diseases
2. Neoplasms
3. Endocrine, Nutritional and Metabolic Diseases
4. Diseases of the Blood and Blood-forming Organs
5. Mental Disorders
6. Diseases of the Nervous System and Sense Organs
7. Diseases of the Circulatory System
8. Diseases of the Respiratory System
9. Diseases of the Digestive System
10. Disease of the Genitourinary System
11. Complications of Pregnancy, Childbirth and the Puerperium
12. Diseases of the Skin and Subcutaneous Tissue
13. Diseases of the Musculoskeletal System and Connective Tissue
14. Congenital Anomalies
15. Certain Causes of Perinatal Morbidity and Mortality
16. Accidents Poisonings and Violence (Nature of Injury)
17. Accidents, Poisonings and Violence (External Cause)

Our analysis proceded along the coarsest level of grouping possible, so as to leave no question remaining about the effect of low levels of ionizing radiation on the probability of death by cancer. Therefore, we grouped the data into two classes for cause of death. These were. neoplasms (noting that benign neoplasms rarely cause death and that general population data are available for the U.S. census on death by malignant neoplasms for purposes of comparison), and other. We realize that this is not the customary division. We point, however, to the success obtained by the method

Cause of Death Classification

in establishing the relationship of radiological exposure to subsequent death by cancer as its justification.

The Nuclear Regulatory Commission provided data on the cause of death for each individual in the cohort. The causes of death were classified according to the International Classification of Diseases (ICDA - adapted for use in the United States - 8th edition).

Univariate Summary

In order to familiarize ourselves and the reader with the data provided by The Nuclear Regulatory Commission, we performed basic tabulations of the important variables and computed the following statistics:

- mean
- variance
- median
- mode

The variables studied in the statistics listed above were:

- age at death
- total years of employment
- primary cause of death examined in two ways:
  - by whether or not the person died of cancer
  - by the seventeen major categories of the International Classification of Diseases, adapted for use in the United States.
- race
- sex
- exposure
- cumulative lifetime dose
- cumulative dose at 3, 5 .....25 years before death
- year of death
- maximum radiation dose in a given year
- average radiation dose in a given year

1393 010

Table 1

NRC HANFORD LOW LEVEL RADIATION DATA -- UNIVARIATE SUMMARY

-- Age at Death

Age at Death	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)
20 - 24	11	0.3	0.3
25 - 29	37	0.9	1.2
30 - 34	73	1.8	3.0
35 - 39	109	2.7	5.8
40 - 44	199	5.0	10.7
45 - 49	318	8.0	18.7
50 - 54	404	10.1	28.8
55 - 57	534	13.4	42.2
60 - 64	539	13.5	55.7
65 - 69	603	15.1	70.8
70 - 74	485	12.1	83.0
75 - 79	379	9.5	92.5
80 - 84	186	4.7	97.1
85 - 89	97	2.4	99.5
90 - 94	14	0.4	99.9
95 - 97	2	0.1	99.9
100 - UP	2	0.1	100.0
TOTAL	3992	100.0	

Mean 59.5

Variance 176.1

Median 60.3

Mode 65

1393 011

Table 2

NRC HANFORD LOW LEVEL RADIATION DATA — UNIVARIATE SUMMARY

-- NUMBER OF YEARS WORKED AT HANFORD

NUMBER OF YEARS WORKED	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)
0 - 4	2072	51.9	51.9
5 - 9	721	18.1	70.0
10 - 14	501	12.6	82.5
15 - 19	378	9.5	92.0
20 - 24	219	5.5	97.5
25 - 29	82	2.1	99.5
30 - 34	14	0.4	99.9
TOTAL	3987	100.0	

Missing Cases 5

Mean 5.3

Median 2.3

Mode 0-4

1393 012

Table 3

NRC HANFORD LOW LEVEL RADIATION DATA -- UNIVARIATE SUMMARY

- CANCER

	PRESENCE OF CANCER	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	ADJUSTED FREQ (PCT)	CUM FREQ (PCT)
NOT CANCER	0	3177	79.6	79.6	79.6
CANCER	1	815	20.4	20.4	100.0
	TOTAL	3992	100.0	100.0	

1393 013

Table 4

NRC HANFORD LOW LEVEL RADIATION DATA --- UNIVARIATE SUMMARY

- ICD CODE FOR CAUSE OF DEATH

	ICD	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)
INFECTIVE PARASITIC	1.	37	0.9	0.9
NEOPLASMS	2.	815	20.4	21.3
ENDO NUTRI METABOLIC	3.	69	1.7	23.1
BLOOD ORGANS	4.	6	0.2	23.2
MENTAL DISORDERS	5.	17	0.4	23.6
NERVOUS SENSE ORGANS	6.	28	0.7	24.3
CIRCULATORY	7.	2022	50.7	75.0
RESPIRATORY	8.	207	5.2	80.2
DIGESTIVE	9.	164	4.1	84.3
GENITOURINARY	10.	49	1.2	85.5
SKIN SUBCUTANEOUS	12.	2	0.1	85.6
MSKEL CONNECTIVE	13.	11	0.3	85.8
CONGENITAL ANOMALIES	14.	10	0.3	86.1
SYMTOMS CONDITIONS	16.	40	1.0	87.1
ACCIDENT POISON VIOL	17.	515	12.9	100.0
	TOTAL	3992	100.0	

1393 014

Table 6

NRC HANFORD LOW LEVEL RADIATION DATA - UNIVARIATE SUMMARY

MALE OR FEMALE

	Sex	Absolute Freq	Relative Freq (PCT)	Cum Freq (PCT)
FEMALE	0.	382	9.6	9.6
MALE	1.	3610	90.4	100.0
	TOTAL	3992	100.0	

1393 015

Table 7

NRC HANFORD LOW LEVEL RADIATION DATA - UNIVARIATE SUMMARY

EXPOSED OR NOT EXPOSED

	Exposure	Absolute Freq	Relative Freq (PCT)	Cum Freq (PCT)
NOT EXPOSED	0.	1638	41.0	41.0
EXPOSED	1.	2354	59.0	100.0
	TOTAL	3992	100.0	

1393 016

Table 8

NRC HANFORD LOW LEVEL RADIATION DATA -- UNIVARIATE SUMMARY

- YEAR OF DEATH

YEAR OF DEATH	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)
45 - 49	73	1.8	1.8
50 - 54	303	7.6	9.4
55 - 59	478	12.0	21.4
60 - 64	797	20.0	41.4
65 - 69	1003	25.1	66.5
70 - +	1338	33.5	100.0
TOTAL	3992	100.0	

1393 017

Table 5

NRC HANFORD LOW LEVEL RADIATION DATA - UNIVARIATE SUMMARY

RACE CODE

	Race	Absolute Freq	Relative Freq (PCT)	Cum Freq (PCT)
NONWHITE	0.	28	0.7	0.7
WHITE	1.	3964	99.3	100.0
	TOTAL	3992	100.0	

1393 018

Table 9

NRC HANFORD LOW LEVEL RADIATION DATA — UNIVARIATE SUMMARY

- LIFETIME RADIATION DOSE

CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)
0 - 9	1867	46.8	46.8
10 - 19	330	8.3	55.0
20 - 29	216	5.4	60.4
30 - 39	205	5.1	65.6
40 - 49	164	4.1	69.7
50 - 59	123	3.1	72.8
60 - 69	108	2.7	75.5
70 - 79	86	2.2	77.6
80 - 89	92	2.3	79.9
90 - 99	73	1.8	81.8
100 - 199	252	6.3	88.1
200 - 299	183	4.6	92.7
300 - 399	75	1.9	94.5
400 - 499	40	1.0	95.5
500 - 599	23	0.6	96.1
600 - 699	35	0.9	97.0
700 - 799	14	0.4	97.3
800 - 899	12	0.3	97.6
900 - 999	8	0.2	97.8
1000 - 1999	26	0.7	98.5
2000 - 2999	30	0.8	99.2
3000 - 3999	25	0.6	99.9
4000	5	0.1	100.0
TOTAL	3992	100.0	

