A Reanalysis of the Atomic Bomb Survivor's Cancer Rates Using a Monte Carlo Simulation

By
Nenad Ilic

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Physics
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

A Reanalysis of the Atomic Bomb Survivor's Cancer Rates Using a Monte Carlo Simulation

BY

Nenad Ilic

A Practicum submitted to the Faculty of Graduate Studies of The University

of Manitoba in partial fulfillment of the requirements of the degree

of

MASTER OF SCIENCE

Abstract

It is known that at high doses ionizing radiation can cause cancer or leukemia. The functional relationship between cancer (leukemia) induction and received dose of ionizing radiation is still unknown, particuiarly in a low dose region.

In this thesis atomic bomb survivors data are used to test two models, a linear threshold model for solid cancers and leukemia data and a liner-quadratic model for leukemia data only.

Atomic bomb survivors data used in this thesis include data for stomach, lung, all solid cancers (all cancers excluding leukemia), and leukemia. Cancer and leukemia mortality rates and excess mortality rates are investigated as function of received dose using the standard Chi-square and a non-standard Monte Carlo simulation method.

Using empirical data points one thousand simulated data sets were generated. Each simulated data set was fitted with a straight line, and intercept to dose axis, threshold, was calculated. This procedure gives one thousand threshold values. Statistical analysis of threshold values is used as a test of linear no-threshold and threshold models. In addition to a linear fit, a linear-quadratic fit was performed for leukemia data. In order to test a hormesis hypothesis Zero equivalent points (*ZEP*) have been calculated.

Upper threshold limits obtained by Monte Carlo simulation are 0.037 Sv and 0.061 Sv for all solid cancers, and 0.154 and 0.193 Sv for leukemia data sets. Investigation of mortality rates shows that the threshold and quadratic models do not fit data significantly better than the linear model.

Acknowledgments

The author wishes to express his appreciation to Prof. J. V. Jovanovich for his encouragement and advice during the course of this study. The author is also very grateful to Dr. Ken Mount from The Statistics Department for the time spent reading early drafts of the thesis, making necessary corrections, and providing valuable advice. Without their help this work would have never been completed.

List of contents

List of Tables

List of Figures

Introduction

Ionizing radiation is one of the most common phenomena in nature. Ionizing radiation is produced in transformation of radionuclides residing in environment as well as in biological systems. In addition to radiation produced by radio-nuclides various particles and photons from cosmic rays are also present. Specifically, neutrons, charged particles, gamma and x, rays have been coming continuously from the outer space. The whole evolution of living organisms had occurred in an environment filled by radiation. At present, a variety of human-made sources of radiation are present, besides natural radiation.

Depending on the amount of energy imparted to biological tissue it is common to divide received doses of ionizing radiation into low, intermediate, and high dose range. Low doses are usually defined as the region below 0.2 Sv [Brill 82, UNSCEAR 94, Schi 95], intermediate as between 0.2 Sv and 2.5 Sv, and high dose range is above 2.5 Sv [Brill 82].

Ionizing radiation acts on genetic material of all living organisms. In doing so radiation might have had an important role in the evolution of species. A knowledge of effects of low doses of ionizing radiation is very important, because live organisms are exposed mainly to low doses delivered at low rates. The effects of these low doses on live organisms are still insufficiently understood.

Most of our knowledge of radiation effects on human health are derived from data from explosions of atomic bombs over Hiroshima and Nagasaki, accidents in nuclear industry, follow up of uranium miners and follow up of children and adults who have

received high doses for therapeutic purposes [Shap 90].

Additional knowledge about the low dose influence on human health has been obtained through studies of influence of radon concentration in air on death excess due to lung cancers, investigation of mortality in regions with high levels of natural exposure and studies of professionally exposed persons in the nuclear industry.

Two main biological effects of ionizing radiation are genetic mutations and induction of cancers. A linear no-threshold hypothesis is generally assumed for induction of all solid cancers in the low dose region. This means, that even the smallest exposures received by someone causes risk of cancer developing.

The linear no-threshold model has implication for regulation of ionizing radiation protection. Over the past few decades protective measures have been getting stricter and stricter. As a result, levels of released radiation in the environment have been repeatedly lowered [Ncrp 93]. The use of strict protective measures required building expensive protective barriers surrounding sources of radiation, applying sophisticated procedures in using, processing, transporting, storing and disposing of radioactive substances, all costing significant amounts of money.

Validity of the linear no-threshold model in the low dose region has been contradicted by some relatively recent environmental and biological studies. Some experimental evidence showed that relatively low doses of ionizing radiation can produce adaptive response that stimulates repair mechanisms of cells [Okam 92, UNSCEAR 94]. Studies of populations that live in regions with high levels of natural radiation did not find an increase in carcinoma mortality, when compared with regions with low levels

[UNSCEAR 94]. Also, studies of incidence and mortality due to radon-induced lung cancers did not find any significant difference between areas with high and low concentration of radon [UNSCEAR 94, Losal 95]. Some authors have found a negative correlation between lung cancer mortality and concentration of radon in dwellings [Cohe 97]. These recent findings indicate that low levels of ionizing radiation may, in fact, be beneficial. This beneficial effect is called hormesis [Lucke 91, Lucke 92, Kondo 93]. Clearly, the hormesis model is incompatible with the linear no-threshold model in the low level region of ionizing radiation.

The investigation of cancer induction includes epidemiological, and biological studies. An epidemiological study investigates connection between some variable of interest and cancer incidence observed in a group of people. Variable of interest can be a chemical or physical agent, heredity, social status of observed people, or some other parameters. The study of atomic bomb survivors is an example of epidemiological study. Epidemiological methodology of assessing carcinoma risk is reviwed in [Bres 80, Bres 87, Este 94]. A biological study investigates direct biological effects of ionizing radiation on cell (or tissue) [Kondo 93, Heid 97].

This thesis studies a possible existence of a threshold in the linear model for cancer and leukemia induction, and possible existence of hormesis effect for leukemia. It is done by applying a standard least-squares fit (Chi-square analysis) and performing a Monte Carlo simulation on the data of the atomic bomb survivors. Statistically significant results of ionizing radiation effects in the low dose region are very difficult to obtain because that effect is very small and conclusive research would require large number of subjects to

observe. Because of high relative uncertainties, standard formulae for calculation of estimator uncertainties can not be used. In this work, we used a Monte Carlo simulation that started from the empirical data sets and have generated, using appropriate computer programs, many artificial data sets. Each generated data set represents one artificial experiment. Statistical analysis of many (in our case one thousand) artificial experiments gives opportunity to obtain values for estimators and their uncertainties in a more reliable manner than using standard formulae.

In this thesis, in agreement with common statistical terminology [Shesk 97, p.6, Neter 90, chapter 1], Greek letters correspond to parameters of a model which describes parent population. When a specific sample is described, values that correspond to parameters are called estimators. Specific estimator values are estimates. Estimators and estimates are labeled with Latin letters. For example, if we assume that a relationship between variables y and x in a parent population is linear, we write it in the form $E[y] = \alpha + \beta \cdot x$. Values α and β are parameters of model. When a specific sample is described, the linear model has a form $\hat{y} = a + b \cdot x$. Values a and b are estimators of the sample. The $\hat{y}$ (hat) refers to a y value on the fitted line, y refers to a data point.

Chapter I

Induction of cancers by ionizing radiation

1.1 Cancer mortality rates

Cancer induction due to ionizing radiation is investigated through two kinds of relationships. One is the relationship between cancer incidence and received dose of ionizing radiation, and the other is the relationship between cancer mortality and the dose. The investigation of the incidence has, in the case of cancers that have higher survival rates (skin cancer excluding melanoma, thyroid cancers etc.), advantages in comparison to the investigation of cancer mortality. The investigation of mortality for cancers that have higher survival rates, can lead to the wrong conclusion that the impact of ionizing radiation to develop cancer is smaller than it actually is. This is not a problem for cancers with lower survival rates (lung, liver carcinoma) because the incidence and mortality data are very close.

Let us look at a group of persons exposed to a certain dose of ionizing radiation. A control group is a group of unexposed persons. Denote the observed number of deaths in the exposed group with O, the number of persons in the exposed group with N, the observed number of deaths in the control group with O_o, and the number of persons in the control group with No [Losal 95 p. 99].

The cancer mortality rate MR, for the observed group is defined by

$$MR = \frac{O}{N} \tag{1.1}$$

and for the control group is

$$MR_0 = \frac{O_0}{N_0}. \tag{1.2}$$

The excess mortality rate is defined as

$$Y = \frac{O}{N} - \frac{O_0}{N_0}. \tag{1.3}$$

1.2 Model forms

A proposed relationship between the effect of ionizing radiation on human health ($Y(d)$) and received dose (d) is defined as expected value of $Y(d)$;

$$E[Y(d)] = \alpha + \beta \cdot d + \gamma \cdot d^2 \tag{1.4}$$

in low and medium dose ranges [Brill 82].

In this thesis the effect of ionizing radiation is the excess cancer mortality rate Y. This proposed functional form includes a linear term $\alpha + \beta \cdot d$, and quadratic term $\gamma \cdot d^2$.

A linear function can describe two models. One is the linear no-threshold model ($\alpha = 0$). The other is the linear threshold model ($\alpha \neq 0$).

The linear no-threshold model has the form:

$$E[Y(d)] = \beta \cdot d. \tag{1.5}$$

In this model even the smallest dose received increases risk of developing cancer. Most authors have commonly used the linear no-threshold model for cancer induction [Beir V 90, UNSCEAR 94, Epa 94] in low and medium dose ranges.

Some authors have assumed that the dose response function for a low dose region can have different forms than the form obtained by the linear no-threshold model. One of those forms assumes the existence of a threshold dose below which radiation has no effect on human health [Heid 97, Hoel 98]. Another assumes the existence of hormesis, namely that radiation has beneficial effect on human health below some dose. Other possible forms of dose response functions below 0.2 Sv are qualitatively represented in figure 1.1

The threshold model has the form:

$$E[Y(d)] = \begin{cases} 0 & d \le T \\ \alpha + \beta \cdot d, \quad \beta > 0, & d > T \end{cases}, \tag{1.6}$$

where T is the threshold dose.

The hormesis effect in its simplest form can be described by a linear-quadratic function of the form:

$$E[Y(d)] = \beta \cdot d + \gamma \cdot d^2 \tag{1.7}$$

with parameter β negative, and γ positive. The constant term α in equation 1.7 is omitted because at zero dose there is no effect due to ionizing radiation. If $\beta \neq 0$, the linear-quadratic function 1.7 has another intercept with the dose axis.

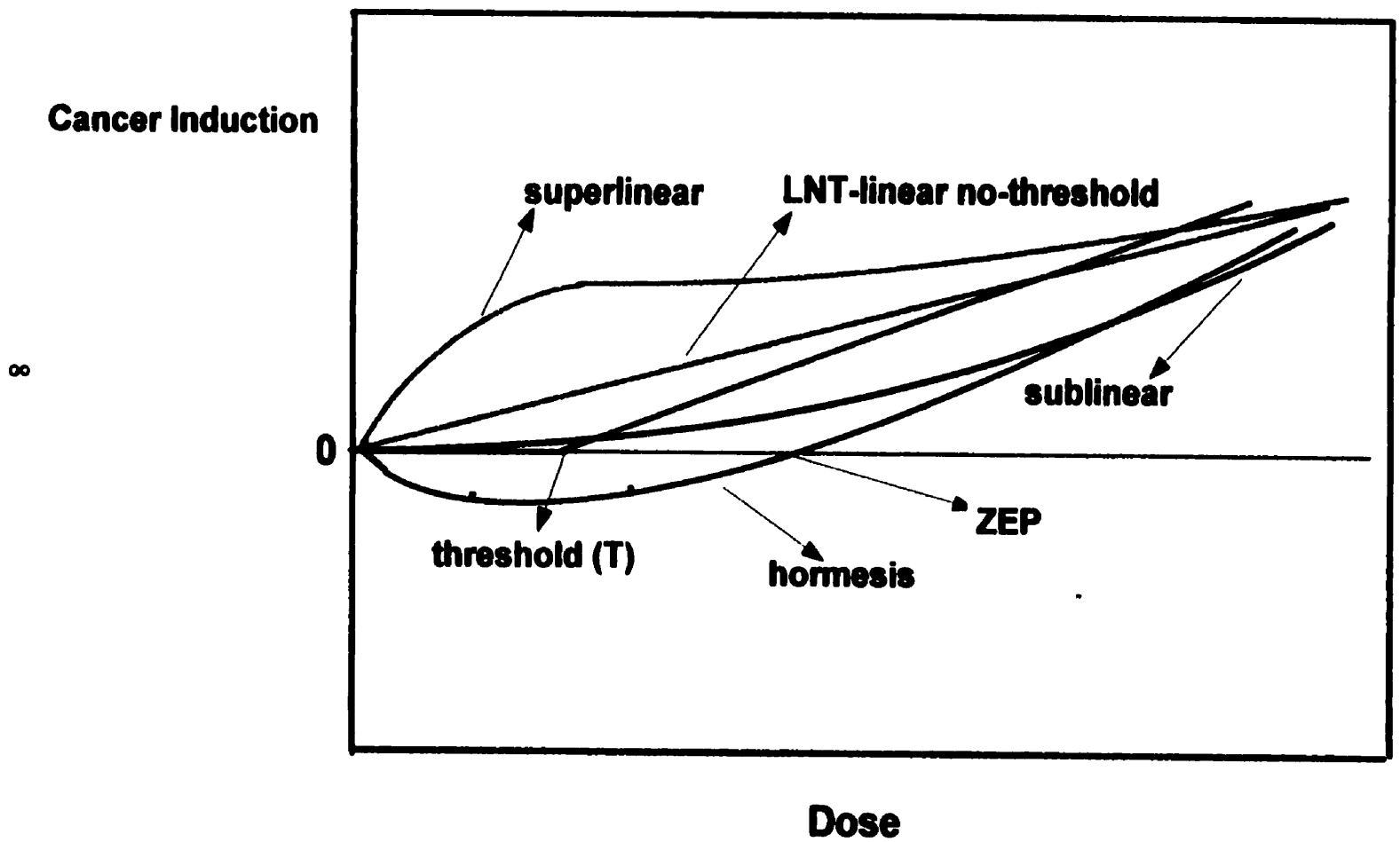

Figure 1.1 Possible shapes for dose-response function in low dose region.

If this intercept is on the positive side of dose axis it is called the zero equivalent point (*ZEP*). At this point ionizing radiation has no effect on human health. Below the *ZEP* value the effect is beneficial (hormesis), above it radiation is harmful. This beneficial effect might be related to the organism's adaptive response caused by ionizing radiation [Kondo 93, Cohe 97].

Depending on signs of parameters β and γ, the linear-quadratic function can describe other models presented on table 1.1. Graphical presentations of these models is shown in Figure 1.2.

Table 1.1 Special cases of the linear-quadratic model. Three combinations of β and γ values are of no physical interest and are labeled "Not of interest".

	Parameter β	Parameter γ	Model
1	$\beta>0$	$\gamma>0$	Linear-quadratic, (no-threshold)
2	$\beta=0$	$\gamma>0$	Pure qudratic
3	$\beta<0$	$\gamma>0$	Hormesis
4	$\beta>0$	$\gamma<0$	Not of interest
5	$\beta=0$	$\gamma<0$	Not of interest
6	$\beta<0$	$\gamma<0$	Not of interest

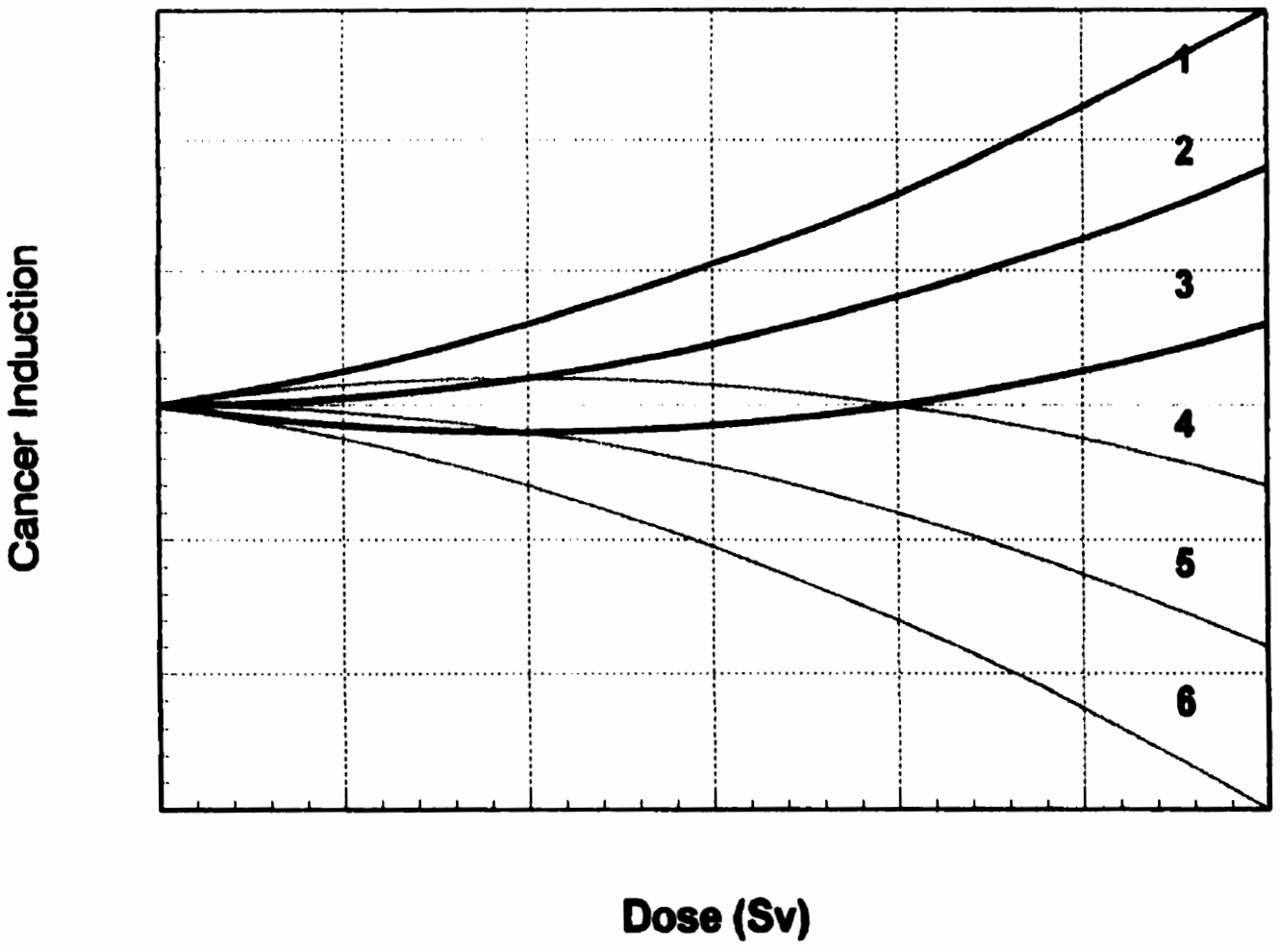

Figure 1.2 Forms of dependence of excess mortality rates versus received dose which can be described by a linear quadratic function. Meaning of numbers attached to curves is explained in Table 1.1.