Mean 98.5
Variance 119637.4
Median 8.9

1393 019

Table 10

NRC HANFORD LOW LEVEL RADIATION DATA -- UNIVARIATE SUMMARY

- TOTAL DOSE 3 YEARS BEFORE DEATH

CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)		
0 - 9	[illegible]	48.7	48.7		
10 - 19	324	8.1	56.8		
20 - 29	221	5.5	62.3		
30 - 39	206	5.2	67.5		
40 - 49	165	4.1	71.6		
50 - 59	121	3.0	74.6		
60 - 69	109	2.7	77.4		
70 - 79	89	2.2	79.6		
80 - 89	93	2.3	81.9		
90 - 99	75	1.9	83.8		
100 - 199	208	5.2	89.0		
200 - 299	179	4.5	93.5		
300 - 399	65	1.6	95.1		
400 - 499	37	0.9	96.1		
500 - 599	19	0.5	96.5		
600 - 699	32	0.8	97.3		
700 - 799	7	0.2	97.5		
800 - 899	13	0.3	97.8		
900 - 999	8	0.2	98.0		
1000 - 1999	29	0.7	98.8		
2000 - 2999	26	0.7	99.4		
3000 - 3999	19	0.5	99.9	Mean	87.5
4000 -	4	0.1	100.0	Variance	99174.6
Total	3992	100.0		Median	6.6

1393 020

Table 11

TOTAL DOSE 5 YEARS BEFORE DEATH

CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)
0 - 9	2011	50.4	50.4
10 - 19	325	8.1	58.5
20 - 29	233	5.8	64.4
30 - 39	200	5.0	69.4
40 - 49	167	4.2	73.5
50 - 59	123	3.1	76.6
60 - 69	108	2.7	79.3
70 - 79	80	2.0	81.3
80 - 89	90	2.3	83.6
90 - 99	69	1.7	85.3
100 - 199	220	5.5	90.8
200 - 299	145	3.6	94.5
300 - 399	53	1.3	95.8
400 - 499	25	0.6	96.4
500 - 599	25	0.6	97.0
600 - 699	23	0.6	97.6
700 - 799	12	0.3	97.9
800 - 899	10	0.3	98.2
900 - 999	9	0.2	98.4
1000 - 1999	23	0.6	99.0
2000 - 2999	27	0.7	99.6
3000 - 3999	13	0.3	100.0
4000 -	1	0.0	100.0
TOTAL	3992	100.0	

Mean 76.

Variance 73949.

Median 4.

1393 021

Table 14

NRC HANFORD LOW LEVEL RADIATION DATA - UNIVARIATE SUMMARY

TOTAL DOSE 20 YEARS BEFORE DEATH

CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)
0 - 9	3274	82.0	82.0
10 - 19	196	4.9	86.9
20 - 29	114	2.9	89.8
30 - 39	88	2.2	92.0
40 - 49	61	1.5	93.5
50 - 59	43	1.1	94.6
60 - 69	34	0.9	95.4
70 - 79	25	0.6	96.1
80 - 89	32	0.8	96.9
90 - 99	24	0.6	97.5
100 - 199	54	1.4	98.8
200 - 299	36	0.9	99.7
300 - 399	8	0.2	99.9
400 - 499	1	0.0	99.9
500 - 599	2	0.1	100.0
TOTAL	3992	100.0	

MEAN 9.1

VARIANCE 1007.7

MEDIAN 1.1

1393 022

Table 12

NRC HANFORD LOW LEVEL RADIATION DATA -- UNIVARIATE SUMMARY

- TOTAL DOSE 10 YEARS BEFORE DEATH

CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)	
0 - 9	2328	58.3	58.3	
10 - 19	305	7.6	66.0	
20 - 29	223	5.6	71.5	
30 - 39	189	4.7	76.3	
40 - 49	158	4.0	80.2	
50 - 59	114	2.9	83.1	
60 - 69	95	2.4	85.5	
70 - 79	66	1.7	87.1	
80 - 89	78	2.0	89.1	
90 - 99	53	1.3	90.4	
100 - 199	145	3.6	94.0	
200 - 299	99	2.5	96.5	
300 - 399	45	1.1	97.6	
400 - 499	19	0.5	98.1	
500 - 599	16	0.4	98.5	
600 - 699	16	0.4	98.9	
700 - 799	7	0.2	99.1	
800 - 899	4	0.1	99.2	
900 - 999	3	0.1	99.3	
1000 - 1999	18	0.5	99.7	Mean 42.9
2000 - 2999	10	0.3	100.0	Variance 23239.3
3000 - +	1	0.0	100.0	Median 3.57
TOTAL	3992	100.0		

1393 023

Table 13

NRC HANFORD LOW LEVEL RADIATION DATA - UNIVARIATE SUMMARY

TOTAL DOSE 15 YEARS BEFORE DEATH

CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)
0 -9	2733	68.5	68.5
10 -19	266	6.7	75.1
20 -29	181	4.5	79.7
30 -39	163	4.1	83.7
40 -49	118	3.0	86.7
50 -59	72	1.8	88.5
60 -69	63	1.6	90.1
70 -79	46	1.2	91.2
80 -89	51	1.3	92.5
90 -99	38	1.0	93.5
100 -199	116	2.9	96.4
200 -299	77	1.9	98.3
300 -399	33	0.8	99.1
400 -499	9	0.2	99.3
500 -599	8	0.2	99.5
600 -699	8	0.2	99.7
700 -799	3	0.1	99.8
800 -899	1	0.0	99.8
900 -999	3	0.1	99.9
1000 -1999	3	0.1	100.0
TOTAL	3992	100.0	

1393 024

MEAN	23.0
VARIANCE	5175.6
MEDIAN	2.3

Table 15

NRC HANFORD LOW LEVEL RADIATION DATA---UNIVARIATE SUMMARY

TOTAL DOSE 25 YEARS BEFORE DEATH

CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	CUM FREQ (PCT)
0 - 9	3750	93.9	93.9
10 - 19	87	2.2	96.1
20 - 29	36	0.9	97.0
30 - 39	25	0.6	97.6
40 - 49	20	0.5	98.1
50 - 59	10	0.3	98.4
60 - 69	8	0.2	98.6
70 - 79	9	0.2	98.8
80 - 89	11	0.3	99.1
90 - 99	9	0.2	99.3
100 - 199	18	0.5	99.8
200 - 299	8	0.2	100.0
300 - 399	1	0.0	100.0
TOTAL	3992	100.0	

MEAN 2.5

VARIANCE 221.2

MEDIAN 0.32

1393 025

The Method of Proportional Mortality

The method of proportional mortality, a statistical technique, is used for a major portion of the analysis in this report. The method, while highly useful, must be applied carefully.

It is particularly useful in cases whose morbidity or mortality data is available but not data on the population at risk where the diseases or deaths occurred. This is the situation with the NRC data.

The analysis cannot, therefore, consider absolute rates of death from a particular cause. Instead, the relative death rates from a cause or group of causes can be shown.

In some cases, the method can artifically show a high death rate. For instance, where two populations have identical cancer death rates, and the first is fortunate in a particularly low rate of death from other causes. There, the first group would falsely appear to have a proportionally higher cancer death rate. In the Hanford data, such a death rate may appear proportionally higher in the exposed group than the non-exposed.

In spite of the drawbacks of the method, the nature of the data necessitated its use, and the results contained herein must be viewed with caution.

Mantel and others have pointed out that while the method is widely used it should only be taken to provide leads for rigorous research. In general, a single retrospective study should not be taken as conclusive.

Tests for Covariates

There are radiological and demographic variables in the data file provided by The Nuclear Regulatory Commission for this study. This section describes statistical tests which were performed in order to determine which of the demographic variables are related to death by cancer.

It is widely known that cancer death rates differ for various subgroups of the population of the United States. For example, the U.S. Decennial Life Tables for 1969-1971 show that the chance of eventually dying of a malignant neoplasm is:

- 16.3 percent for the total population
- 16.9 percent for white males
- 15.9 percent for white females
- 15.3 percent for non-white males
- 13.5 percent for non-white females.

These variations indicate that sex and race should be taken into account in the analysis to follow.

Another important source of heterogeniety in cancer death rates is age. Death rates, generally, increase with age and, in particular, cancer death rates rise rapidly with age. The cancer death ratio (i.e., the fraction of total deaths which are cancer deaths) also varies with age.

The question which we address in this section is a simple one: what important intervening factors effect the probability of death by cancer? The primary statistical method which was employed to answer this question was that of simple cross tabulation. The cancer vs. not cancer cause of death indicator was cross tabulated by each of the covariates of interest. The results of these cross tabulations together with statistical tests of significance are presented below.

Tests for Covariates

A Chi square test of a four-fold table with not-cancer/cancer vs. male/female was performed:

Table 16

	female	male	
not cancer	265	2912	3177
cancer	117	698	815
	382	3610	3992

The results of the Chi square test of the hypothesis that sex and death by cancer are unrelated were:

. Chi square = 26.4

. df = 1

. p (Chi square > 26.4) < .0001

This indicates that sex is a significant covariate of death by cancer in the Hanford cohort. It is to be noted that the table contains thirty-nine more cases of death by cancer for women than would be expected if there were no relationship between cancer death and sex.

It is also of interest to note the cancer death rate for the women of the Hanford cohort is higher than thirty percent. This is substantially higher than the 15.9 percent for white females which was reported in the Decennial Life Table as mentioned earlier.

In order to determine if race is a significant covariate of cancer death, a Chi square test was performed on a four-fold table with not-cancer/cancer vs. race as non-white/white. The resulting table is shown below.

1393 028

Tests for Covariates

Table 17

	non-white	white	
not-cancer	25	3152	3177
cancer	3	812	815
	28	3964	3992

A Chi square test of the hypothesis that race and cancer death are unrelated was performed. The results of that test were:

. Chi square = 1.09

. df = 1

. p (Chi square > 1.09) = .297

This result must be regarded with caution for two reasons:

. The lowest cell frequency is only 3, and the chi square test generally requires 5 or more in each cell to be accurate.

. In general, there are very few non-whites in the cohort.

For practical purposes the race variable will be disregarded in subsequent analysis, but in no sense is this a generalizable conclusion.

Next, stratifying age into moderately broad intervals, the relationship between cancer death and age was investigated. Since sex was found to be an important covariate, this test was performed for each of the sexes.

The contingency tables for age at death vs. not-cancer/cancer are presented along with the Chi square test of the hypothesis that age at death is unrelated to the cancer death ratio (i.e. probability of death by cancer given death at age t). It is seen that, for both sexes, the cancer death ratio is strongly dependent on age at death. Therefore, all subsequent

Table 18

MALES

Age at Death

	25-34	35-34	45-54	55-64	65-74	75-84	85-94	95-104	105-115	
Not Cancer	66	169	439	717	797	564	154	5	1	2912
Cancer	2	31	92	191	253	113	16	0	0	698
	68	200	531	908	1050	677	170	5	1	3610

. Chi square = 47.12

. df = 8

. p (Chi square > 47.12) < .00009

Table 19

FEMALES

Age at Death

	25-34	34-45	45-54	55-64	65-74	75-84	85-94	105-115	
Not Cancer	11	30	58	47	63	43	12	1	265
Cancer	2	18	29	40	16	12	0	0	117
	13	48	87	87	79	55	12	1	382

- Chi square = 24.18
- df = 7
- p (Chi square > 24.18) = .0011

Tests for Covariates

analysis shall statistically adjust for age at death.

To summarize the results of this section, the subsequent analysis will:

- Control for sex-based differences in cancer death.
- Control for age at death.
- Disregard race since few non-whites are in the cohort.

Basic Statistical Tests

This section describes the application of various commonly used statistical tests to the Hanford data. Its primary purpose is to explore the relationship of certain of the independent variables which describe the character of exposure to low level ionizing radiation in the cohort to death by cancer. The analyses described in this section are primarily bivariate tests of one of the exposure variables at a time as it relates to death by cancer. These analyses are grouped under the specific questions which each is designed to investigate, and which were set forth in the section on the goals of the study and specific questions to be investigated.

One statistical procedure was also performed to assess the relationship between exposure and subsequent death by cancer as it is shown in a set of four-fold tables which result from the stratification of the cohort into several age brackets. This is the Mantel-Haenzel procedure. It is used to assess two factors of interest to this study. These are:

- The average degree of association of exposure with death by cancer.
- The degree of homogeniety across the age brackets into which the cohort was stratified.

The details of the Mantel-Haenzel procedure will be summarized in the text of the report. Further explanation can be found in Fleiss which is listed in the bibliography.

One of the simplest and most important questions under investigation in this study is: is the probability of death by cancer different for exposed versus unexposed populations in the Hanford cohort. First the relationship in the aggregate population was investigated.

Basic Statistical Tests

A Chi square test of a four-fold table having not-cancer versus cancer crossed with not-exposed versus exposed was performed. The table which resulted follows:

Table 20

	Exposed	Not-Exposed	
Not-Cancer	1311	1866	3177
Cancer	327	488	815
	1638	2354	3992

A Chi square test of the hypothesis that exposure and cancer death are unrelated in the total population under study was performed. The results of that test were:

. Chi square = .304

. df = 1

. p (Chi square > .304) = .58

This can be interpreted to mean that simple occupational exposure is not significantly related to subsequent death by cancer or that any such relationship as exists is not clear until some of the covariates which effect cancer death rates are taken into account.

Next,the cohort was stratified by sex to clarify the nature of relationship between exposure and cancer death rates for each of the sexes.

The data provided on the Hanford cohort contains almost four thousand cases and over eight hundred deaths by cancer. This allows the question under investigation to be investigated using asymtotically normal procedures. That is to say, the difference in the mean rates of death by cancer for the exposed vs. the non-exposed populations is normally distributed. This

difference can be normalized to unit variance and zero mean in the usual way. The statistical tests of hypothesis which is thus generated are equivalent to Chi square test of the corresponding four-fold tables. However, the rate tables given more easily lend themselves to interpretation. These rates are described in the following table:

Table 21

	$\hat{p}$ (cancer)	σ	n
Not-exposed	.1996	.3998	1638
Exposed	.2073	.4055	2354
Total	.2042	.4031	3992

$d = \hat{P}$ not exposed $- \hat{P}$ exposed

The normalized difference observed for the two proportions and the test of the hypothesis that they are equal resulted in the following:

. $d/\sigma_d = -.5935$

. $p(|d/\sigma_d| > .5935) = .55$

This can be interpreted to indicate that cancer death rates were not significantly different for exposed versus non-exposed populations. As might be expected, this statistic is in nearly perfect agreement with the previous chi square test. However, sex was found to be a significant intervening variable with death by cancer. Therefore, the above test was repeated for males only, and females only, with the following results.

1393 035

Basic Statistical Tests

Table 22

Males Only $d = \hat{P}$ not-exposed - $\hat{P}$ exposed

	$\hat{p}$ (cancer)	σ	n
Not-exposed	.1786	.3831	1372
Exposed	.2024	.4019	2238
Total	.1934	.3950	3610

The normalized difference and the test of the hypothesis that the cancer rates are equal for the exposed and unexposed population resulted in:

- $d/\sigma_d = -1.75$
- $p\ (|\ d/\sigma_d\ | > 1.75) = .08$

This may be interpreted to indicate that there is a difference in the cancer rates, which is at most marginally significant for exposed versus non-exposed males.