According to present published evidence, there is no clear answer as to which model of dose response function is most appropriate for a description in the low dose region. This uncertainty in the shape of dose response function is due to weak impact of ionizing radiation on the excess carcinoma or leukemia.

Chapter II

Data sets used in analysis

The analysis in this thesis was done using data from two studies of mortality of atomic bomb survivors (ABS). The first study is:

D. Perce, Y. Shimizu, D. Preston, M. Vaeth, and K. Mabuchi;

"*Studies of the Mortality of Atomic Bomb Survivors. Report 12, Part I Cancer:1950-1990*", Radiation Research 146, 1-27 (1996).

This study ([Pier 96]) was chosen because it has the most recent set of data for solid cancers, and leukemia. It covers the period between1950 and 1990. The solid cancer data set was taken from Table II on page 5, which we designate here "[Pier 96] solid cancers" (reproduced in Table 2.1), the leukemia data set was taken from Table V on page 7 and was designated "[Pier 96] leukemia" (reproduced in Table 2.2).

The other study is: Y. Shimizu, H. Kato, W. Schull, K. Mabuchi;

"*Dose-response analysis among atomic-bomb survivors exposed to low-level radiation*", published in " Low dose irradiation and biological defense mechanisms" Elsevier Science Publishers B.V., 1992. This study ([Shim 92]) is interesting because it has more data points in a the dose region below 0.5 Sv. This set of data reports on cancer rates for the period between 1950 and 1985. Data sets were taken from the original table on page 72 of the original paper. This paper includes data for the stomach (named "[Shim 92] stomach"), the lung (named "[Shim 92] lung"), all solid cancers ("[Shim 92] all solid"), and leukemia ("[Shim 92] leukemia"). The [Shim 92] stomach, lung, all solid cancer data sets are reproduced in Table 2.3 in this thesis, the [Shim 92] leukemia data set is in Table 2.4.

The words "dose received" in this thesis refer to the dose equivalent in Siverts (Sv).

The first column in tables 2.1-2.4 "Observed group j" labels seven different groups of the ABS data. All individuals in a group are assumed to have received the same mean dose for that group.

The group in the lowest dose region was taken as the control group. For [Pier 96] the control group (i.e. background cancer mortality rates) is a group with received doses below 0.005 Sv. For [Shim 92] the control group is a group that received doses in interval 0.010-0.019 Sv. The control groups are labeled $j = 0$. The highest observed groups labeled $j = 6$ and are not considered in this thesis because of uncertainties in finding mean dose.

In all four tables column "Dose range (Sv)" lists dose ranges received by observed groups of survivors. Column "Number of subjects N_{d_j}" contains the number of people in each dose range group and column "Number of observed deaths O_{d_j}" contains the number of observed deaths due to a particular cancer category in each dose group.

The values in column "Mean dose, d_j (Sv)" were taken as mid-points of dose ranges in column "Dose range (Sv)". Dose uncertainties were assumed to be standard deviations equal to 25% of the width of the corresponding dose range. This value may be an overestimate but it was so chosen to put an upper limit on the effect of dose uncertainties (see chapter 6).

"Excess Cancer Mortality rate Y_j" is computed using equations 1.1-1.3. Specifically, for a group j

$$Y_j = \frac{O_{d_j}}{N_{d_j}} - \frac{O_{d_0}}{N_{d_0}}, \tag{2.1}$$

where

O_{d_j} is the number of observed deaths due to cancer in j -dose group,

N_{d_j} is the number of persons in j - dose group,

O_{d_0} is the number of observed deaths in the control group, and

N_{d_0} is the number of persons in the control group.

It is important to note that the subtraction in 2.1 introduces correlation among the excess mortality rates because the same value (the control mortality rate) is subtracted from each one. This subtraction of background rate is quite common practice in nuclear physics experiments because correlation introduced this way has usually negligible effect on final result.

Following the practice of [Heid 97], [Losal 95 p. 107], [Winkl 75 p. 227] uncertainties in the observed number of deaths are assumed to be Poisson distributed, thus the standard deviation in mortality rate is equal to

$$s_{MR_j} = \frac{\sqrt{O_{d_j}}}{N_{d_j}}, \tag{2.2}$$

and the standard deviation in excess mortality rates is equal to

$$s_{Y_j} = \sqrt{\frac{O_{d_j}}{N_{d_j}^2} + \frac{O_{d_0}}{N_{d_0}^2}} \,. \tag{2.3}$$

To illustrate the procedure used we shall look at an example in Table 2.2 [Pier 96] leukemia data. The dose group labeled $j = 1$ has a range of received doses between 0.005 Sv and 0.1 Sv. The mean received dose is: d_1 = (0.005+0.1)/2=0.05 Sv. In order to illustrate effect of dose uncertainties, the standard deviation of the mean dose was taken as 0.25 x (0.1-0.005) = 0.024 Sv. The number of subjects in this group was N_{d_1} =32915 persons and the number of observed deaths due to leukemia was O_{d_1} = 59 deaths. The control group has N_{d_0} = 35458 persons and O_{d_0} = 73 observed deaths. The excess leukemia mortality rate according to equation 2.1 is $Y_1 = -0.27 \cdot 10^{-3}$ *deaths / person* (column six, Table 2.2). Uncertainty for the excess leukemia mortality rate according to equation 2.3 is $s_{Y_1} = 0.33 \cdot 10^{-3}$ *deaths / person* (column six, Table 2.2).

Table 2.1 Number of observed deaths for solid cancers, and the excess in cancer mortality rate. Columns 2,4 and 5 were taken from [Pier 96, Table II, page 5] and columns 3 and 6 were calculated as explained in the text.

1	2	3	4	5	6
Obs. group j	Dose range (Sv)	**Mean Dose** d_j **(Sv)**	Num. of Subjects N_{d_j}	Num. of Obs. Deaths O_{d_j}	**Excess Cancer Mortality Rate** (Y_j) $(10^{-3}\frac{deaths}{person})$
0	0 (<0.005)	**0.0025**	36459	3013	**0**
1	0.005-0.1	**0.05 ± 0.02**	32849	2795	**2.44 ± 2.20**
2	0.1-0.2	**0.15 ± 0.02**	5467	504	**9.55 ± 4.37**
3	0.2-0.5	**0.35 ± 0.07**	6308	632	**17.5 ± 4.3**
4	0.5-1.0	**0.75 ± 0.12**	3202	336	**22.3 ± 5.9**
5	1.0-2.0	**1.5 ± 0.25**	1608	215	**51.1 ± 9.2**
6	>2.0		679	83	**39.6 ± 13.5**
	Total		86572	7578	

Table 2.2 Number of observed deaths for leukemia, and the excess in leukemia mortality rate. Columns 2,4 and 5 were taken from [Pier 96, Table V, page 7] and columns 3 and 6 were calculated as explained in the text.

1	2	3	4	5	6
Obs. group j	Dose range (Sv)	**Mean Dose (Sv)** d_j	Num. of Subjects N_{d_j}	Num. of Obs. Deaths O_{d_j}	**Excess Leukemia Mortality Rate** (Y_j) $(10^{-3}\frac{deaths}{person})$
0	0 (<0.005)	**0.0025**	35458	73	**0**
1	0.005-0.1	**0.05 ± 0.02**	32915	59	**-0.27 ± 0.33**
2	0.1-0.2	**0.15 ± 0.02**	5613	11	**-0.10 ± 0.64**
3	0.2-0.5	**0.35 ± 0.07**	6342	27	**2.20 ± 0.85**
4	0.5-1.0	**0.75 ± 0.12**	3425	23	**4.66 ± 1.42**
5	1.0-2.0	**1.5 ± 0.25**	1914	26	**11.5 ± 2.7**
6	>2.0		905	30	**31.1 ± 6.1**
	Total		86572	249	

Table 2.3 The number of observed deaths for stomach, lung and all solid cancers, and corresponding excess mortality rates (Y_j). Columns 2,4 and 5 were taken from [Shim 92, page 72] and columns 3 and 6 were calculated as explained in the text.

1	2	3	4	5			6		
j	Dose range (Sv)	**Mean Dose (Sv)** d_j	Num. of Subjects N_{d_j}	Numb. of observed deaths (O_{d_j}) due to:			Y_j $(10^{-3}\frac{deaths}{person})$		
				Stomach cancer	Lung cancer	All cancers	**Stomach**	**Lung**	**Total solid**
0	0		45148	1153	338	3246	**0**	**0**	**0**
1	0.010-0.019	**0.014 ± 0.002**	7430	175	55	498	**-1.98 ± 1.93**	**-0.08 ± 1.08**	**-4.87 ± 3.26**
2	0.020-0.049	**0.034 ± 0.007**	9235	254	87	717	**1.96 ± 1.88**	**1.93 ± 1.09**	**5.74 ± 3.16**
3	0.050-0.099	**0.074 ± 0.012**	6439	168	56	516	**0.55 ± 2.15**	**1.21 ± 1.23**	**8.24 ± 3.75**
4	0.100-0.199	**0.150 ± 0.025**	5316	147	44	400	**2.11 ± 2.40**	**0.79 ± 1.31**	**3.35 ± 3.97**
5	0.200-0.499	**0.350 ± 0.075**	6271	187	65	533	**4.28 ± 2.31**	**2.88 ± 1.35**	**13.1 ± 3.9**
6	>0.500		6681	194	82	573	**3.50 ± 2.22**	**4.79 ± 1.41**	**13.9 ± 3.8**
	Total		86520	2278	727	6501			

Table 2.4 Number of observed deaths for leukemia, and corresponding excess leukemia rates. Columns 2,4 and 5 were taken from [Shim 92, page 72] and columns 3 and 6 were calculated as explained in the text.

1	2	3	4	5	6
Obs. group j	Dose range (Sv)	**Mean Dose (Sv)** d_j	Num. of Subjects N_{d_j}	Num. of Obs. Deaths O_{d_j}	**Excess Leukemia Mortality Rate** (Y_j) ($10^{-3}\frac{deaths}{person}$)
0	0		45148	81	**0**
1	0.010-0.019	**0.014 ± 0.002**	7430	11	**-0.31 ± 0.49**
2	0.020-0.049	**0.034 ± 0.007**	9235	14	**-0.28 ± 0.45**
3	0.050-0.099	**0.074 ± 0.012**	6439	8	**-0.55 ± 0.48**
4	0.100-0.199	**0.150 ± 0.025**	5316	11	**0.27 ± 0.65**
5	0.200-0.499	**0.350 ± 0.075**	6271	21	**1.55 ± 0.76**
6	>0.500		6681	75	**9.43 ± 1.31**
	Total		86520	211	

Chapter III

Statistical methods used in data analysis

In this study two models were used; linear, and linear-quadratic. The method of the least-squares was applied to both models. The estimators for both models and their uncertainties were computed. Then the confidence intervals were computed for the fitted estimators by analyzing Chi-square curves. Next the statistical analysis of the results obtained by the Monte Carlo simulation was performed. In this chapter dose uncertainties are not taken in account.

3.1 Least-Squares fit

3.1.1 Fit by a linear function

Set (d_j, Y_j) of the data points can be fitted with a straight line by making standard weighted least-squares fit [Bevi 92, p. 103]:

$$\hat{Y}(d) = a + b \cdot d \quad . \tag{3.1}$$

The Chi-square function for the linear fit is defined as

$$X^2 = \sum_{j=1}^{n} \left(\frac{Y_j - a - b \cdot d_j}{s_j} \right)^2 , \tag{3.2}$$

where

Y_j are data points (excess mortality rates) to fit,

d_j are dose values,

a, b are estimators of linear fit,

s_i are uncertainties in Y_j ,

n is number of data points to be fitted.

Estimators of the fit, intercept a and slope b, are obtained by using formulas from [Bevi 92, p. 104, eqn. 6.12]. These equations in our notation are:

$$a = \frac{1}{\Delta} \cdot \left(\sum_{j=1}^{n} \frac{d_j^2}{s_j^2} \cdot \sum_{j=1}^{n} \frac{Y_j}{s_j^2} - \sum_{j=1}^{n} \frac{d_j}{s_j^2} \cdot \sum_{j=1}^{n} \frac{d_j \cdot Y_j}{s_j^2} \right), \tag{3.3}$$

and

$$b = \frac{1}{\Delta} \cdot \left(\sum_{j=1}^{n} \frac{1}{s_j^2} \cdot \sum_{j=1}^{n} \frac{d_j \cdot Y_j}{s_j^2} - \sum_{j=1}^{n} \frac{d_j}{s_j^2} \cdot \sum_{j=1}^{n} \frac{Y_j}{s_j^2} \right), \tag{3.4}$$

where

$$\Delta = \sum_{j=1}^{n} \frac{1}{s_j^2} \cdot \sum_{j=1}^{n} \frac{d_j^2}{s_j^2} - \left(\sum_{j=1}^{n} \frac{d_j}{s_j^2} \right)^2 . \tag{3.5}$$

3.1.2 Fit by a linear-quadratic function

As discussed in section 1.2, the linear-quadratic model has the form

$$\hat{Y}(d) = b \cdot d + c \cdot d^2 . \tag{3.6}$$

The Chi-square function for linear-quadratic fit is

$$X^2 = \sum_{j=1}^{n} \left(\frac{Y_j - b \cdot d_j - c \cdot d_j^2}{s_j} \right)^2 . \qquad (3.7)$$

Estimators b and c were obtained by minimizing the Chi-square function 3.7. Equations 3.8-3.10 were derived using formulas for least-squares fit to a polynomial which is given in chapter 7 of [Bevi 92, p115]. These estimators when applied to eq. 3.6 are

$$b = \frac{1}{\Delta_1} \cdot \left(\sum_{j=1}^{n} \frac{d_j^4}{s_j^2} \cdot \sum_{j=1}^{n} \frac{d_j \cdot Y_j}{s_j^2} - \sum_{j=1}^{n} \frac{d_j^3}{s_j^2} \cdot \sum_{j=1}^{n} \frac{d_j^2 \cdot Y_j}{s_j^2} \right), \qquad (3.8)$$

$$c = \frac{1}{\Delta_1} \cdot \left(\sum_{j=1}^{n} \frac{d_j^2}{s_j^2} \cdot \sum_{j=1}^{n} \frac{d_j^2 \cdot Y_j}{s_j^2} - \sum_{j=1}^{n} \frac{d_j^3}{s_j^2} \cdot \sum_{j=1}^{n} \frac{d_j \cdot Y_j}{s_j^2} \right), \qquad (3.9)$$

where

$$\Delta_1 = \sum_{j=1}^{n} \frac{d_j^2}{s_j^2} \cdot \sum_{j=1}^{n} \frac{d_j^4}{s_j^2} - \sum_{j=1}^{n} \frac{d_j^3}{s_j^2} \cdot \sum_{j=1}^{n} \frac{d_j^3}{s_j^2} . \qquad (3.10)$$

3.1.3 Estimation of errors

Uncertainties of estimators can be determined by calculating error matrices [Bevi 92, p. 123]. The error matrix for a linear fit is

$$E = \begin{bmatrix} \sum_{j=1}^{n} s_j^2 \cdot \left(\frac{\partial a}{\partial Y_j}\right)^2 & \sum_{j=1}^{n} s_j^2 \frac{\partial a}{\partial Y_j} \cdot \frac{\partial b}{\partial Y_j} \\ \sum_{j=1}^{n} s_j^2 \frac{\partial a}{\partial Y_j} \cdot \frac{\partial b}{\partial Y_j} & \sum_{j=1}^{n} s_j^2 \cdot \left(\frac{\partial b}{\partial Y_j}\right)^2 \end{bmatrix} \tag{3.11}$$

and for a linear-quadratic fit is

$$E = \begin{bmatrix} \sum_{j=1}^{n} s_j^2 \cdot \left(\frac{\partial b}{\partial Y_j}\right)^2 & \sum_{j=1}^{n} s_j^2 \frac{\partial b}{\partial Y_j} \cdot \frac{\partial c}{\partial Y_j} \\ \sum_{j=1}^{n} s_j^2 \frac{\partial b}{\partial Y_j} \cdot \frac{\partial c}{\partial Y_j} & \sum_{j=1}^{n} s_j^2 \cdot \left(\frac{\partial c}{\partial Y_j}\right)^2 \end{bmatrix}. \tag{3.12}$$

Diagonal elements of error matrix E are variances of estimators a and b (s_a^2 and s_b^2 , the linear fit), or b and c (s_b^2 and s_c^2 , the linear-quadratic fit). Off-diagonal elements represent covariance of estimators ($s_{a,b}$ for the linear fit and $s_{b,c}$ for the linear-quadratic fit). Covariance terms are written without squares according to the [Neter 90, p.5]. Numerical values of error matrices 3.11-3.12 were computed using Maple programs (see Appendices B-1 and B-2).