The classical values used to indicate significance are, of course, .05 and .01. This p-value is not as small as either of these. In cases where loss is very high (such as increased cancer deaths) such a result might at least prompt interest and further investigation even though risk seems low or uncertain.

The cancer rates in the females of the cohort were tested by identical means. The results of that test are presented below:

1393 036

Basic Statistical Tests

Table 23

Females only $d = \hat{P}$ not-exposed - $\hat{P}$ exposed

	$\hat{p}$ (cancer)	σ	n
Not-Exposed	.3083	.4626	266
Exposed	.3017	.4610	116
Total	.3063	.4622	382

The normalized difference and the test of the hypothesis of equality of cancer rates for exposed versus non-exposed women resulted in the following:

. $d/\sigma_d = .13$

. $p\ (|d/\sigma_d| > .13) = .9$

This indicates that females who are exposed to low level ionizing radiation die of cancer less frequently than those who are not. However, there is no reason to reject the null hypothesis of equality of the rates. Therefore, the observed difference is quite possibly accidental.

To summarize the above results: Simple occupational exposure to low-level ionizing radiation does not appear to be related to the chance of death by cancer.

The final analysis of this section which is addressed to the question of the relatedness of simple exposure to death by cancer employs the Mantel-Haenzel procedure. This will provide even further stratification of the cohort to remove the effects of the significant covariates discovered earlier. To this end a set of four-fold tables were generated. One of these for each of a set of moderately broad age strata, and of course, for each sex. The tables were then combined into the following layouts and the Mantel-Haenzel statistics computed. The Mantel-Haenzel (M-H) procedure addresses three questions:

1393 037

Basic Statistical Tests

- Is there evidence that the degree of association is consistent from one age group to another?
- If the degree of association is consistent is it also statistically significant?
- Assuming that the common degree of association is significant, what is the best estimate of its magnitude?

To answer these questions the M-H procedure involves the computation the three Chi square distributed statistics. These are:

- $X^2_{\text{total}} = \sum_{i=1}^{g} W_i M_i^2$ with g degrees of freedom
- $X^2_{\text{assoc}} = (\sum_{i=1}^{g} W_i M_i)^2 / (\sum_{i=1}^{g} W_i)$ with one degree of freedom
- $X^2_{\text{homog}} = X^2_{\text{total}} - X^2_{\text{assoc}}$ with g-i degrees of freedom

This becomes the M-H procedure with the defintion of:

$$m_i = d_i = \frac{n_{i.} - 1}{n_{i.}} \quad \frac{P_{i1} - P_{i2}}{\bar{P}_i \bar{Q}_i}$$

and

$$w_i = \frac{\bar{P}_i \bar{Q}_i n_{i1} n_{i2}}{N_i - 1}$$

This requires the definition of:

$$\bar{P}_i = \frac{N_{i1} P_{i1} + N_{i2} P_{i2}}{N_i}$$

and

$$\bar{q}_i = 1 - \bar{P}_i$$

1393 038

Further in the ith group, n_{i1} is the number of people not exposed and P_{i1} is the proportion of the unexposed subjects with cancer as cause of death. The quantity n_{i2} is the number of subjects in the ith group who were exposed, and P_{i2} is the proportion of exposed subjects with cancer as cause of death. The total number of subjects in the ith group is given

as $n_i = n_{i1} + n_{i2}$.

Proceeding now by sex we obtained the following results by the application of the M-H procedure:

Table 24

Males

	Not Exposed		Exposed				
Age Group	n_{i1}	P_{i1}	n_{i2}	P_{i2}	$n_{i.}$	d_i	w_i
25-34	29	.069	39	0	68	2.41	3.7
35-44	84	.143	116	.164	200	-.16	6.38
45-54	203	.118	328	.207	531	-.621	17.94
55-64	340	.215	568	.208	908	.042	35.28
65-74	375	.235	675	.244	1050	-.049	44.1
75-84	267	.146	410	.180	677	-.244	22.49
85-94	70	.1	100	.09	170	.177	3.51
95-104	4	0	1	0	5	0	0
105-114	0	0	1	0	1	0	0

From this we have (discarding ages 95 and above for insufficient data):

. X^2 total = $\sum_{i=1}^{7} W_i d_i^2 = 30.13$

. X^2 assoc = $(\sum_{i=1}^{7} W_i d_i)^2 / (\sum_{i=1}^{7} W_i) = 9^2/133.4 = .607$

. X^2 homog = X^2total - X^2assoc = 30.13 - .607 = 29.52

Testing the hypothesis that the degree of association is homogenous from age bracket to age bracket we have:

. Chi square = 29.52

1393 039

Basic Statistical Tests

. df = 7-1 = 6

. p (Chi square > 29.52) < .001

This strongly indicates that the degree of association varies from age bracket to age bracket. Therefore, it is not possible to analyze the common degree of association. The degree of significance of association between exposure and subsequent cancer death must be examined for each individual age bracket. Since the actual four-fold tables can be reconstructed from the above layout only the Chi square statistics and the significance of each is shown below.

Table 25

Males

Age Group	Chi Square	Significance
25-34	.88	.347
35-44	.042	.837
45-54	6.34	.011
55-64	.027	.869
65-74	.078	.779
75-84	1.14	.285
85-94	.002	.962

From the above it can be seen that there is a statistically significant degree of association in only one age bracket: namely 45-54. The four-fold table for that age bracket is presented below.

Table 26

Males Age 45-54

	Not Cancer	Cancer	
Not Exposed	179	24	203
Exposed	260	68	328
	439	92	531

1393 040

Basic Statistical Tests

As mentioned above:

- Chi square 6.34
- df = 1
- p (Chi square > 6.34) = .011

The odds ratio for this age bracket is given as:

$$O = \frac{P_{i1}(1-P_{i2})}{P_{i2}(1-P_{i1})} = \frac{.118\ (1-.207)}{.207\ (1-.118)} = .51$$

This indicates that subjects who were not exposed in this age bracket died of cancer only fifty-one percent as often as those who were exposed. More directly: 11.8 percent of those who were not exposed died of cancer, while 20.7 percent of those exposed died of cancer. To summarize: given death between the ages of 45-54, simple occupational exposure is associated with a two-fold higher cancer death rate than the rate for unexposed men in the cohort.

Applying the M-H procedure to the age stratified women of the cohort, the following layout was obtained: (see next page).

1393 041

TABLE 27

Females

Age Group	not exposed		exposed				
	n_{i1}	P_{i1}	n_{i2}	P_{i2}	n_i	d_i	W_i
25-34	8	0	5	.4	13	-2.83	.434
35-44	40	.375	8	.375	48	0	1.596
45-54	54	.37	33	.273	87	.432	4.6
55-64	59	.475	28	.429	87	.185	4.77
65-74	53	.189	26	.231	79	-.26	2.86
75-84	41	.22	14	.214	55	.035	1.81
85-94	12	0	2	0	12	0	0
105-114	1	0	0	0	1	0	0

From this we have (discounting ages 85 and above for insufficient data):

. x^2 total $= \sum_{i=1}^{7} W_i d_i^2 = 4.69$

. x^2 assoc $= (\sum_{i=1}^{7} W_i d_i)^2 / (\sum_{i=1}^{7} W_i) = (.961)^2/(16.07) = .057$

. x^2 homog $= x^2$ total $- x^2$ assoc $= 4.96 - .057 = 4.63$

Testing the hypothesis that the degree of association is homogeneous from age bracket to age bracket we have:

. Chi square = 4.63

. df = 7 - 1 = 6

. p(chi square >4.63) $\approx$.59

1393 042

Basic Statistical Tests

We find no reason to reject the hypothesis that the degree of association between exposure and subsequent cancer death is homogeneous from age bracket to age bracket.