In order to determine the goodness of fit, t values were computed using the values of the estimators a, b, c and their standard deviations s_a , s_b , s_c . The t values for estimators a, b, and c were computed in the following manner:

$$t_a = \left|\frac{a}{s_a}\right| , t_b = \left|\frac{b}{s_b}\right| , t_c = \left|\frac{c}{s_c}\right|. \tag{3.13}$$

Using the t values, and tables of t -student distribution, the goodness of fit can be estimated. The discussion of the goodness of fit tests is given in Appendices A-1 and A-2.

3.2 Chi-square analysis

The Chi-square function can be used to compute confidence regions of fitted parameters. Denote with n the number of points to be fitted. Denote with m the number of parameters. The Chi-square defined with eq. 3.2 (linear fit) and eq. 3.7 (linear-quadratic fit) can be rewritten in a general form as

$$X^2 = \sum_{j=1}^{n} \frac{\left(Y_j - E[Y(\varepsilon_k, d_j)]\right)^2}{s_j^2} \tag{3.14}$$

where

d_j is received dose in a dose group labeled j

Y_j is excesses mortality rate in a dose group labeled j

X^2 is the chi-square function calculated for a given set of data (d_j, Y_j),

ε_k are parameters of fit. For a linear fit ε_k are α, and β. For linear-quadratic fit ε_k are β, and γ.

$E[Y(\varepsilon_k, d_j)]$ is a function of fit. For a linear fit it has the form: $\alpha + \beta \cdot d_j$.

For a linear-quadratic fit it has the form: $\beta \cdot d_j + \gamma \cdot d_j^2$.

(The regression parameter β is often called the linear effect coefficient, while γ is called the quadratic effect coefficient [Neter 90 p.316]).

Minimum of Chi-square X_0^2 is obtained by minimizing equation 3.14. It is

$$X_0^2 = \sum_{j=1}^{n} \frac{\left(Y_j - E[Y(e_{0,k}, d_j)]\right)^2}{s_j^2} \tag{3.15}$$

where

$e_{0,k}$ are specific values of estimators obtained by minimizing X^2. For linear fit $e_{0,k}$ are a_0, and b_0. For a linear-quadratic fit $e_{0,k}$ are b_0, and c_0.

$E[Y(e_{0,k}, d_j)]$ is function of the "best" fit which is obtained by minimizing Chi-square function 3.14.

The new function "delta Chi-square"(ΔX^2) is defined as [Pres 92, p. 692]:

$$\Delta X^2 = X^2 - X_0^2 \ . \tag{3.16}$$

Function ΔX^2 is distributed as the Chi-square distribution with m degrees of freedom [Press 96, page 690]. For the linear, and the linear-quadratic function used in this work m is equal to two (two parameters). ΔX^2 can be used to obtain the confidence regions for the parameters of fit ε_k [Press 92, p.687].

Equation 3.14 describes the functional dependence of the Chi-square function versus parameters ε_k. This dependence has a paraboloid form (see Figure 4.7). Intersections of that paraboloid with constant Chi-square planes (values are given in Table 3.1) give curves that define confidence regions for the parameters. The projection of these curves onto the parameter's axes give confidence intervals for each parameter of fit.

Table 3.1 Confidence levels as function of chi-square. The table is reproduced from [Press 92, p.692]

ΔX^2 as s a function of :						
Confidence level	degrees of freedom					
	1	2	3	4	5	6
68.3%	1.00	2.30	3.53	4.72	5.89	7.04
90%	2.71	4.61	6.25	7.78	9.24	10.6
95.4%	4.00	6.17	8.02	9.70	11.3	12.8
99%	6.63	9.21	11.3	13.3	15.1	16.8
99.73%	9.00	11.8	14.2	16.3	18.2	20.1
99.99%	15.1	18.4	21.1	23.5	25.7	27.8

3.3 Monte Carlo simulation

Assume that a set of data $Do\,(d_j, Y_{0j})$ is obtained by a measurement, (index j denotes j-th measured value) where d stands for dose and Y for excess in cancer (leukemia) mortality rate. The set Do is our information about the true set of data D_{true} that is only "known" to nature. Denote by $e_0(d_j, Y_j)$ the set of estimators which are obtained by minimizing the Chi-square function (in equation 3.14). Experimental uncertainties for measured values are $\sigma_{Y_{0j}}$ [Pres 92 p. 684]. Do is not the only possible realization of D_{true}. Repeated measurements would give other sets of data $D_i\,(d_j, Y_j)$ with estimators $e_i\,(d_j, Y_j)$ (index i denotes i-th experiment). In order to investigate other possible experimental outcomes one can simulate new events $D_i^s(d_j, Y_j)$ using a computer. This can be done using experimental set of data $Do(d_j, Y_{0j})$, and generating values of the dependent variable Y_j for each independent variable point d_j.

The generation of simulated points has to be done with experimental values of the uncertainties, $s_{Y_{0j}}$, because the true uncertainties, $\sigma_{Y_{0j}}$, are unknown. Thus, we assume $s_{Y_{0j}} = \sigma_{Y_{0j}}$. A diagram of this procedure named Monte Carlo, is shown in Figure 3.1. Then, each generated event has to be fitted by minimizing the Chi-square using eq. 3.14. This procedure, gives new sets of estimators $e_i^s(d_j, Y_j)$ which are subject of the statistical analysis.

For the simulation of new "measurements" in this thesis empirical values of the cancer and leukemia mortality rates were used from tables 2.1-2.4. The Gaussian distribution of excess mortality rates was assumed. The Gaussian distribution had a mean value equal to the experimented excess mortality rate and the standard deviation was taken

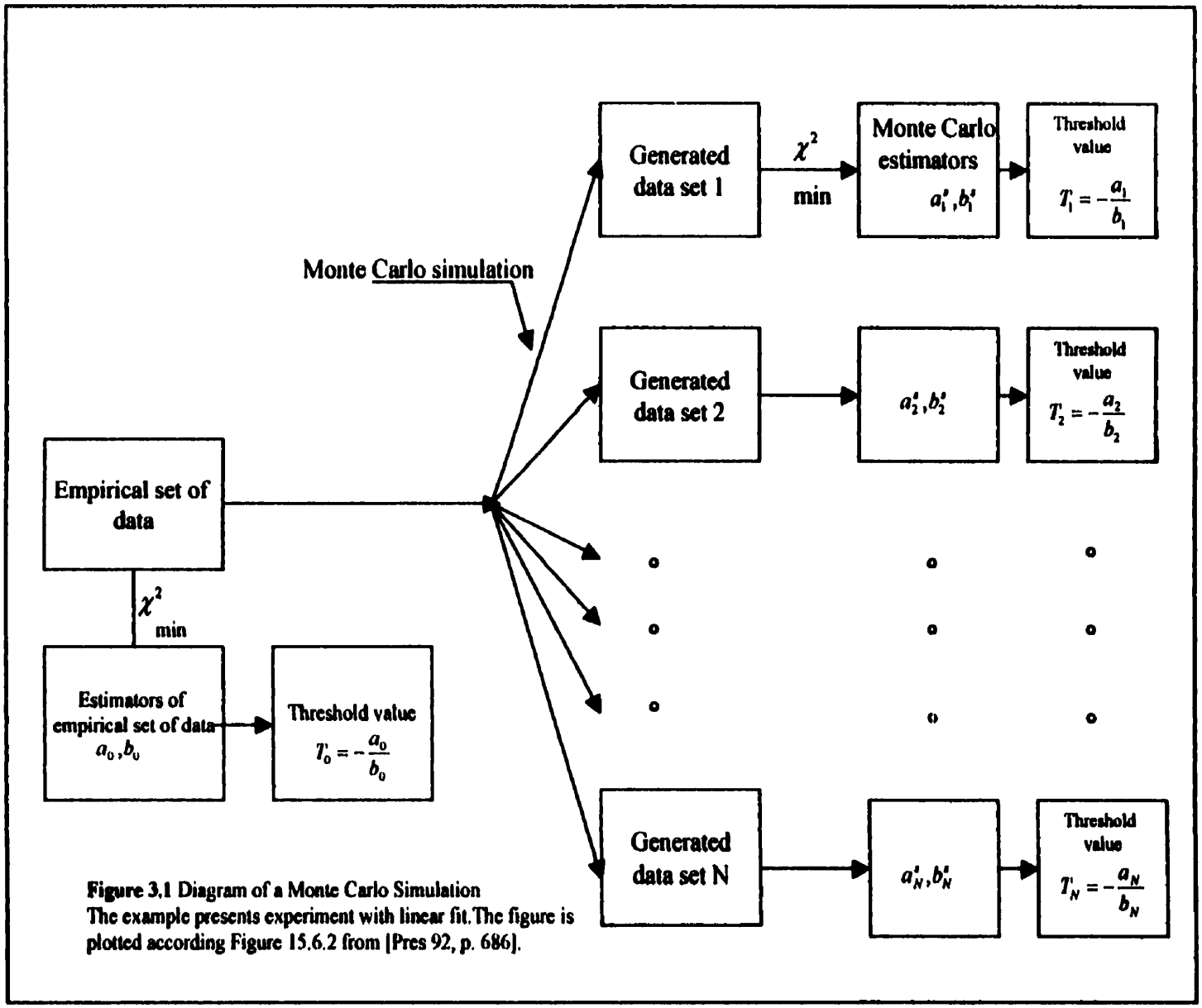

Figure 3.1 Diagram of a Monte Carlo Simulation
The example presents experiment with linear fit.The figure is plotted according Figure 15.6.2 from [Pres 92, p. 686].

to be equal to the uncertainty of corresponding cancer (leukemia) mortality rate (Tables 2.1-2.4, column 6).

In order to illustrate above, we consider the data from table 2.1 ([Pier 96] solid cancer data). Table columns "Mean Dose d_j", and "Excess cancer mortality rate Y_j" are of interest for simulation. Excess cancer mortality rates Y_j were simulated. The point with the mean dose 0.05 Sv ($j = 1$) was first. For this data point one thousand new values of excess mortality rates were generated as the Gaussian distributed values with mean value 2.44, and with a standard deviation which is equal to 2.20.This was done using the Minitab statistical program. This procedure was repeated using the corresponding values of the excess mortality rates and their uncertainties for the dose groups labeled j = 2,3,4,5.

In this way one thousand generated events were obtained for the excess cancer mortality rates. This one thousand generated events make simulated data set. Each simulated event has five values of cancer mortality rates. In the process of fit the mean dose values were kept the same as in the original data set. For each simulated event a least-squares fit was done and the values of the estimators a and b for linear, or b and c for linear-quadratic, model were computed.

3.3.1 Fitting by a linear function

Assuming that the linear dependence between excess of deaths due to solid cancers (or leukemia) and received dose exists, each simulated set of data was fitted by a straight line. The least-squares fit was used to find estimators for each set of simulated points (section 3.1.1).

Specifically, for each simulated set of points (d_j, Y_{ji}) (i denotes the simulation

number, and j denote a specific dose point in the i-th simulation) the slope and the segment of the straight line was calculated using equations 3.3-3.4. Repeating this for one thousand new simulated events generates one thousand values of an intercept a and slope b.For this task a Maple program was written.

One thousand values for an intercept a and slope b determine one thousand straight lines. For each line intercept to the dose axis was calculated using

$$T_i = -a_i / b_i \tag{3.17}$$

Solutions of 3.17 are called threshold. These solutions can be separated into two groups. One group of the solutions are the threshold values that are positive in sign, and the other group includes solutions with a negative sign. The group of solutions (with positive T) are values permissible under the linear-threshold model. The group solutions with negative T describe non-threshold models. For these one thousand thresholds the mean value, the trimmed mean value, the standard deviation, the standard error of the mean, the first and third quartiles are calculated using program Minitab. The formulae for these calculations are given in Appendix A-3.

3.3.2 Fitting by a linear-quadratic function

Besides a linear fit, a linear-quadratic function was also used in fitting leukemia data. The procedure is basically the same as the linear fit described in section 3.3.1. The linear-quadratic function was used in the form given in equation 3.6 (section 3.1). Equations 3.8-3.9 were used to calculate estimators b and c for each set of simulated data. One thousand simulations determined one thousand parabolas forced through the

coordinate origin (equation 3.6).The models described with these parabolas were discussed in section 1.2.

The intersection of a parabola to the dose axis is called the zero equivalent point (*ZEP*) (see section 1.2). The values of *ZEP* were calculated by solving the equation

$$d_{ZEP_i} \cdot (b_i + c_i \cdot d_{ZEP_i}) = 0, \tag{3.18}$$

giving

$$d_{ZEP_i} \equiv ZEP_i = -b_i / c_i. \tag{3.19}$$

The statistical analysis of *ZEP* values was done in the same manner as the statistical analysis of the threshold values for the linear fit.

3.4 Uncertainties in the independent variable

Sections 3.1-3.2 described the fitting procedure with the uncertainties only in the dependent variable (*Y*). This method is valid only if the uncertainties in the independent variable (*d*) were much smaller than the uncertainties in the dependent variable. If this were not the case, or if someone wanted to study influence of uncertainties in the independent variable on the results, those uncertainties can be taken in account by combining *d* and *Y* uncertainties as described in [Bevi 92, page 100].

Let s_d be the uncertainty in the independent variable and s_Y in the dependent variable. Let $\hat{Y}(e_k, d)$ be the fitted value at dose *d* and e_k be the estimators of the fit (*a* and *b* for linear fit, *b* and *c* for linear-quadratic fit, see section 3.2). The total uncertainty in the dependent variable *Y*, which is labeled $s_{Y,T}$, can be calculated by adding the uncertainties for the only independent s_d and the only dependent s_Y variables in the following manner:

$$s_{Y,T}^2 = s_Y^2 + \left(\frac{\partial \hat{Y}(e_k, d)}{\partial d} \right)^2 \cdot s_d^2 . \tag{3.20}$$

The total uncertainty in the dependent variable can be used for minimizing the Chi-square function (equation 3.14, section 3.1). By using equation 3.20, equation 3.14 becomes

$$X^2 = \sum_{j=1}^{n} \frac{\left(Y_j - a - b \cdot d_j \right)^2}{s_{Y,j}^2 + b^2 \cdot s_{d,j}^2} \tag{3.21}$$

for the linear fit. For the linear-quadratic fit, chi-square function (3.14) becomes

$$X^2 = \sum_{j=1}^{n} \frac{\left(Y_j - b \cdot d_j - c \cdot d_j^2 \right)^2}{s_{Y,j}^2 + (b + 2 \cdot c \cdot d_j)^2 \cdot s_{d,j}^2} . \tag{3.22}$$

In order to find the values of estimators a and b for linear model, and b and c for linear-quadratic model the Chi-square function in equations 3.21 and 3.22 has to be minimized, and estimators corresponding to Chi-square minimum be numerically found.

Minimizing the Chi-square in equations 3.21 or 3.22 was done by finding derivatives by parameters of fit e_k:

$$\frac{\partial X^2}{\partial e_k} = 0 . \tag{3.23}$$

After partial derivatives were found, the system of equations 3.23 was solved for estimators e_k. System 3.23 was computed and solved numerically for estimators of fit e_k, by writing a Maple program. Comparison of the results with and without dose uncertainties is included in chapter 6.

Chapter IV

Results for the linear fit

This chapter presents results of the standard statistical analysis and of the Monte Carlo simulation of data. Method used was summarized in chapter 3. Dose uncertainties are not included in this analysis.

4.1 Least-squares fit

The excess cancer and leukemia mortality rates with their errors are all taken from column six of Tables 2.1, 2.2, 2.3 and 2.4. The graphical representation of the excess cancer (leukemia) mortality rates as a function of received dose (column 3 in all tables) are presented in Figures 4.1-4.6. The error bars are equal to the uncertainties of the excess mortality rates. Each set of these empirical data is fitted by a linear function using equations 3.3-3.4. The result is represented by the heavy line in Figures 4.1-4.6.

Table 4.1 contains results for best line fits shown in figures 4.1-4.6. The specific values of estimators (estimates) a and b are in columns 3 and 4. Threshold values are in column 5. Standard deviations of estimators s_a and s_b are in columns 6 and 7. Standard deviations are equal to square root of variances s_a^2 and s_b^2. The variances s_a^2 and s_b^2 and covariance $s_{a,b}$ of a and b were calculated using error matrix 3.11. The t values were calculated using 3.13. The p probabilities are obtained using the table of the t distribution [Neter 90, p.1128]. More details are provided in Appendix A-1.