The hypothesis that this average degree of association is zero resulted in the following:

- chi square = .057
- df = 1
- p(chi square > .057) $\approx$ .81

This does not indicate that, for the women of the cohort, simple occupational exposure is associated with subsequent cancer death.

Next, we shall describe the analysis done to investigate the question: Is the rate of dosage related to the incidence of death by cancer? The rate of dosage was derived from the data available on the cohort as follows:

dose rate = (cumulative lifetime dose)/(total years of exposure)

It must be observed that this indicator is, at best, a crude estimate, and that the dosages involved were certainly not accumulated uniformly over the course of exposure. However, in the interest of such information as is contained in this index, we performed the following analysis. The difference in mean dosage rates per year for the cancer vs. the non-cancer groups was tested for statistical significance. We proceeded, as usual, with the total population first, then the male and female populations separately. The analysis and results follows:

1393 043

TABLE 28

Total Population ($d = \bar{R}$ not cancer $- \bar{R}$ cancer)

	mean rate	σ	n
not cancer	7.8	25.93	3177
cancer	8.2	24.93	815
all	7.8	25.73	3992

The test which was computed and the statistical significance of the observed difference between the mean rates for the cancer vs. the non-cancer groups were as follows:

- $d/\sigma_d = -.395$
- $P(|d/\sigma_d| = .395) = .692$
- H_0: $d = 0$ vs. H_1: $d \neq 0$

There is a slight difference in the mean rates of exposure for the cancer vs. the non-cancer groups. However, we would expect a result which was this different, or more so, seven tries in ten by chance alone. This definitely gives no indication that H_0 should be rejected in favor of H_1. Proceeding as above for males only we have:

1393 044

TABLE 29

Males Only ($d = \bar{R}$ not cancer $- \bar{R}$ cancer)

	mean rate	σ	n
not cancer	8.36	26.97	3177
cancer	9.15	26.61	815
all	8.51	26.9	3992

The test statistic and the results of the test of the hypothesis that the observed difference between the males who died of cancer vs. those who did not is due to chance alone are presented below:

- $d/\sigma_d = -.697$
- $P(|d/\sigma_d| > .697) = .1038$
- H_o: $d = o$ vs. $d \neq o$

This result indicates a marginal but not classical degree of significance between the cancer vs. the non-cancer groups. Again, if the risk is great, this degree of difference certainly prompts further research. Now, for the female groups in the cohort we obtain:

TABLE 30

Females Only ($d = \bar{R}$ not cancer $- \bar{R}$ cancer)

	mean rate	σ	n
not cancer	1.82	5.35	3177
cancer	2.60	8.14	815
all	2.06	6.33	3992

1393 045

Basic Statistical Tests

The test statistic and the results of the test of the hypothesis that the observed difference is due to chance factors alone for females who did not die of cancer vs. those who did are as follows:

. $d/\sigma_d = -1.11$

. $P(|d\sigma d|>1.11) = .2669$

. H_o: $d = o$ vs. $d \neq o$

Another non-significant result. Therefore we cannot reject H_o with any confidence.

To summarize the results of the above analysis: The average rate at which occupational exposure to low-level ionizing radiation was incurred does not appear to be significantly different for those who die of cancer vs. those who do not. It is to be remembered that the variable used above is only a crude estimator and that the dosages involved almost certainly did not occur uniformly across the interval of exposure.

Next, we shall describe the analysis which was performed to investigate the question: Is the total lifetime dose of radiation related to the probability of death by cancer?

Since it is difficult to compute the probability of death by cancer as a function of the independent variable at hand, we will again stratify the variable for those cases who died from cancer vs. those who did not. Tnis means we will test the hypothesis that the mean life-time dose for the two groups is, in fact, equal. The difference of the mean lifetime exposure for cancer vs. not cancer population was tested:

Table 31

	Mean life dose	σ	n
Not cancer	95.0	333.3	3177
Cancer	113.7	382.8	815
All	98.8	344.1	3992

1393 046

Basic Statistical Tests

The results were:

- $d/\sigma d = -1.38$
- $P(|d/\sigma_d| > 1.38) = .1662$
- $H_o : d = o$ vs. $d \neq o$

Which provides no evidence for the rejection of the null hypothesis that the group means are identical.

Proceeding as above for the male population in the cohort we found:

Table 32

Males Only

$d = \bar{D}$ not cancer - $\bar{D}$ Cancer

	Mean Life Dose	σ	n
Not Cancer	101.77	346.81	2912
Cancer	126.41	407.34	698
All	106.53	359.38	3610

The test statistic and the results of the test of the hypothesis that the observed difference in the mean lifetime dose differs for males who died vs. those who did not die of cancer resulted in the following:

- $d/\sigma_d = -1.627$
- $P(|d/\sigma_d| > 1.627) = .104$
- $H_o : d = o$ vs. $H1: d \neq o$

We can interpret this to give us at best marginal reason to reject the hypothesis that the lifetime dose of radiation in males differs by cause of death.

Proceeding as above for females in the Hanford cohort we obtained:

1393 047

Basic Statistical Tests

Table 33

Females Only

$d = \bar{D}$ not cancer - $\bar{D}$ cancer

	Mean Life Dose	σ	n
Not Cancer	20.85	70.08	265
Cancer	38.31	157.63	117
All	26.2	105.03	382

The test statistic and the results of the test of the hypothesis is that the mean dose for females who died of cancer is equal to the mean dose for those who did not die of cancer are below:

- $d/\sigma_d = -1.497$
- $P(|d/\sigma_d| > 1.497) = .1344$
- $H_o : d = o$ vs. $H_1 : d \neq o$

Which again provides no definite evidence of a difference in the mean lifetime doses of the cancer vs. not cancer groups. However, it must be noted that the standard deviations in the lifetime dose for the cancer vs. the non-cancer groups are highly different (70.80 vs 157.63). The statistical test which was used above is quite sensitive to differences in the group standard deviations. The result which it gives cannot be accurately interpreted. Therefore, an additional test statistic was computed for this table. That statistic is the Welch-Alpin t-test. The statistic is computed as follows:

$$t = \frac{\bar{R}n - \bar{R}c}{\sqrt{\frac{\sigma n^2}{Rn} + \frac{\sigma c^2}{Kc}}}$$

1393 048

Basic Statistical Tests

For the data in question this statistic evaluates :

$$t = \frac{20.85 - 38.31}{\sqrt{\frac{(70.08)^2}{265} + \frac{(157.63)^2}{117}}} = \frac{-17.46}{15.19} = -1.15$$

That statistic is distributed as Student's distribution for degrees of freedom which depend on the standard deviation of the two groups. However, even for infinite degrees of freedom the value -1.15 will not allow us to reject the hypothesis that the group means are equal with the same degree of confidence that the above procedure allowed. We find, therefore, no reason to reject the hypothesis of equality.

To summarize the results of the above analysis: The total lifetime dose incurred by the cancer vs. the non-cancer groups of the population are not statistically different and we see no reason to claim that the simple lifetime dose is related to death by cancer. This is, of course, not to claim that the same would be true for all lifetime doses at all non-lethal levels only that it is true for the occupational levels encountered by the Hanford cohort.

The final basic question with which this section shall deal is: Does age at death differ for exposed vs. non-exposed populations? We performed two basic types of analyses to investigate the question.

First the mean age at death was computed for the exposed vs. the non-exposed groups with the following results:

Basic Statistical Tests

Table 34

$d = \bar{A}$ not exposed - $\bar{A}$ exposed

	Mean Age At Death	σ	n
Not Exposed	59.17	13.66	2354
Exposed	59.72	12.87	1638
All	59.50	13.20	3992

The test statistic and the results of the test of the hypothesis that the observed difference in the mean age at death for exposed vs. non-exposed population are shown below:

- $d/\sigma_{\bar{d}} = -1.29$
- $P(|d/\sigma_{\bar{d}}| > 1.295) = .1953$
- $H_o : d = o$ vs. $H_1 : d \neq o$

Again we have no reason to reject the hypothesis that the age at death is, on average, equal.

In addition, we correlated age at death with lifetime dose with the following results:

Table 35

	r	r^2	P(r)	n
All	-.0104	.00011	.37	3992
Male	.00729	.00005	.41	3610
Female	.0822	.00675	.05	382

We note that over all and in the male sub-population that there is no significant correlation between life span and lifetime dose. However, there is a slight positive correlation in women at the .05 level of confidence. This indicates that longer life spans are weakly associated with

1393 050

higher lifetime dosages. The source of this relationship is unclear at present. As further data accumulate from other sites the issue will, presumably, be resolved.

With one noteworthy exception, the analysis described in this section can be characterized as uninformative with regard to who will die of cancer.

To summarize the analytical results of this section we found one significant result and many non-significant results:

- Men who died at ages from 45-54 and were exposed died from cancer almost twice as often as similar men who were not exposed. There were 531 men in that age bracket and 20.7 percent of the exposed men died of cancer while 11.8 percent of the non-exposed men died of cancer.
- No significant relationship between exposure and cancer death in any other age bracket for men.
- No significant relationship between exposure and cancer death for women.
- No difference in the rate of dosage for subjects who died of cancer versus subjects who did not die of cancer. With rate of dosage being computed as:

  (lifetime dose) / (years exposed).
- No difference in the average age at death for exposed versus non-exposed populations.
- No correlation between lifetime dose and life span for males.
- A slight positive correlation between life span and total lifetime dose for females. It is to be emphasized that this correlation is very weak and that it is certainly not clear that higher lifetime doses cause longer life.

Combined Impact

This section describes the application of multivariate discriminant analysis to the prediction of which subjects in the cohort died of cancer. It does not attempt to develop a dose response for the population because of the multivariate nature of the predictive model. The presence of non-radiological variables in the model also makes the interpretation of any dose response relationship difficult and error prone.

The primary purpose of the section is to explore the relationship of certain of the independent variables as they act in concert to predict who in the cohort died of cancer. As was shown in the section on basic statistical tests, these variables taken one at a time have little power to predict who died of cancer except in one age bracket for the men of the cohort. However, it is sometimes the case that the more complete description provided by several variables will allow good predictions to be made even when those same variables, individually, do not. There are a variety of variables in the Hanford data which may act--both simply and jointly--in explaining variations in the risk of cancer death observed in the cohort. Since there are several relevant variables and some of these are continuous in nature the method of multi-way contingency tables which is often employed for multi-variate analysis is not practical. In this case it would result in more cells in the multi-way cross classification than there are cases in the Hanford data. We will therefore employ a discriminant analysis which will develop a linear combination of variables which will maximally separate the cancer from the non-cancer groups. We will then use this linear combination to classify the cases as cancer vs. non-cancer as cause of death. Several analyses of this type were performed in the attempt to identify those variables in the data that have the best ability to separate the cancer from the non-cancer deaths. The set of variables with which we began our discriminant analysis is

Combined Impact

comprised of:

- Total life time dose
- The peak exposure rate
- Years exposed
- The cancer death ratio for women
- The cancer death ratio for men
- The year at death
- Average rate of exposure.

Several of these variables were derived from the data provided by the commission. How each of these was derived is described below:

- The peak exposure rate - The cumulative lifetime dose is available at death, three years prior, five years prior, ten years prior, fifteen years prior, twenty years prior and twenty-five years prior
- to death. The average dose rate for each of the intervals defined by these cumulative doses was computed as: (incremental exposure in interval)/(years in interval). The maximum rate of the above set is taken as peak exposure rate.

- Years exposed - Years exposed is found by examining each of cumulative lifetime exposures at the above described points. For example, if the dose twenty-five years before death is non-zero then the years exposed variable is nominally defined as 25. If the twenty-five years prior to death exposure is zero and the twenty years prior to death exposure is non-zero then the years exposed variable is defined as 20.

Combined Impact

. The cancer death ratio for women - The cancer death ratio for women is computed as a function of age at death. For each of the age brackets:

  - 25-34
  - 35-44
  - 45-54
  - 55-64
  - 65-74
  - 75-84
  - 85-94
  - 95-104
  - 105-114

  The cancer death ratio (CDF) was computed as follows:

. CDRF= (# of female cancer deaths)/(# of female deaths)

. The cancer death ratio for men - The cancer death ratio for men was computed as above, except that male deaths were used.

. The average rate of exposure - The average rate of exposure was computed as: (lifetime dose)/(years exposed).

Though this method is well established in the field of cardiovascular epidemiology, as can be found in the references, there are certain hazards which must be guarded against in its use.

One problem which may cause difficulty is the form of the cancer death ratio as a function of age at death. If this ratio is increasing with age at death, and the total lifetime dose is increasing with age at death, increasing risk of death by cancer may be falsely attributed to dose when the better