In this section, errors for threshold values were not calculated because the commonly used formula for combining uncertainties [Bevi 92 p. 50]

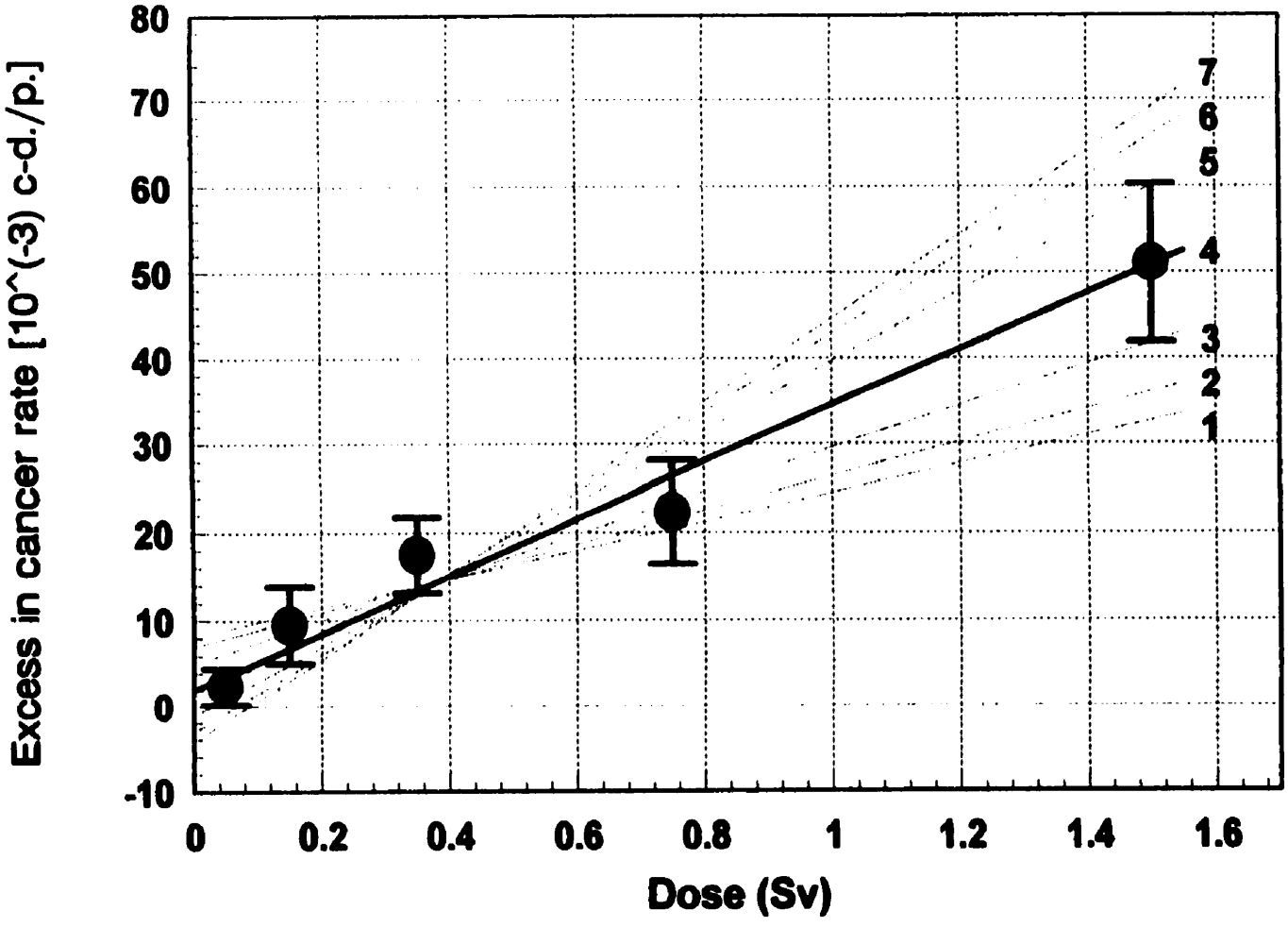

Figure 4.1 Excess in cancer rates for [Pier 96] solid cancers as a function of received dose. The heavy line is the best fit line, light lines are plotted using parameters fromTable 4.2. Labeling is according to Table 4.2. (c-d./p. refers to cancer deaths per person.)

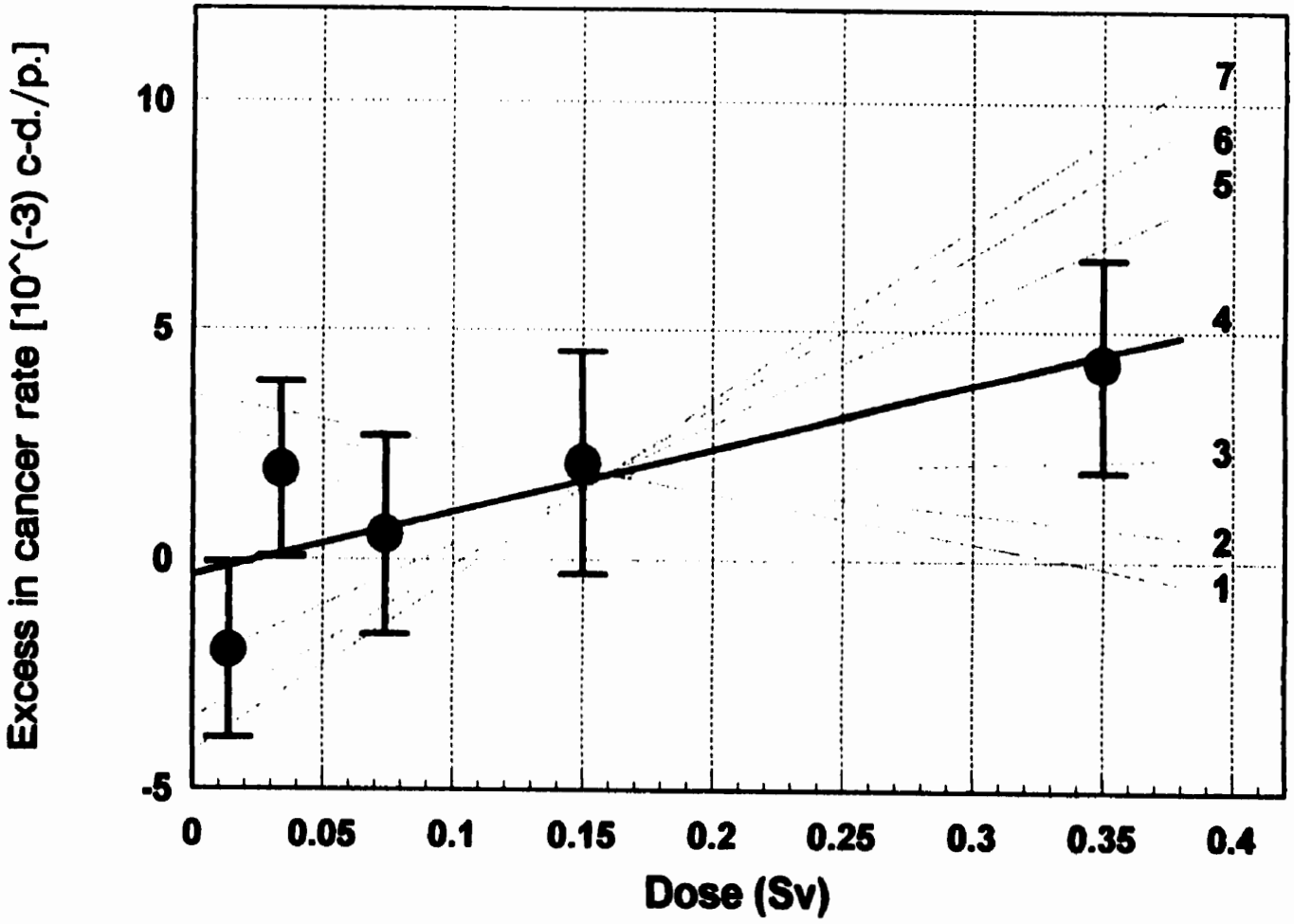

Figure 4.2 Excess in cancer rates for [Shim 92] stomach cancers as a function of received dose. The heavy line is the best fit line, light lines are plotted using parameters fromTable 4.2. Labeling is according to Table 4.2. (c-d./p. refers to cancer deaths per person.)

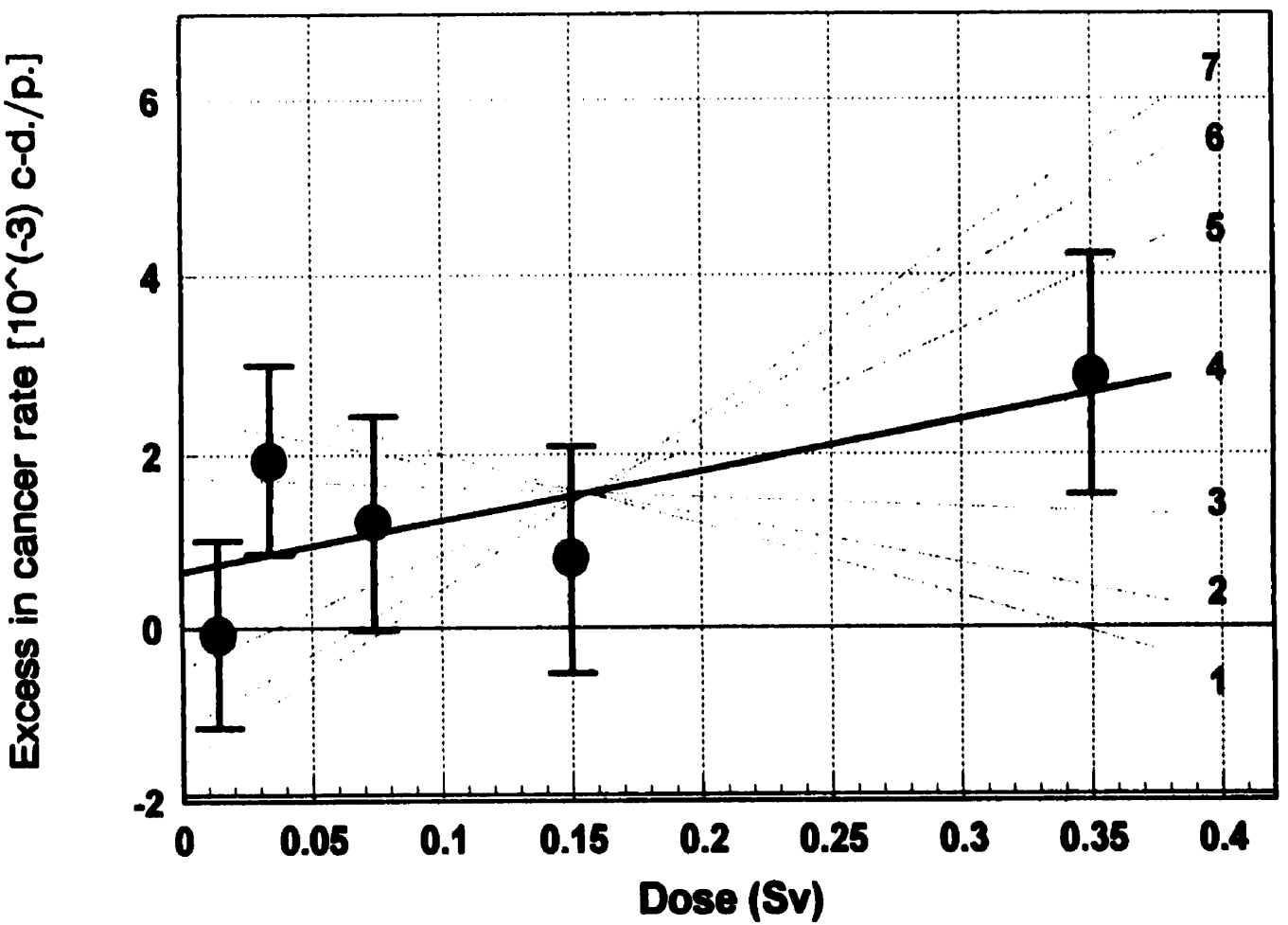

Figure 4.3 Excess in cancer rates for [Shim 92] lung cancers as a function of received dose. The heavy line is the best fit line, light lines are plotted using parameters fromTable 4.2. Labeling is according to Table 4.2.
(c-d./p. refers to cancer deaths per person.)

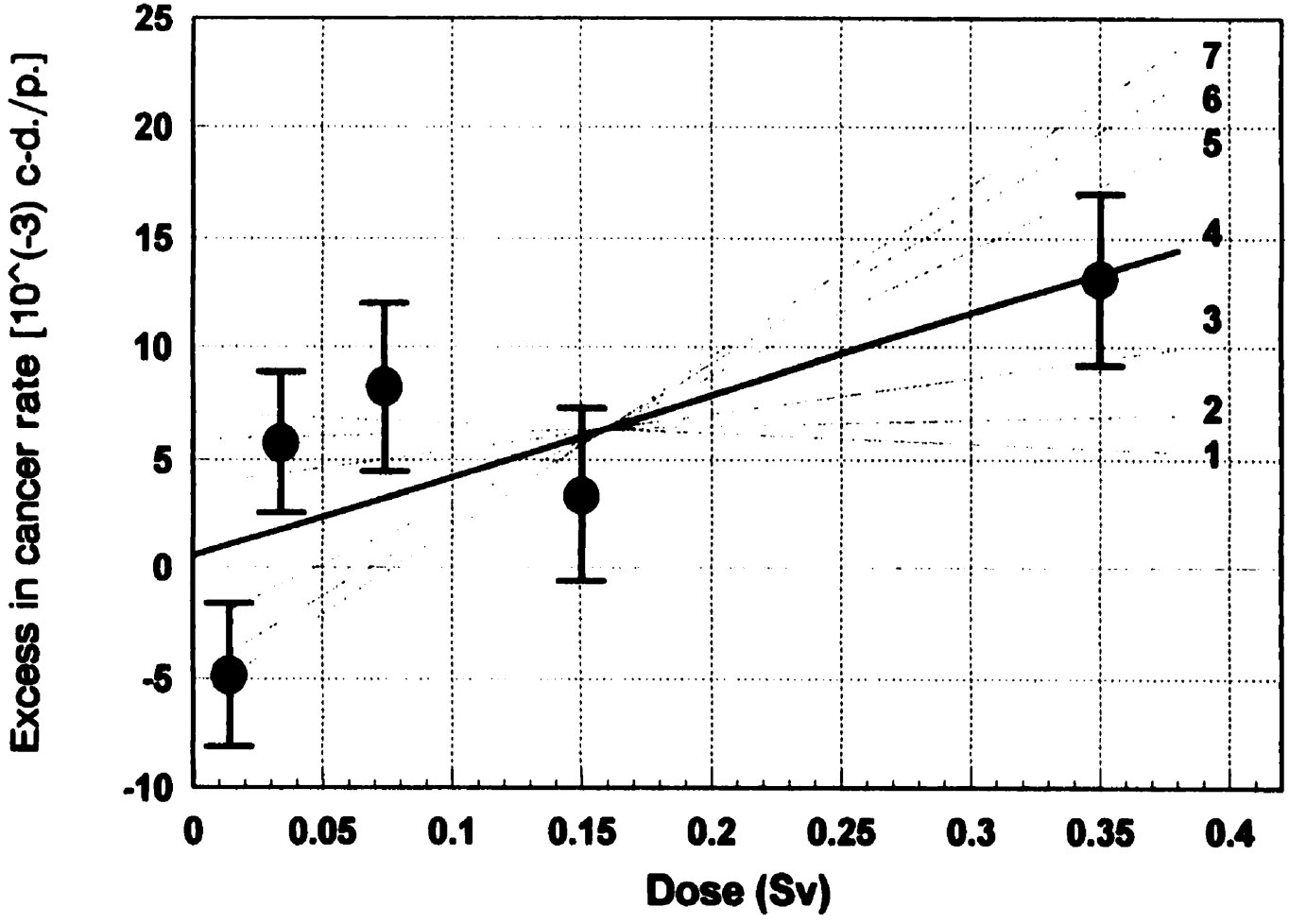

Figure 4.4 Excess in cancer rates for [Shim 92] all solid cancers as a function of received dose. The heavy line is the best fit line, light lines are plotted using parameters fromTable 4.2. Labeling is according to Table 4.2. (c-d./p. refers to cancer deaths per person.)

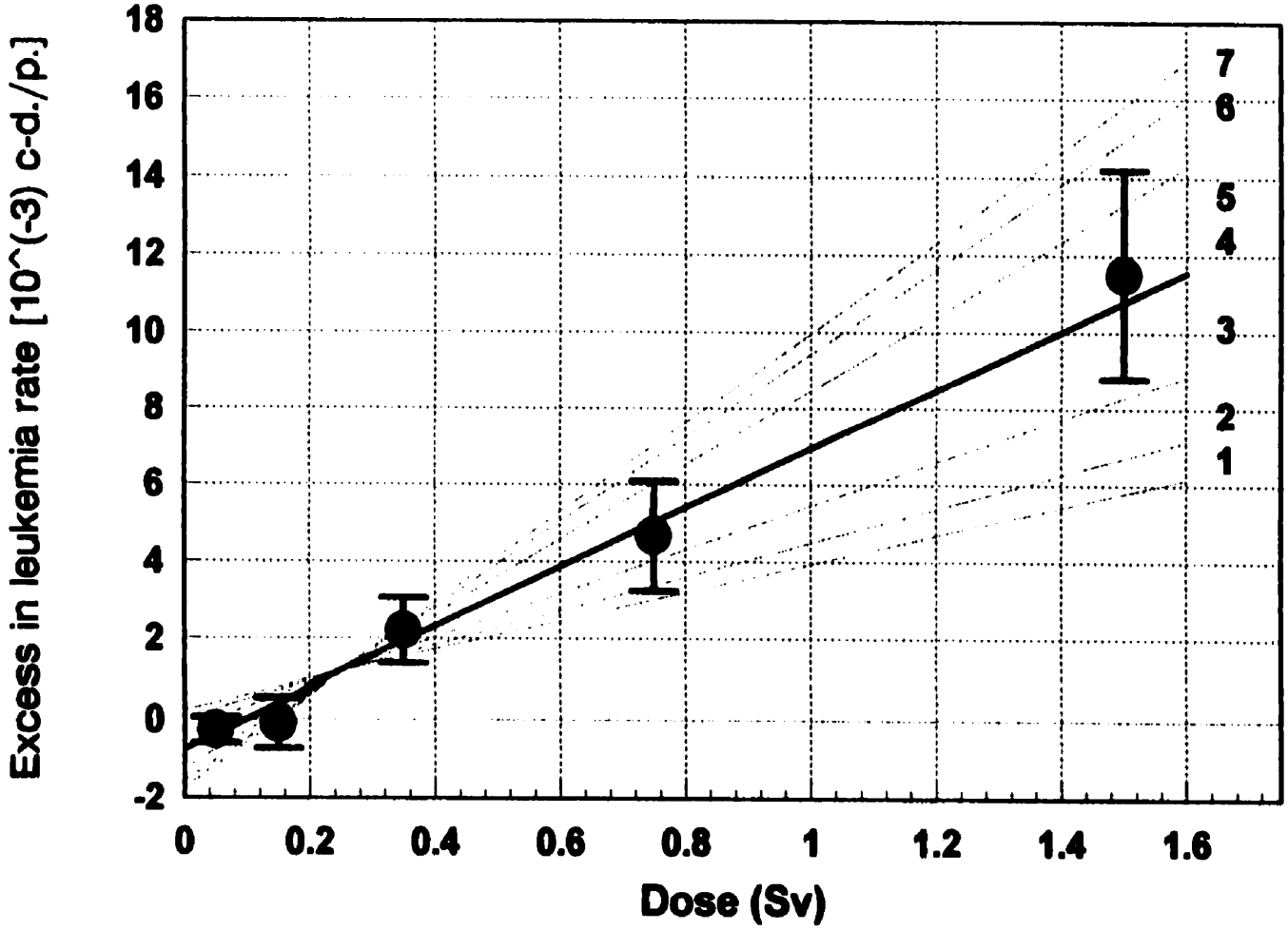

Figure 4.5 Excess in leukemia rates for [Pier 96] leukemia as a function of received dose. The heavy line is the best fit line, light lines are plotted using parameters fromTable 4.2. Labeling is according to Table 4.2.

(c-d./p. refers to leukemia deaths per person.)

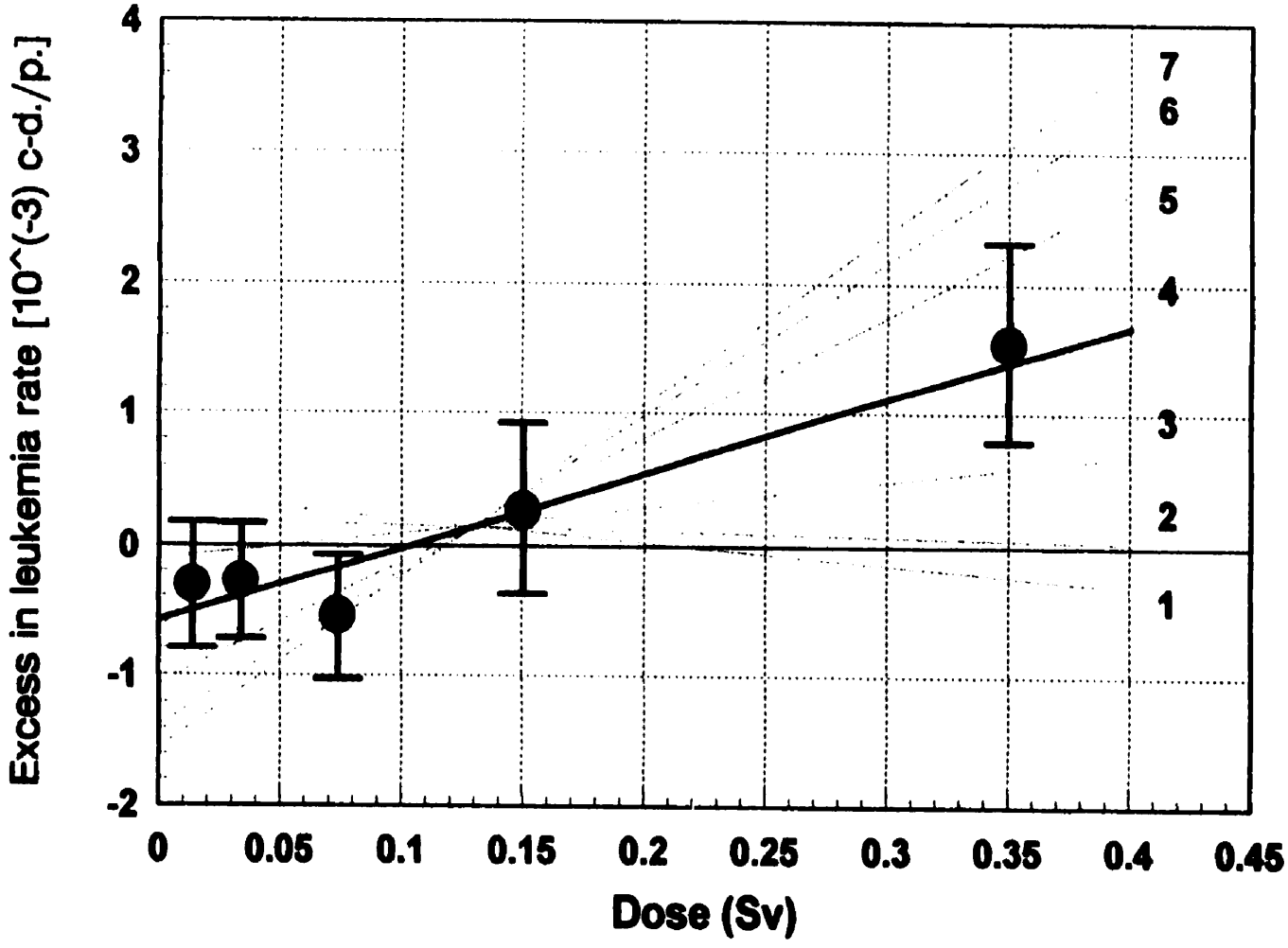

Figure 4.6 Excess in leukemia rates for [Shim 92] leukemia as a function of received dose. The heavy line is the best fit line, light lines are plotted using parameters fromTable 4.2. Labeling is according to Table 4.2. (c-d./p. refers to leukemia deaths per person.)

Table 4.1 The values for estimators of fit and their corresponding uncertainties obtained by the least-squares method. Corresponding threshold values ($T = -a / b$) are included, for explanations of errors and goodness of fit see text.

Figure	Data set	Intercept *(a)*	Slope *(b)*	T (Sv)	s_a	s_b	$s_{a,b}$	t_a	t_b	$p_{(a)}$	$p_{(b)}$
4.1	[Pier 96] solid cancers	1.99	32.6	-0.061	2.04	5.38	-6.25	0.97	6.06	0.60	0.01
4.2	[Shim 92] Stomach cancer	-0.33	13.9	0.024	1.28	8.04	-6.94	0.26	1.73	0.80	0.20
4.3	[Shim 92] Lung cancer	0.64	5.88	-0.108	0.73	4.66	-2.29	0.88	1.26	0.80	0.30
4.4	[Shim 92] All solid cancers	0.56	36.6	-0.015	2.16	13.6	-19.9	0.26	2.70	0.80	0.10
4.5	[Pier 96] leukemia	-0.76	7.70	0.099	0.33	1.32	-0.24	2.29	5.84	0.20	0.02
4.6	[Shim 92] leukemia	-0.59	5.68	0.104	0.32	2.46	-0.52	1.87	2.31	0.20	0.20

$$\Delta T = \frac{a}{b} \cdot \sqrt{\frac{s_a^2}{a^2} + \frac{s_b^2}{b^2} - 2 \cdot \frac{s_{a,b}}{a \cdot b}} \qquad (4.1)$$

is not valid because relative errors $\Delta a / a$ and $\Delta b / b$ are not much smaller than one. For example, ratio $\Delta a / a$ for [Pier 96] solid cancer is 2.04 / 1.99 >1 (see Table 4.1)

4.2 Chi-square analysis

In addition to the values of the estimators a and b and their standard deviations for the best fit, confidence regions for these parameters were calculated. Confidence regions are represented with areas enclosed by delta Chi-square ellipses. Regions were calculated using the confidence levels of ΔX^2 .The chosen confidence levels are 68.3%, 95.4%, and 99% (Table 3.1). Appropriate tail values of the delta Chi-square statistic as a function of the confidence levels, and degrees of freedom are given in Table 3.1. From that table, for two degrees of freedom the delta Chi-square values which correspond to 68.3%, 95.4%, and 99% of the confidence are 2.30, 6.17, and 9.21 for parameters α , and β jointly [Pres 92, p. 688].

Figure 4.7 represents the delta Chi-square paraboloid for the [Pier 96] solid cancer data set. Horizontal planes cut the paraboloid at ΔX^2 equal to 2.30, 6.17, and 9.21. The intersections have a form of ellipses which are shown in Figure 4.8. The projections of the ellipses onto coordinate axes give the joint confidence intervals for parameters α and β. In Figure 4.8 projections labeled 3, and 5 correspond to 68.3% confidence, projections labeled 2, and 6 correspond to 95.4% confidence, and projections labeled 1, and 7 correspond to 99% confidence. The point labeled $j = 4$ corresponds to the values of

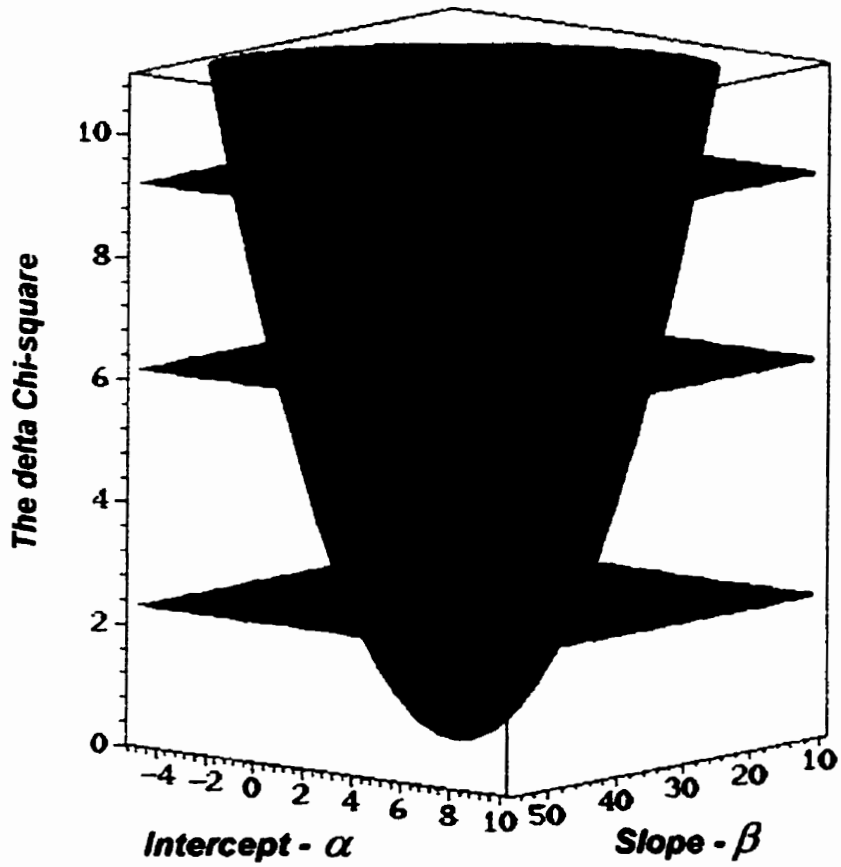

Figure 4.7 The delta Chi-square paraboloid (equation 3.16) for [Pier 96] solid cancer data (Table 2.1). The data set is fit by a linear function with intercept α and slope β. The intersections of the horizontal planes and the paraboloid correspond to 68.3%, 95%, and 99% joint confidence regions for the parameters α and β. These intersections have the form of ellipses (Figure 4.8).

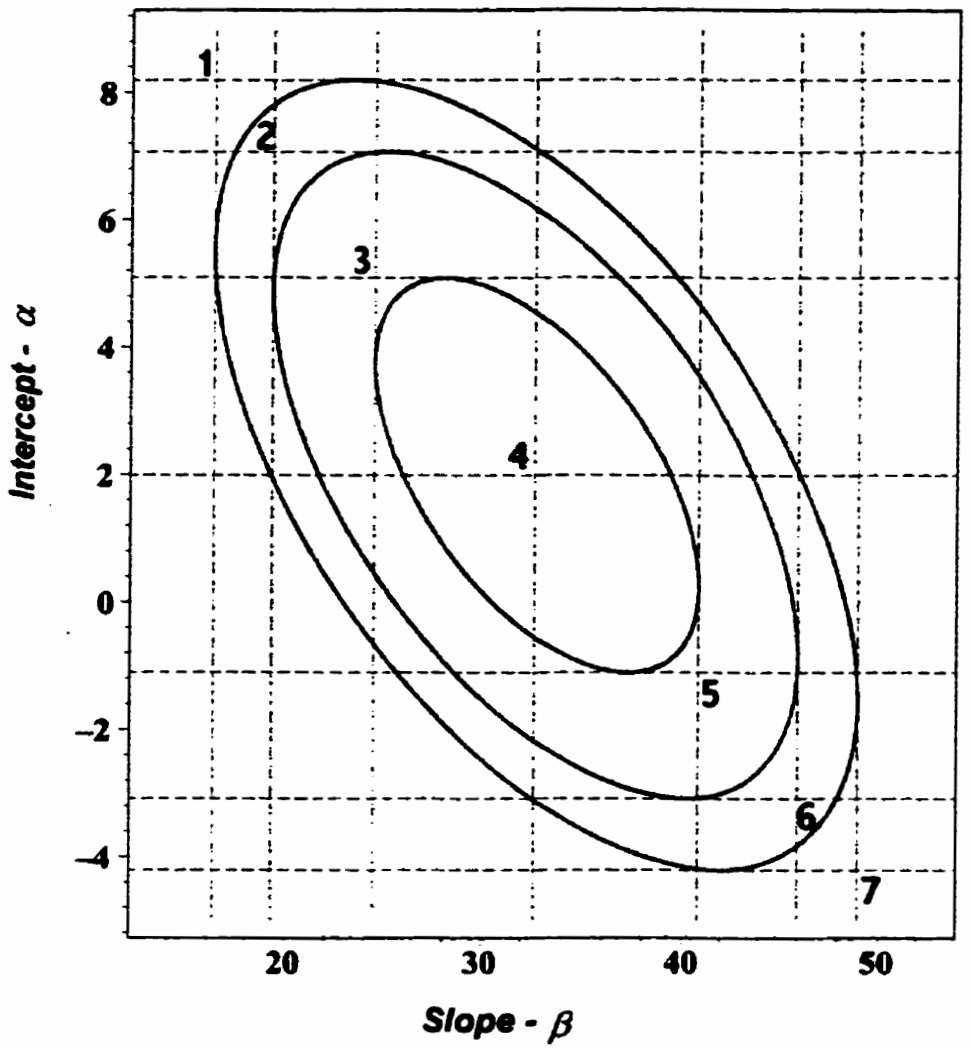

Figure 4.8 The delta Chi-square ellipses for [Pier 96] solid cancer data, Table 2.1 plotted as function of intercept *a*, and slope *b*. The ellipses are obtained by projecting the intersections of the delta Chi-square paraboloid (Figure 4.7) and the constant delta Chi-square planes equal to the 2.30, 6.17, and 9.21. The areas enclosed by ellipses are 68.3%, 95%, and 99% confidence regions for parameters α and β jointly. Values of projections labelled 1,..,7 are given in Table 4.2.

parameters for the "best" fit ($\Delta X^2 = 0$).

Table 4.2 lists end points for the confidence limits of estimators a and b. The relevant ellipses are shown in Figure 4.8 for [Pier 96] solid cancer data. This table also includes the confidence limits for all other data sets computed in the same manner as [Pier 96] solid cancer data set. These numerical values of the projections of the delta Chi-square ellipses onto coordinate axes were obtained using a program written in Maple. In addition to the confidence levels for estimators, the threshold values, the Chi-square value, and the reduced Chi-square value ($X^2 / 3$) are listed. The threshold values were computed using equation 3.17 for each line in Figures 4.1-4.6. The threshold values (labeled $j = 4$) correspond to the intercepts of the "best" fit lines and the dose axis. The other threshold values given (j=1,2,3,5,6,7) correspond to the (α, β) pairs obtained from projections of the appropriate delta Chi-square ellipses onto axes (see Figure 4.8). The dose intercept of the appropriate line is the threshold.

The value of Chi-square was calculated using equation 3.2. To obtain the point estimates, reduced Chi-square was calculated for three degrees of freedom. Chi-square defined in equation 3.2 has $\nu = n - m = 5 - 2 = 3$ degrees of freedom. Labeling is done according to figure 4.8. The straight lines using estimators from Table 4.2 are presented in Figures 4.1-4.6. The lines labeled with 4 are "best" fit lines. The lines which are labeled with 1,2,3,5,6,7 have the values of α , and β as shown in Table 4.2, and are labeled with the same indices j.

Table 4.2 The estimates of intercept ***a*** and slope ***b*** are obtained by projecting the Chi-square ellipses onto the corresponding axes. The threshold, the Chi-square, and the reduced Chi-square values are also in the table below.