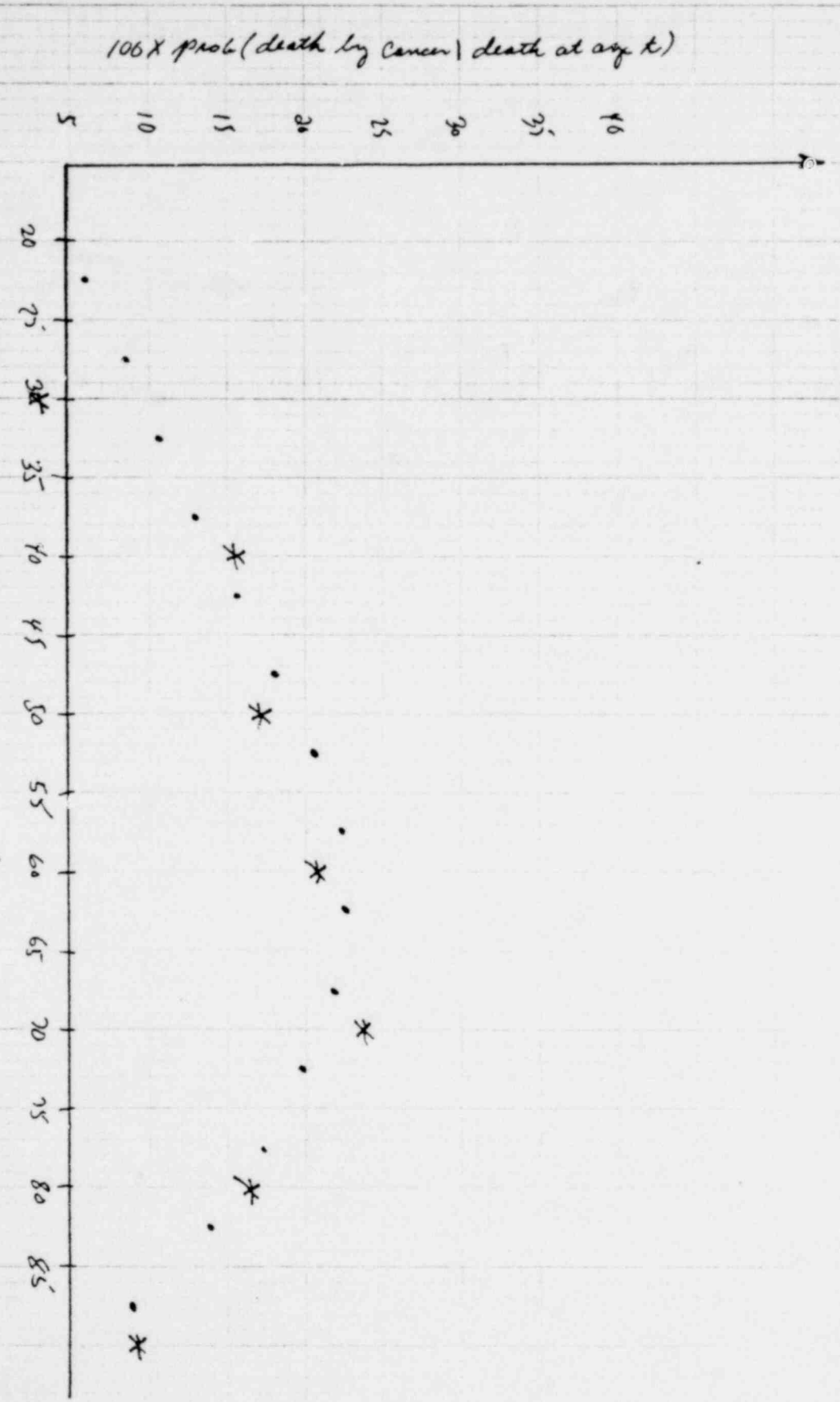

FIGURE 1

1393 055

Combined Impact

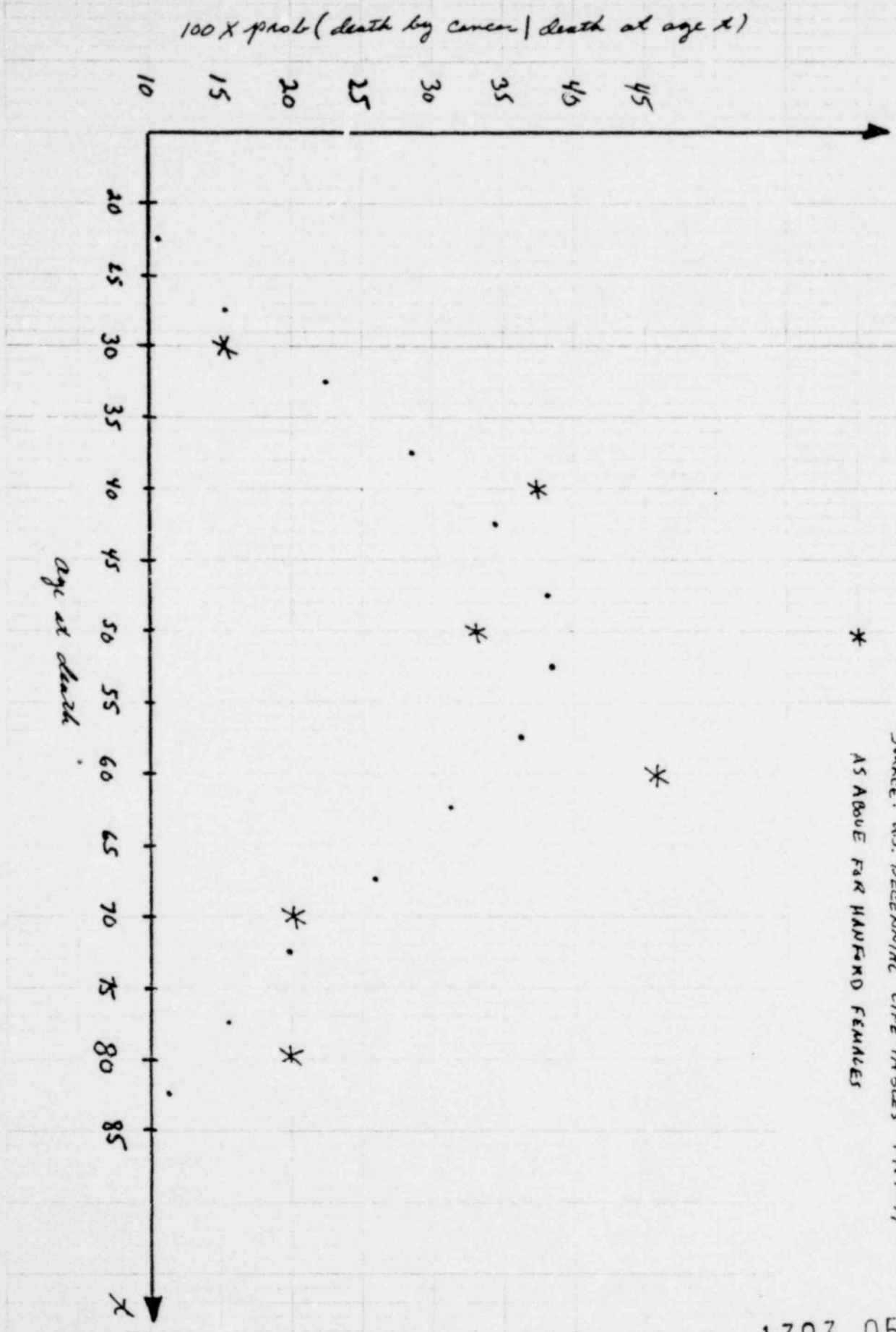

FIGURE 2

1393 056

explanation is age.

Another problem is that the individual variables may be highly inter-correlated leading to results which are difficult to interpret.

Regarding the first possible problem, the form of the cancer death ratio is of primary interest. The cancer death ratio for the male and female general populations of the United States was computed. The results are shown on the following pages. It is clearly observable that the cancer death ratio is not increasing with age at death after age 60 for men and after age 50 for women. The cancer death ratio which was derived for the men and women of the cohort is shown on the same axes. This implies that, even if lifetime dose and age at death are correlated, that confunding of the effects of increasing dosage and increasing age is not a problem of practical concern for this study.

To address the problem of the degree to which the independant variables in the discriminant model are correlated, a matrix of correlation coefficients was computed for the independent variables used in the discriminant models to be derived. The correlation matrix for males is given below.

TABLE 36

Correlation Matrix for Males*

	life dose	average rate	peak rate	years exposed	age at death	year of death
life dose	1	.854	.929	.257	-.034	.182
average rate		1	.907	.129	-.128	.082
peak rate			1	.239	-.07[illegible]	.158
years exposed				1	.24	.494
age at death					1	.3827
year of death						1

*All correlation coefficients significant at .05

Combined Impact

Some important observations on this correlation matrix are:

- Three of the radiological variables are strongly intercorrelated. These are lifetime dose and the two variables which were derived from it, so intercorrelation is not surprising.

- Age at death shows slight negative correlations with lifetime dose, average rate, and peak rate. This again points up the fact that age and age correlated factors will not be confounded with these three radiological variables.

- Age at death is correlated with years of exposure for the male subpopulation of the cohort.

- Year at death and age at death are correlated.

The correlation matrix for the independant variables used in predicting which females in the cohort died of cancer is shown on the accompaning Table 36. Some important observations about this correlation matrix are:

- There is no significant correlation between age at death and lifetime dose, average rate of exposure, peak rate of exposure, and years exposed.

- As with the males, the radiological variables are inter-correlated among one another.

- Years exposed is correlated with age at death

These observations will be recalled in order to clarify the interpretation of the discriminant models which have been contructed for this study.

1393 058

Combined Impact

TABLE 37

Correlation Matrix for Females

	life dose	average rate	peak rate	years exposed	age at death	year of death
life dose	1	.881*	.945*	.426*	.074	.09
average rate		1	.889*	.389*	.04	.039
peak rate			1	.423*	.064	.065
years exposed				1	.095	.221*
age at death					1	.095*
year of death						1

*Significant at the .05 level of confidence.