Data Set	Proj. (j)	Intercept (a_j)	Slope (b_j)	Threshold ($-\frac{a_j}{b_j}$) (Sv)	X_o^2	$X^2_{o\,red}$
[Pier 96] solid cancers	1	8.19	16.3	-0.503	11.3	
	2	7.07	19.3	-0.367	8.23	
	3	5.09	24.5	-0.208	4.36	
	4	**1.99**	**32.6**	**-0.061**	**2.06**	**0.69**
	5	-1.11	40.8	0.027	4.36	
	6	-3.09	46.0	0.067	8.23	
	7	-4.21	48.9	0.086	11.3	
[Shim 92] stomach cancer	1	3.54	-10.5	0.338	11.1	
	2	2.84	-6.04	0.471	8.06	
	3	1.60	1.74	-0.921	4.19	
	4	**-0.33**	**13.9**	**0.024**	**1.89**	**0.63**
	5	-2.27	26.1	0.087	4.19	
	6	-3.50	33.9	0.103	8.06	
	7	-4.21	38.3	0.110	11.1	
[Shim 92] lung cancer	1	2.84	-8.26	0.344	11.1	
	2	2.44	-5.69	0.429	8.06	
	3	1.74	-1.18	1.47	4.19	
	4	**0.64**	**5.88**	**-0.108**	**1.89**	**0.63**
	5	-0.46	12.9	0.036	4.19	
	6	-1.17	17.4	0.067	8.06	
	7	-1.57	20.0	0.078	11.1	

Table 4.2 continued.

Data Set	Proj. (j)	Intercept (a_j)	Slope (b_j)	Threshold ($-\frac{a_j}{b_j}$) (Sv)	X_o^2	$X_{o\,red}^2$
[Shim 92] all solid cancers	1	7.12	-4.63	1.54	16.3	
	2	5.93	2.85	-2.08	13.3	
	3	3.84	16.0	-0.240	9.40	
	4	**0.56**	**36.6**	**-0.015**	**7.10**	**2.37**
	5	-2.72	57.2	0.047	9.40	
	6	-4.81	70.3	0.068	13.3	
	7	-6.00	77.8	0.077	16.3	
[Pier 96] leukemia	1	0.23	3.70	-0.063	10.1	
	2	0.05	4.43	-0.012	7.10	
	3	-0.26	5.70	0.046	3.23	
	4	**-0.76**	**7.7**	**0.099**	**0.93**	**0.31**
	5	-1.25	9.71	0.129	3.23	
	6	-1.57	11.0	0.143	7.10	
	7	-1.74	11.7	0.149	10.1	
[Shim 92] leukemia	1	0.38	-1.77	0.215	10.1	
	2	0.20	-0.42	0.490	7.10	
	3	-0.10	1.96	0.053	3.20	
	4	**-0.59**	**5.68**	**0.104**	**0.90**	**0.30**
	5	-1.07	9.40	0.114	3.20	
	6	-1.38	11.8	0.117	7.10	
	7	-1.55	13.1	0.118	10.1	

Table 4.2 includes some parameters of β which are negative in sign (a negative slope). There is no evidence of negative correlation between cancer mortality and the received dose in whole low, and medium dose ranges, however they are in the limits of required confidence.

Intersections of lines with negative slope do not have meaning of threshold. This is why negative slopes are recognized as not of interest. Figure 4.2 (stomach cancer), 4.3 (lung cancer), and 4.6 ([Shim 92] leukemia) all have negative slope lines. It can be seen from Table 4.2 that the best fit line has the positive threshold for [Shim 92] stomach cancer and both leukemia data sets. The other three sets of solid cancer data have negative threshold values. Negative threshold values are consistent with supra-linear model (see Figure 1.1).

4.3 Discussion of the goodness of fit

In order to estimate the goodness of the "best" fit lines, values of t ratio (equation 3.13) and their corresponding p values, the Chi-square values (equation 3.2 for linear fit), and the reduced Chi-square were calculated. The numerical values of these calculations for the linear fit are presented in Tables 4.1 and 4.2. Table 4.1 contains t and, p values. Table 4.2 contains the Chi-square, and the reduced Chi square values. Index $j = 4$ labels "best" fit lines in these two tables.

The two sided t test was used to test whether or not a linear relationship exists between the excess mortality rates and the received dose (is slope b different from zero). The description of this test is included in Appendix A-1.

The t values t_b, and the probability values $p_{(b)}$ which were obtained using table

of the t distribution are presented in two of the last three columns of Table 4.1. The $p_{(b)}$ value (i.e. probability of observing this or a larger value of $|t_b|$ if the true slope is zero - less evidence of a significant slope) for [Pier 96] solid cancer data set is 0.01 for a linear relationship between the excess mortality rate and the received dose. The $p_{(b)}$ value for [Pier 96] leukemia data is 0.02. The $p_{(b)}$ values for the [Shim 92] sets are higher than for the [Pier 96] data sets (stomach and leukemia 0.20, and all solid cancers 0.10). The highest $p_{(b)}$ value is for the lung cancer data (0.30).

The reduced Chi-square values for a good fit should be close to one [Bevi 92 p.195]. The values greater than one are due to high values of squares of deviations between the points and the fit line. The values which are very small suggest unusually high uncertainties in variables. The reduced Chi-square values for the best fit lines (j =4) are 0.69, 0.63, 0.63, 0.31, 0.30 for [Pier 96] solid cancer set, [Shim 92] stomach, [Shim 92] lung cancer sets, [Pier 96] leukemia, [Shim 92] leukemia data sets, respectively. The fact that all of these reduced Chi-square values are quite a bit smaller than one suggests that the assumed errors are too large. The linear fit of the [Shim 92] all solid cancers set of data has the poorer Chi-square value of 2.37. The best fit lines which are labeled 4 in figures 4.1, 4.5. and 4.6 appear to fit the "data" well, the lines in Figures 4.2 and 4.3 fit somewhat less well. The best line in Figure 4.4 ([Shim 92] all solid cancers), fits least well of all six best lines. Figure 4.4 shows that the best line goes outside of error bars that are determined by the values of excess mortality rates and their uncertainties. Specifically, the best line goes only through two out of five intervals determinated with error bars. This high dispersion around the best fit line gives higher value for the Chi-square.

High Chi-square values, and low $p_{(b)}$ values for this fit suggest a poor fit.

Both [Pier 96] sets of data (solid cancers and leukemia) have good linear fits. The [Shim 92] data sets have worse fits than [Pier 96]. The [Shim 92] stomach, lung, leukemia fits have low Chi-squares, but higher $p_{(b)}$ values. This suggest a weaker linear correlation.

4.4 Using Monte Carlo simulation to determine threshold errors

In order to get an estimate of threshold confidence intervals, a Monte Carlo simulation was done as described in section 3.3. For each of six sets of excess mortality rates (Tables 2.1 to 2.4), one thousand simulated events were generated and fitted with straight lines as described in section 3.3.1. This procedure gives one thousand lines. The intersections of these lines to the dose axis determine one thousand threshold values.

Figures 4.9 and 4.10 present the distributions of the estimators (intercept a and slope b) obtained by the simulation for the [Pier 96] solid cancer set of data. The area under Gaussian distribution curves is normalized to one.

Figures 4.11-4.16 are histograms of the simulated threshold values for each cancer category. Distributions from Figures 4.11-4.16 are asymmetrical and are skewed on the left side. In order to present (almost) all simulated points the frequency axes in Figures 4.12-4.14, and 4.16 are logarithmic. A certain number of threshold values in Figures 4.13, 4.14, and 4.16 are omitted. In order to show threshold distribution produced by lines with positive slope only, the histograms 4.12a-4.14a, and 4.16a were plotted. In order to better present the form of the distribution, a certain number of points from the left side is also omitted in these histograms. All threshold values for [Pier 96] solid cancer, and leukemia data are produced by lines with positive slopes (see Figures 4.11 and 4.15, and Table 4.4).

Some threshold values are extremely high (an example, omitted point 452 Sv , Figure 4.13). In this case the simulation has produced a fit line with a small negative slope. This line intercepts the dose axis on the far right side.

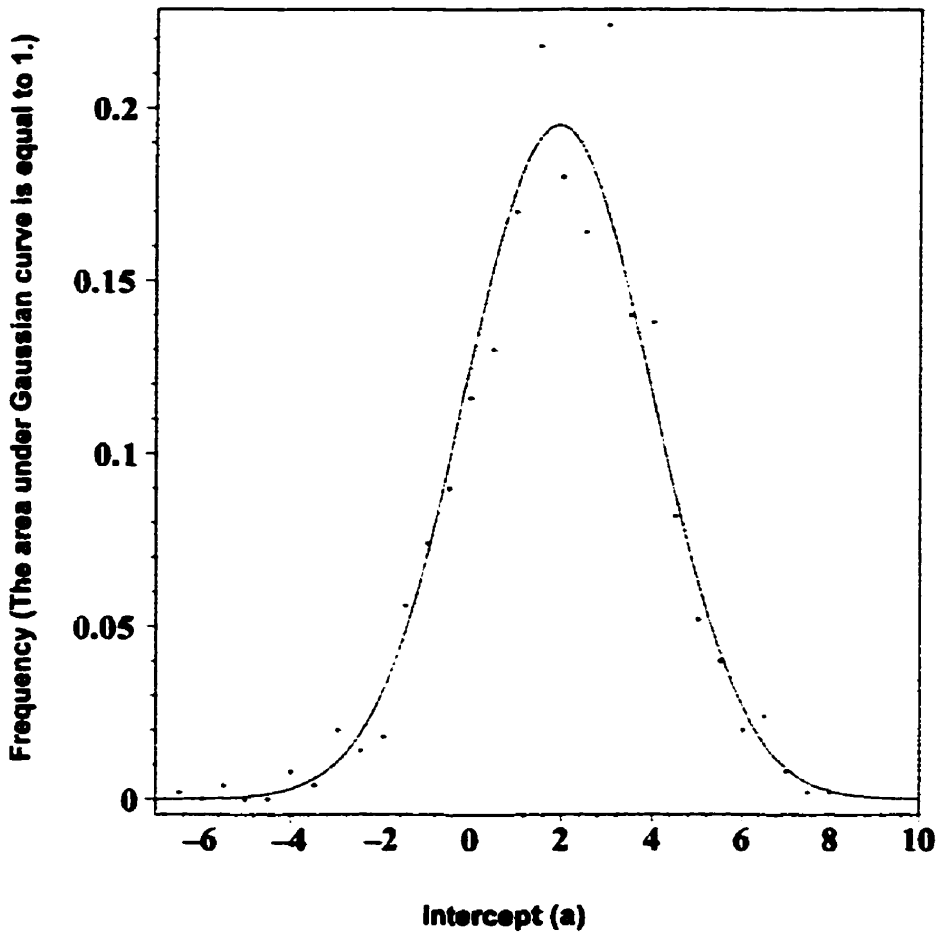

Figure 4.9 The distribution of the simulated intercepts a and corresponding Gaussian fit. The mean value and standard deviation of Gaussian curve are given in Table 4.3. The areas under the Gaussian curve and the histogram are normalized to one. ([Pier 96] solid cancer data).

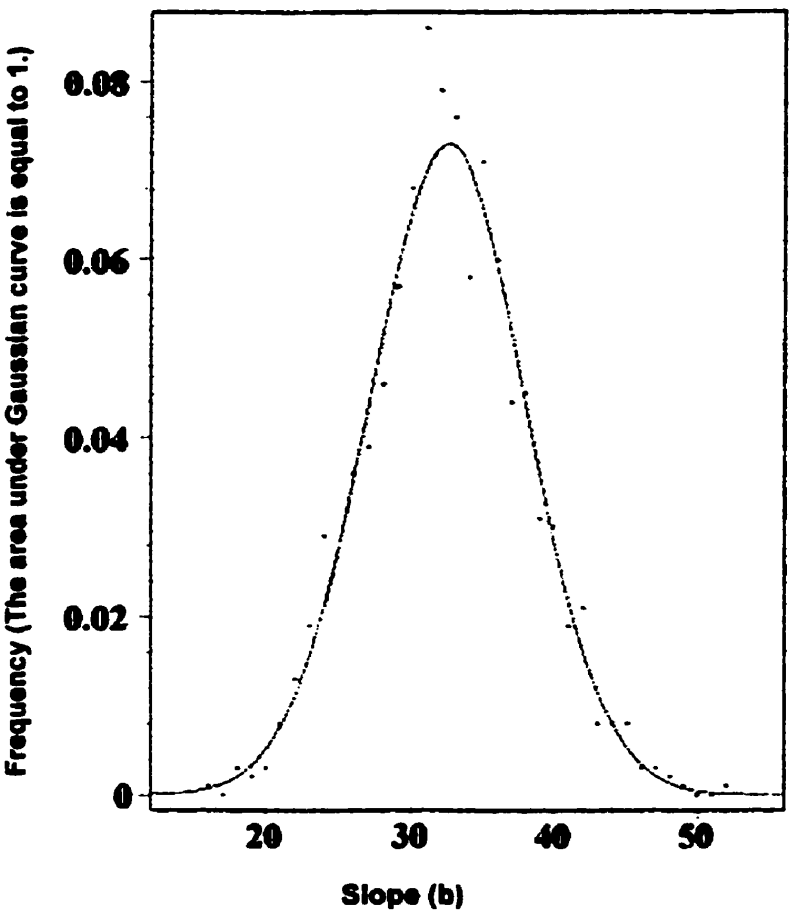

Figure 4.10 The distribution of the simulated slopes *b* and corresponding Gaussian fit. The mean value and standard deviation of Gaussian curve are given in Table 4.3. The areas under the Gaussian curve and the histogram are normalized to one. ([Pier 96] solid cancer data).

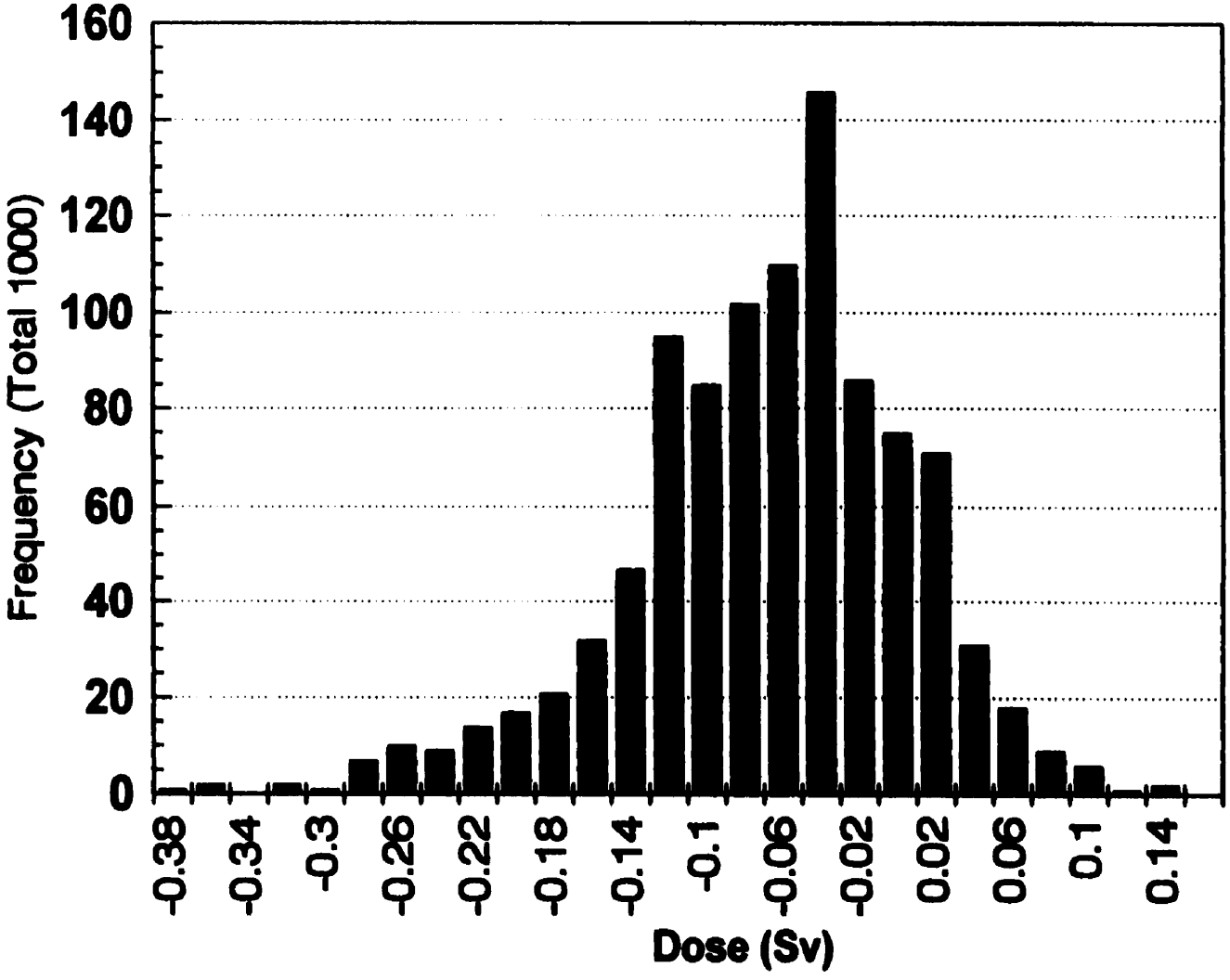

Figure 4.11 The histogram of threshold values for solid cancers obtained by simulation, ([Pier 96] data, Table 2.1 and Figure 4.1).