In many cases, such as the highly correlated radiological variables used here, the complete set of independent variables at hand contain redundant information about the difference between the two groups being investigated. In some cases the variables at hand may not be useful in discriminating the members of one group from the members of the other. Sequential selection procedures for variables to be used in discriminant models have been developed. In this analysis a generalized distance measure (V which was proposed by C.R. Rao) is used. The final discriminant model is constructed in a step-wise manner one variable at a time. First, the variable which produces the greatest distance between the groups is used to create a single variable prediction for group membership. Thereafter, the model is sequentially augmented by the variable from the full set which adds most to the distance between groups already attained with the previous

variables. Often a reduced set of variables can be found which is almost as good, or even better than the full set. When no variable from the full set can be found which increases the distance of the two groups from one another, the analysis is terminated. The interested reader is referred to the references listed under statistical methods in the bibliography. The test of significance for each of the variates as they are added to the model can be found in Cooley and Lohnes (1971 page 175).

The first discriminant analysis which was conducted was based on the male population and the following set of variates:

- Cancer death ratio for males (CDRM)
- Peak exposure rate
- Lifetime dose*
- Average rate of exposure
- Year at death
- Years exposed

The actions taken in the step-wise procedure are summarized on Table 37. Of the six available variables, three entered the model.

---

*Coded as 0=0, 1-99=1, 99+=2. The raw lifetime dosages never entered into the models generated by step-wise methods. Raw dosages in fact decreased the distance between the cancer and not-cancer groups.

1393 060

Combined Impact

TABLE 38

Step-wise Discriminant Analysis-Males

step	variable entered	Rao's V	ΔV	significance
1	CDRM	47.72	47.72	$p<.00009$
2	year at death	51.53	3.81	.05
3	life dose*	52.79	1.26	.26

*Coded as 0=0, 1-99=1, 99+=2. The raw lifetime dosages never entered into the models generated by stepwise methods. Raw dosages in fact decreased the distance between the cancer and not-cancer groups.

For the male subpopulation, in the aggregate, only one radiological variable enters and it is not associated with subsequent death by cancer at an even marginal level of significance. However, since significant variability from age bracket to age bracket was observed for the association of simple exposure with cancer death using the Mantel-Haenzel procedure in the previous section an age stratified analysis was undertaken here as well. For each of the age brackets previously described, a step-wise discriminant analysis was performed. The findings for each age bracket were:

- 25-34 - Only two cases of cancer caused death were found so no significiant findings were possible.
- 35-44 - No significant predictive variables were found.
- 45-54 - Years exposed is correlated (canonically) at the .001 level of confidence with subsequent death by cancer. The average rate of exposure enters the model but is significant only at .09.

1393 061

Combined Impact

- 55-64 - No variables were significantly associated with subsequent cancer death.
- 65-74 - No variables were significantly associated with subsequent cancer death.
- 75-84 - The average rate of exposure was associated with subsequent cancer death with p-value .0001, and year of death at .035.
- 85-94 - No radiological variables were significantly associated with subsequent cancer death.

For each of the two age brackets in which significant findings were uncovered, the step-wise analysis is presented in a table. First, for the 45-54 age bracket we have:

TABLE 39

Step-wise Discriminant Analysis
Males ages 45-54 at death

step	variable entered	Rao's V	ΔV	significance
1	years exposed	10.55	10.55	.001
2	life dose	13.45	2.89	.089

Canonical Correlation: .157, significance: .001

1393 062

Combined Impact

The power of this model to predict who in this age bracket died from cancer is shown in the following four-fold table:

Actual	Predicted	
	not-cancer	cancer
not-cancer	282	157
cancer	46	46

From which it can be seen that the model correctly predicts 328 of 531 or 61.77 percent of the cases.

As was noted earlier, the years exposed variable is correlated with age at death, but in a single age bracket this is of no great concern. It is to be noted that increasing periods of exposure point towards increasing risk of cancer death.

Turning now to the other age bracket in which significant results were focused we have the following:

TABLE 40

Step-wise Discriminant Analysis
Males ages 75-85 at death

step	variable entered	Rao's V	ΔV	significance
1	average rate	15.68	15.68	.0001
2	year of death	20.11	4.4	.0353

Canonical Correlation: .17, significance: .0001

1393 063

Combined Impact

The predictive power of this model is expressed in the following table:

Actual	Predicted	
	not-cancer	cancer
not-cancer	404	160
cancer	68	45

From which it can be seen that the model correctly predicts 449 out of 677 or 66.3 percent of the cases correctly with high average rates pointing in the direction of increasing cancer risk.

The method of step-wise discriminant analysis was applied to the women of the cohort. The variables available for inclusion into the model are the same as were available for the males except for the cancer death ratio (CDRF) which was computed specifically for the female sub-sample. The lifetime dose was again coded as was that for the males. When the raw lifetime dose scores were made available to the step-wise discriminant procedure the variable was not included in the model because it decreased the inter-group separation rather than increasing it - even trivially. The table describing the course of the stepwise discriminant analysis for the female subpopulation is shown on the following page (Table 39). Two of the radiological variables are associated with subsequent cancer death at a p-value of .06. This is not quite significant to a classical degree of one chance in twenty. They are significant to one chance in 16.67. This is obviously a borderline value. It is to be observed, however, that this degree of association arises from

1393 064

TABLE 41

Step-wise Discriminant Analysis-Females

step	variable entered	Rao's V	ΔV	significance
1	CDRF	25.68	25.68	p<.00009
2	peak rate	29.09	3.41	.06
3	years exposed	32.52	3.43	.06
4	average rate	33.86	1.34	.24
5	life dose	36.36	2.5	.11

Canonical Correlation: .295, significance: p<.00009

the rather small set of cases available in the cohort which were female. This result must be called ambiguous at this time. As more data become available the question should be reinvestigated. The addition of a few hundred cases may suffice to settle the issue.

1393 065

BIBLIOGRAPHY

The Mantel-Haenzel Procedure

J.L. Fleiss, Statistical Methods for Rates and Proportions. John Wiley and Sons, New York, 1973.

N. Mantel, W. Haenzel: "Statistical Aspects of the Analysis of Data From Retrospective Studies", Journal of NCI, Vol. 22, No. 4, April 1959, pp. 719-748.

N. Mantel, "Chi-Square Tests with One Degree of Freedom; Extentions of the Mantel-Haenzel Procedure", Journal of Am. Statistical Assoc., Sept. 1963, Vol. 58, pp. 690-700.

The Method of Proportional Mortality

S. Wagner, N. Mantel: "Breast Cancer at a Mental Hospital Before and After the Intorduction of Neuroleptic Agents", Cancer Research, Vol. 38, Sept. 1978, pp. 2703-2708.

Discriminant Analysis and Canonical Correlation

J. Truett, J. Cornfield, W. Kannel, "A Multivariate Analysis of the Risk of Coronary Heart Disease in Framingham", Journal of Chronic Diseases, 1967, Vol. 20, pp. 511-524.

W. Cooley, P. Lahnes, Multivariate Data Analysis. John Wiley and Sons, New York, 1971.

N. Nie, et al., Statistical Package for the Social Sciences. McGraw-Hill.

Yu-Chi Ho, A.K. Agrowala, "On Pattern Classification Algorithms Introduction and Survey", Machine Recognition of Patterns. IEEE Press, A.K. Agrowala, ed., 1977, pp. 247-260.

Radiation Biology

Casarett, Alison P.,Radiation Biology. Prentice Hall, New Jersey, 1968.

Mole, R.H.,"Ionizing Radiation as a Carcinogen: Practical Questions and Academic Persuits", British Journal of Radiology, Vol. 48, pp. 157-169.

General

U.S. Decennial Life Tables for 1969-1971. Vol. 1, No. 5, DHEW Publication No. HRA 75-1150.