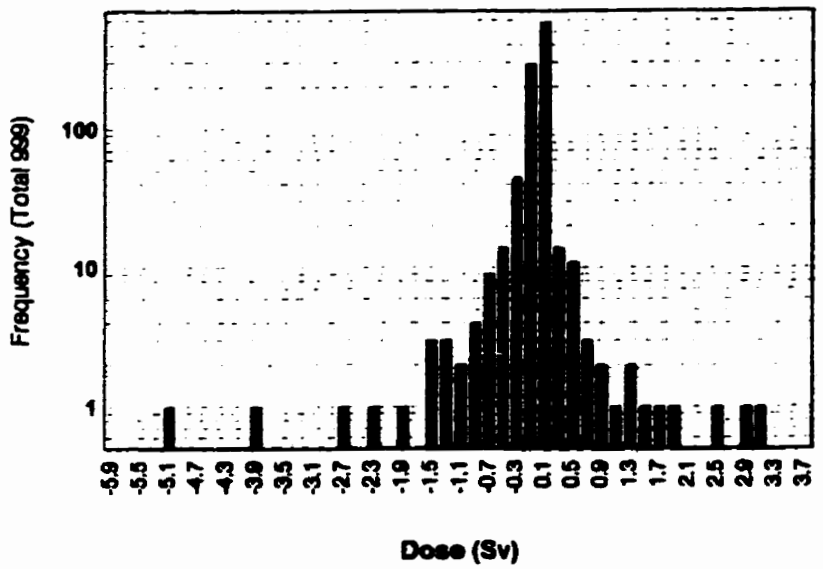

Figure 4.12 Histogram of threshold values for stomach cancers, ([Shim 92], Table 2.3 and Figure 4.2). Threshold value at -13.3 is omitted.

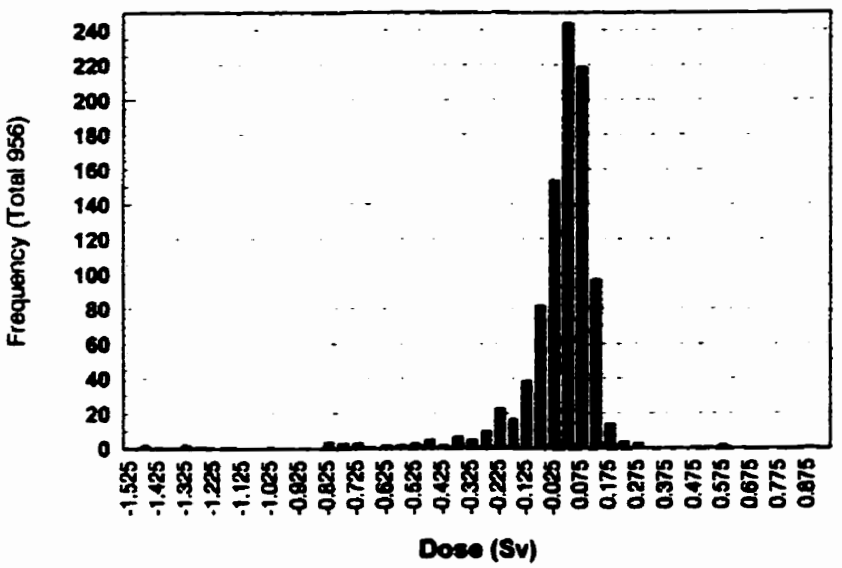

Figure 4.12a Threshold histogram produced by lines with positive slope b. Total 962 thresholds produced by lines with b>0, see Table 4.4). The smallest 6 threshold values are omitted. (stomach cancer data).

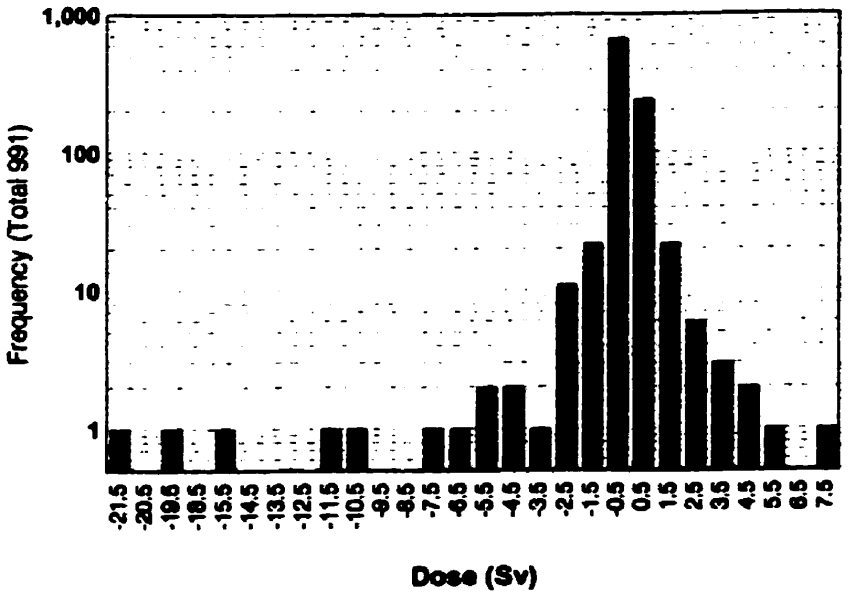

Figure 4.13 Histogram of threshold values for Lung cancer, ([Shim 92], data Table 2.3 and Figure 4.3). Threshold values: -47.9, -27.4, -26.9, 14.5, 23.3, 42.6, 46.7, 59.6, and 452 are omitted.

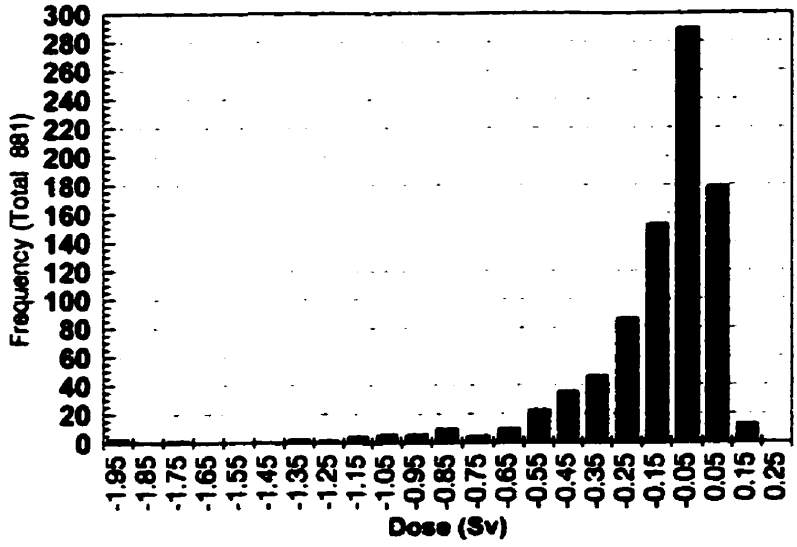

Figure 4.13a Threshold histogram produced by lines with positive slope b. Total 907 thresholds produced by lines with b>0, see Table 4.4). The smallest 26 threshold values are omitted. (lung cancer data).

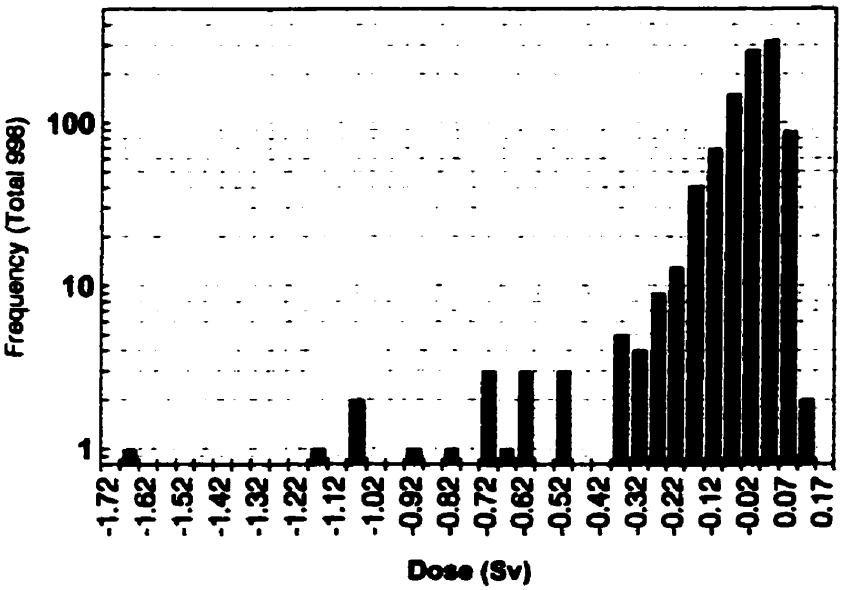

Figure 4.14 Histogram of threshold values for total solid cancers, ([Shim 92], Table 2.3 and Figure 4.4). Threshold values: -9.48 and 3.95 are omitted. The frequency axis is logarithmic.

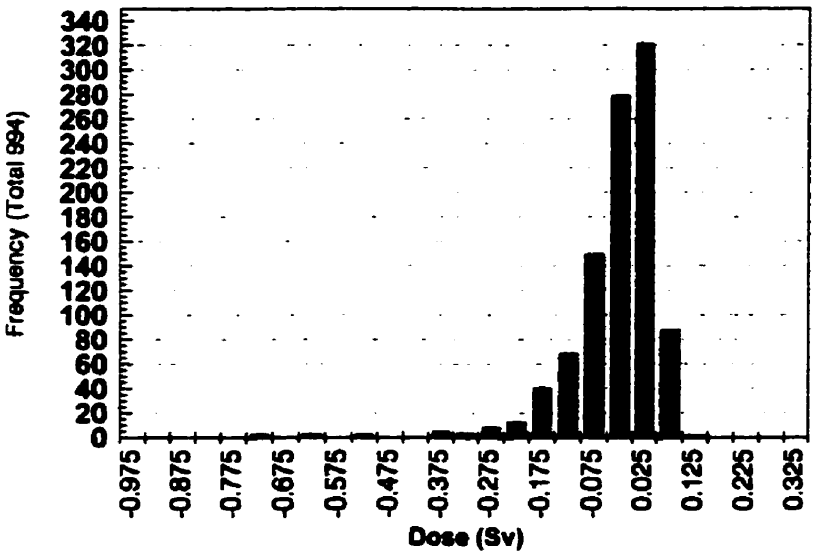

Figure 4.14a Thresholds produced by lines with positive slope b. Total 999 thresholds produced by lines with b>0, see Table 4.4). The smallest 5 threshold values are omitted, ([Shim 92] all solid cancers).

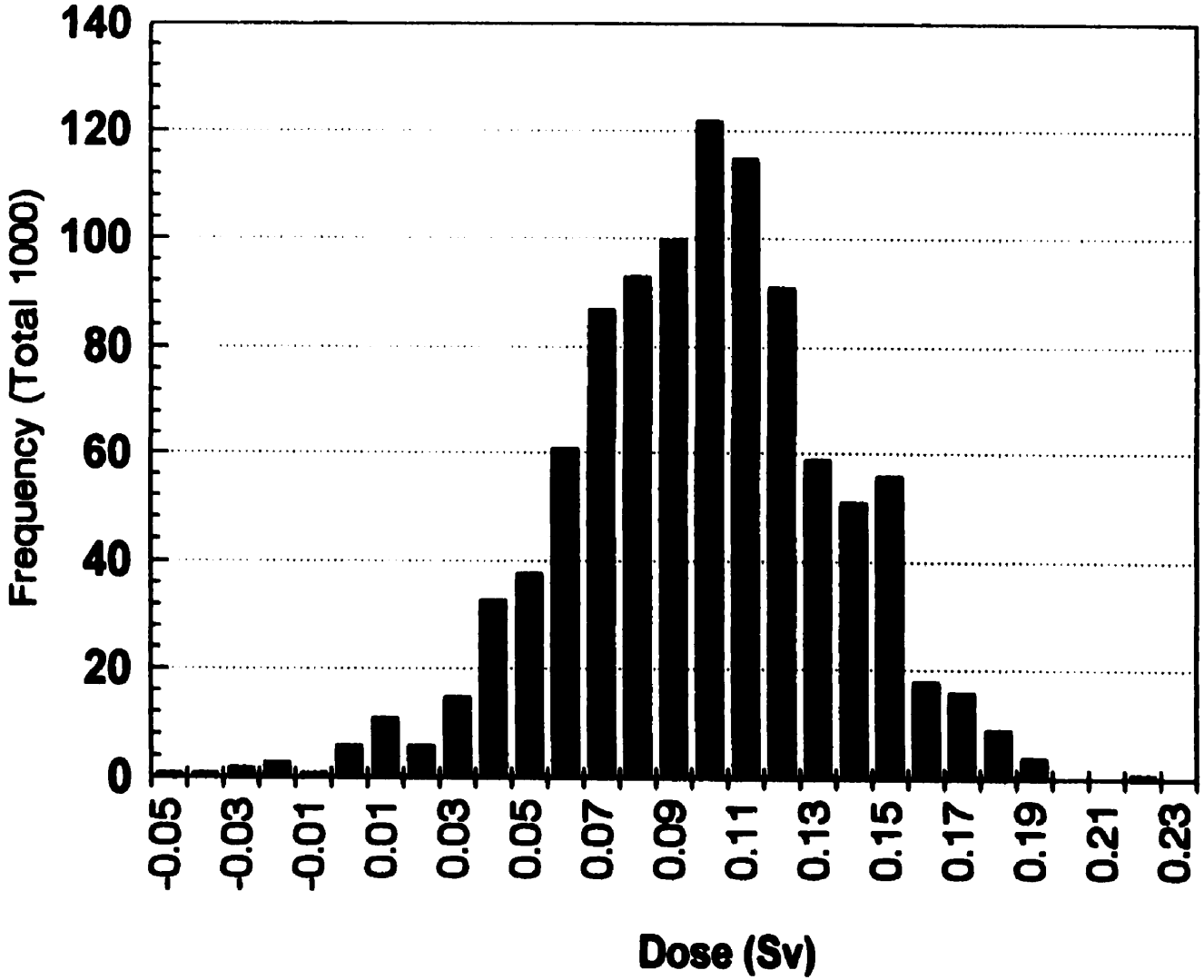

Figure 4.15 Histogram of threshold values for linear fit of leukemia data, ([Pier 96] leukemia data, Table 2.2 and Figure 4.5).

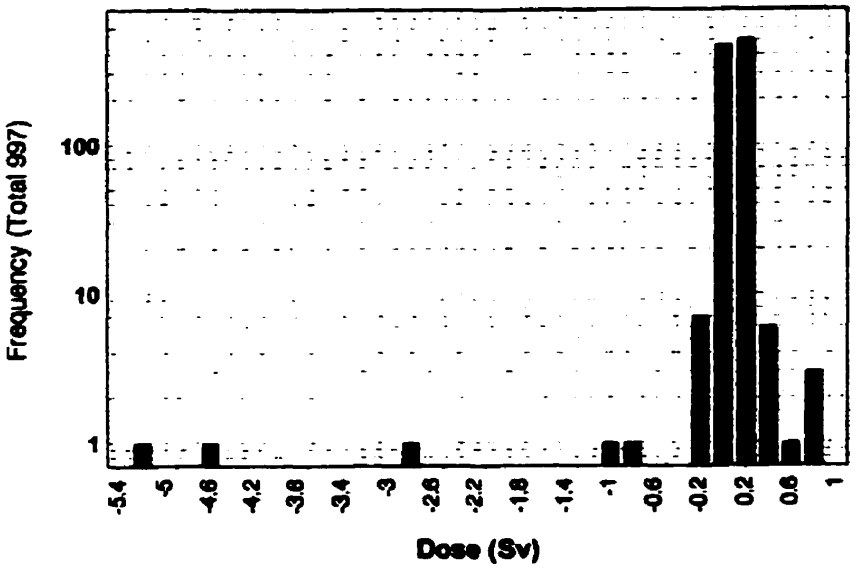

Figure 4.16 Histogram of threshold values for linear fit of leukemia data, ([Shim 92] leukemia data, Table 2.4 and Figure 4.6). Threshold values -39.9, -7.90, and 3.88 are excluded. The frequency axis is logarithmic.

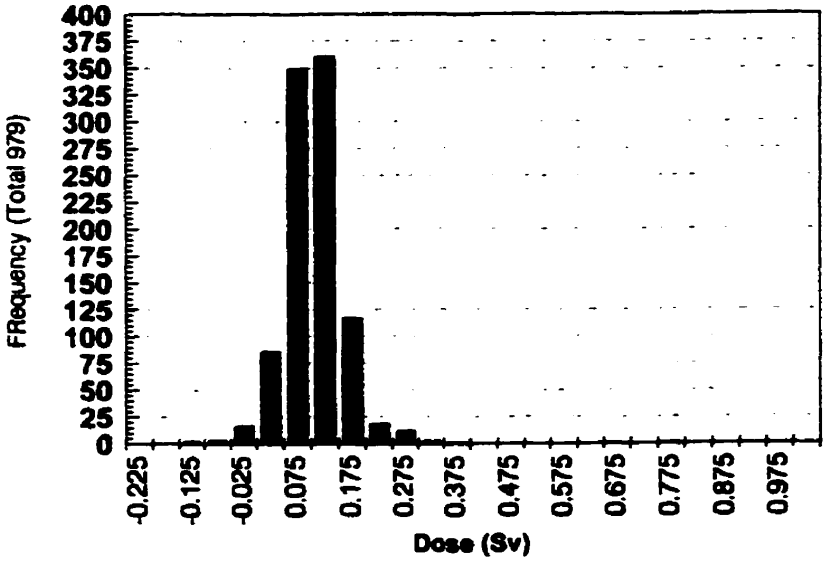

Figure 4.16a Thresholds produced by lines with positive slope b. Total 986 thresholds produced by lines with b>0, see Table 4.4). The smallest six points and the point at 3.88 are omitted, ([Shim 92] leukemia).

Table 4.3 represents the statistical analysis of estimators (intercept a, and slope b), and the analysis of threshold values (T) obtained using simulated data points. Formulae A-9 to A-13 from Appendix A-3 were used for these calculations. The third column contains the mean values, the next columns contain the median, trim mean values, their standard deviations, and their standard mean deviations (see Appendix A-3). The minimal simulated values and maximal simulated values of the estimators are in columns 9, and 10. The next two columns contain the first and third quartiles for the simulated estimators. The last column of table 4.3 contains 95% confidence intervals for median values of estimators a, b, and T for the simulated data sets (explained in Appendix A-3). Values as standard deviation of mean, minimal and maximal simulated values are not of particular physical interest, but they are included in Table 4.3 in order to get better description of simulated distributions.

Simulated data set for stomach cancer has a positive median threshold value of 0.026 Sv. The 95% confidence interval for the threshold median is in positive limits between 0.02 Sv and 0.03 Sv. Both leukemia data sets have a positive median threshold values which are equal to 0.099 Sv, and 0.102 Sv for [Pier 96] and [Shim 92] leukemia data sets, respectively. The 95% confidence intervals for the threshold median also are in positive limits between 0.096 Sv and 0.101 Sv for [Pier 96] leukemia, and between 0.099 Sv and 0.105 Sv for [Shim 92] leukemia.

The simulated data sets (lung cancer, and all solid cancers) have negative median threshold values, and negative limits of their 95 % confidence intervals.

Table 4.4 lists number of simulations with positive slope *b* for each simulated data set (see Figures 4.11-4.16). Column three contains maximum threshold values obtained by lines with positive slope. Column four lists limits obtained by subtracting fifty threshold values produced by lines with positive slope (column three) for each simulated data set. All threshold values in columns three and four are positive. Threshold values in column four (Upper 95) define the upper limit signifying that 95% of simulated data have threshold smaller than that value or negative slope (that violates the model). For [Pier 96] solid cancer and leukemia data sets values in column four give the 95% upper limits of threshold for the next simulation. This column lists upper limits of threshold uncertainties.

Table 4.3 Statistical analysis of a and b estimators of simulation, and threshold values (T) for the linear fit (solid cancer and leukemia data). The threshold values are expressed in Sv. Number (N) of simulations is one thousand.

Column #	1	2	3	4	5	6	7	8	9	10	11
Data Set	Estimator	Mean	Median	Trim. Mean	S	S_{mean}	Min.	Max.	Q_1	Q_3	95%-Sign. conf. Inter. for T
[Pier 96] Solid cancer	a	1.92	1.92	1.94	2.05	0.06	-6.44	7.90	0.62	3.27	
	b	32.5	32.4	32.5	5.43	0.17	16.1	52.3	29.0	36.1	
	T	**-0.067**	**-0.059**	**-0.064**	**0.074**	**0.002**	**-0.380**	**0.144**	**-0.112**	**-0.017**	**-0.066-0.053**
[Shim 92] Stomach cancer	a	-0.30	-0.27	-0.30	1.24	0.04	-4.69	3.53	-1.12	0.53	
	b	13.8	13.9	13.8	7.74	0.24	-10.5	39.8	8.72	19.1	
	T	**-0.017**	**0.026**	**0.011**	**0.560**	**0.018**	**-13.3**	**3.18**	**-0.040**	**0.075**	**0.021-0.032**
[Shim 92] Lung cancer	a	0.64	0.64	0.64	0.73	0.02	-1.65	3.18	0.15	1.15	
	b	5.85	5.93	5.85	4.37	0.14	-9.31	19.3	2.86	8.79	
	T	**0.355**	**-0.074**	**-0.110**	**14.8**	**0.467**	**-47.9**	**452**	**-0.225**	**0.012**	**-0.084-0.062**
[Shim 92] All solid cancers	a	0.53	0.42	0.52	2.17	0.07	-5.78	8.72	-0.91	2.04	
	b	36.9	36.7	36.8	12.8	0.40	-1.40	80.8	28.3	45.4	
	T	**-0.046**	**-0.011**	**-0.025**	**0.349**	**0.011**	**-9.48**	**3.95**	**-0.068**	**0.021**	**-0.016-0.007**

Table 4.3 Continued.

Column #	1	2	3	4	5	6	7	8	9	10	11
Data set	Estimator	Mean	Median	Trim. Mean	S	S_{mean}	Min.	Max.	Q_1	Q_3	**95%-sign Intervals for T**
[Pier 96] Leukemia	a	-0.75	-0.75	-0.75	0.32	0.01	-1.67	0.24	-0.97	-0.54	
	b	7.65	7.64	7.66	1.30	0.04	2.60	11.3	6.82	8.55	
	T	**0.097**	**0.099**	**0.098**	**0.037**	**0.001**	**-0.047**	**0.216**	**0.073**	**0.121**	**0.096-0.101**
[Shim 92] Leukemia	a	-0.56	-0.59	-0.58	0.33	0.01	-1.61	0.48	-0.81	-0.35	
	b	5.64	5.55	5.65	2.52	0.08	-2.91	14.3	3.95	7.47	
	T	**0.046**	**0.102**	**0.102**	**1.32**	**0.042**	**-39.9**	**3.88**	**0.072**	**0.133**	**0.099-0.105**

Table 4.4 Number of simulations with positive slope *b*. Column three contains maximum threshold values obtained by lines with positive slope. Column four (Upper 95%) lists limits obtained by subtracting fifty threshold values produced by lines with positive slope (column three) for each simulated data set. Total number simulations is one thousand (The difference between one thousand and number of simulations with positive slope is number of simulations with negative slope).

Simulated data set	Number simulations with $b>0$	Maximum threshold value for simulations with $b>0$ (Sv)	Upper 95% (Sv)
[Pier 96] solid	1000	0.144	0.037
[Shim 92] stomach	962	0.898	0.128
[Shim 92] lung	907	0.190	0.054
[Shim 92] all solid	999	0.103	0.061
[Pier 96] leukemia	1000	0.216	0.154
[Shim 92] leukemia	986	3.88	0.193

Chapter V

Linear-quadratic fit of leukemia data

Besides the linear fit, the leukemia data were fitted with the linear-quadratic function. This section includes the results of the standard statistical analysis and of the Monte Carlo simulation for the linear-quadratic fit as explained in sections 3.3. Dose uncertainties are not included in analysis in this section.

5.1 Results of the Least-square and Chi-square analysis

Figures 5.1, and 5.2 show the same measured values for the excess in the leukemia mortality rates as a function of received dose as in Figures 4.5 and 4.6. The leukemia data fitted with a linear model in section 4, are used in this section to fit with a linear-quadratic model. Estimates b and c were calculated by performing the weighted least squares fit. These estimates were calculated using equations 3.8 and 3.9. The variance, and covariance of the estimators were calculated as elements of error matrix 3.12. The t values were calculated using 3.13. The p values were computed using the table of t distribution in the same manner as in section 4.1. The results of these calculations are shown in Table 5.1.

The numerical values of the estimates represented in Table 5.2 were obtained by projecting the delta Chi-square ellipses onto corresponding axes in the same manner as for linear fit (section 4). Like for the linear fit in section 4.1, the projections refer to the joint confidence intervals, this time for parameters β and γ. This table includes zero equivalent points (*ZEP*) values obtained using equation 3.26. Table 5.2 also contains values of the Chi-square (equation 3.7), and of the reduced Chi-square.

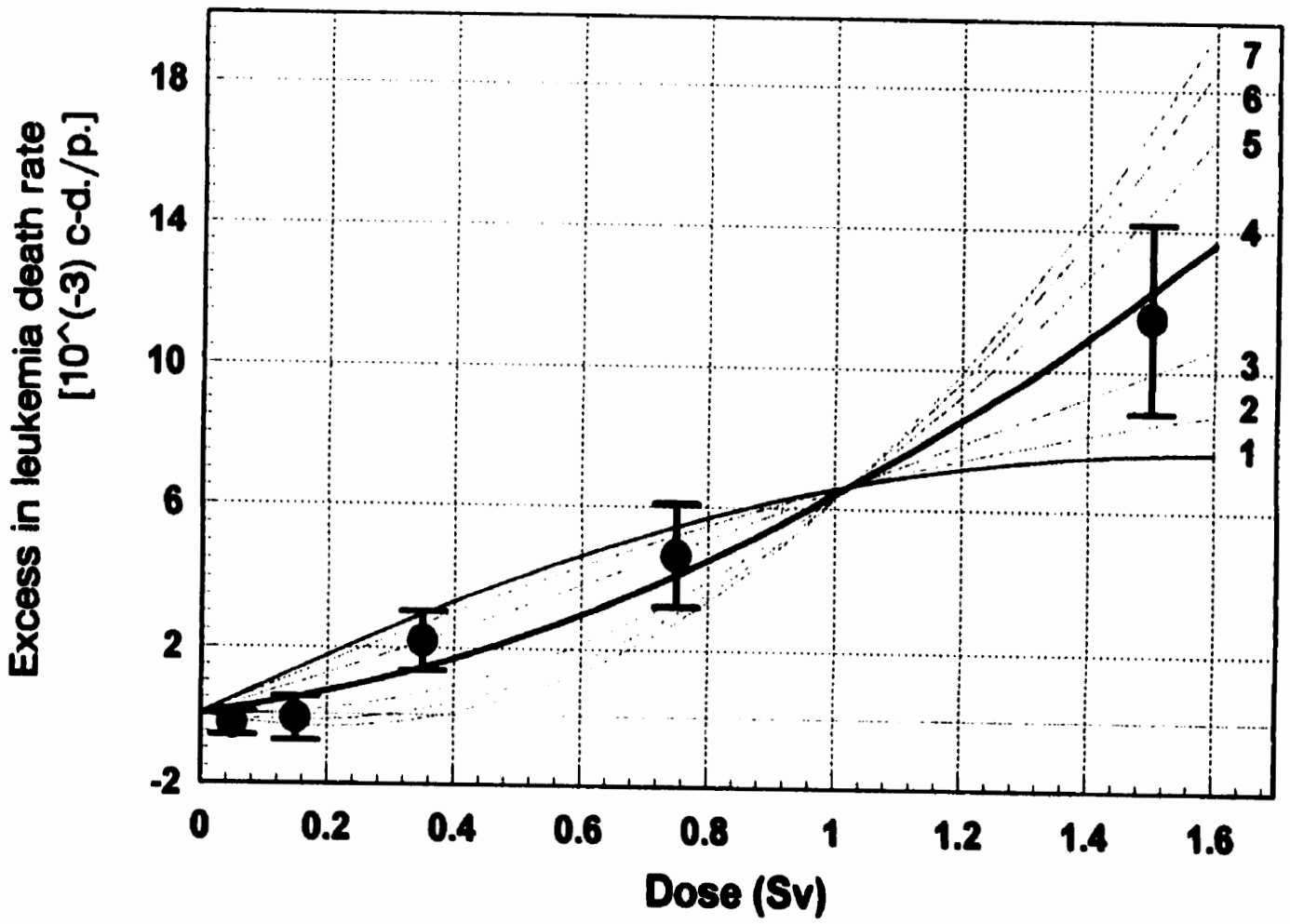

Figure 5.1 Excess in leukemia mortality rates plotted as a function of received dose. The linear-quadratic fit applied on[Pier 96] leukemia data set (Table 2.2).

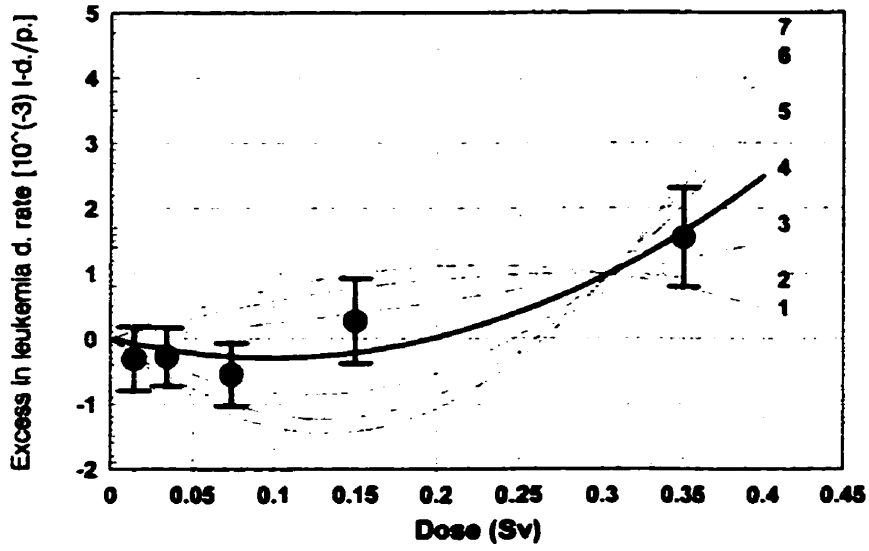

Figure 5.2 Excess in leukemia mortality rates plotted as a function of receiced dos, linear-qudratic fit. ([Shim 92] leukemia data, Table 2.4.)

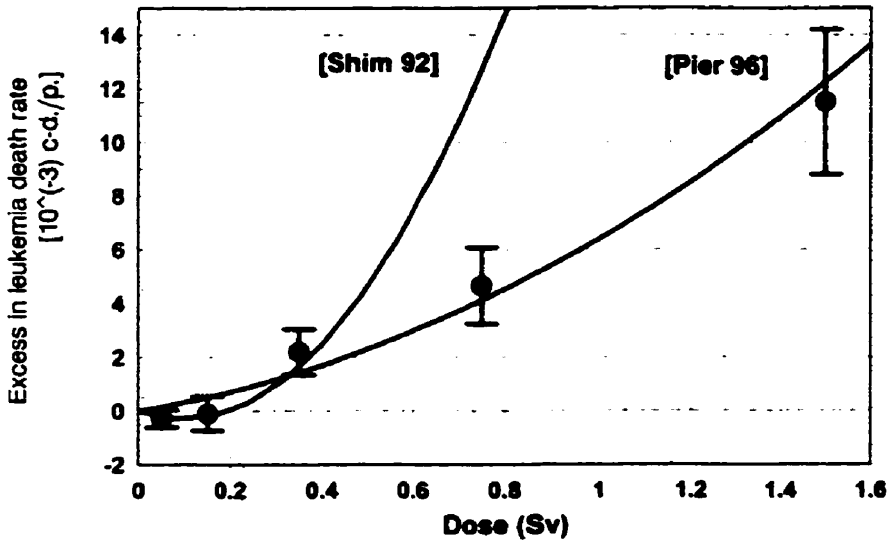

Figure 5.2a Comparison of the best fit lines (labeled 4) represented in Figures 5.1 and 5.2. It is clear that best fit parabola of [Shim 92] leukemia data set (Figure 5.2) does not fit whole range of [Pier 96] leukemia data set (Figure 5.1). Data points and their error bars are from [Pier 96] leukemia data set.

Table 5.1 Fit estimators, their uncertainties, and ZEP for the best linear-quadratic fit of leukemia data.

Leukemia data set	b	c	ZEP (Sv)	s_b	s_c	$s_{b,c}$	t_b	t_c	$P_{(b)}$	$P_{(c)}$
[Pier 96]	2.83	3.56	-0.795	2.20	2.15	-4.11	1.29	1.65	0.30	0.20
[Shim 92]	-6.02	30.6	0.197	5.25	17.3	-85.1	1.15	1.77	0.40	0.20

Figures 5.3 and 5.4 represent the delta Chi-square paraboloid and the corresponding joint confidence ellipses obtained as intersections of the delta Chi-square paraboloid to the constant values of the delta Chi-square planes. Values of these constant planes are same as for linear fit in section 4.2. Both figures were plotted using [Shim 92] leukemia data.

For the [Pier 96] leukemia data the best fit estimates b and c are positive. A linear-quadratic function with positive b, and c corresponds to the no-threshold model according to table 1.1. For the [Shim 92] leukemia data the best fit, the estimate b is negative, while the estimate c is positive.This case corresponds to the hormesis effect, according to table 1.1. Dose range of [Pier 96] leukemia data set is wider (up to 1.5 Sv) than dose range of [Shim 92] leukemia data set (up to 0.35 Sv). Figure 5.2a shows that best fit parabola of [Shim 92] leukemia data set does not fit last two points (at 0.75 Sv and 1.5 Sv) of [Pier 96] leukemia data set.

5.2 Discussion of goodness of fit

Let us consider goodness of fit for the linear-quadratic fits. The analyses of goodness of fit for this fit was done in similar manner as for the linear fit, values of t ratio (equation 3.13) and their corresponding p values, the Chi-square values (equation 3.7 for linear quadratic fit), and the reduced Chi-square for the "best" fit lines were calculated. The numerical values of these calculations for the linear fit are presented in Tables 5.1 and 5.2. Table 5.1 contains t and, p values. Table 5.2 contains the Chi-square, and the reduced Chi square values. Index $j = 4$ labels "best" fit lines in these two tables.

The two sided t test was used to test whether or not a quadratic coefficient (c) is equal to zero (the test of existence of quadratic relationship). The description of this test is included in Appendix A-2.

The reduced Chi-square values (table 5.2 $j = 4$) are 1.19 for the [Pier 96] leukemia data set, and 0.39 for the [Shim 92] leukemia data set which were fitted by linear-quadratic functions. These values are slightly closer to one than the linear fit values. However the fact that the p_c values for the quadratic term in both fits ([Pier 96] leukemia and [Shim 92] leukemia) are 0.2 together with an inspection of Figures 5.1 and 5.2 suggest that in both cases a linear model should be fit.

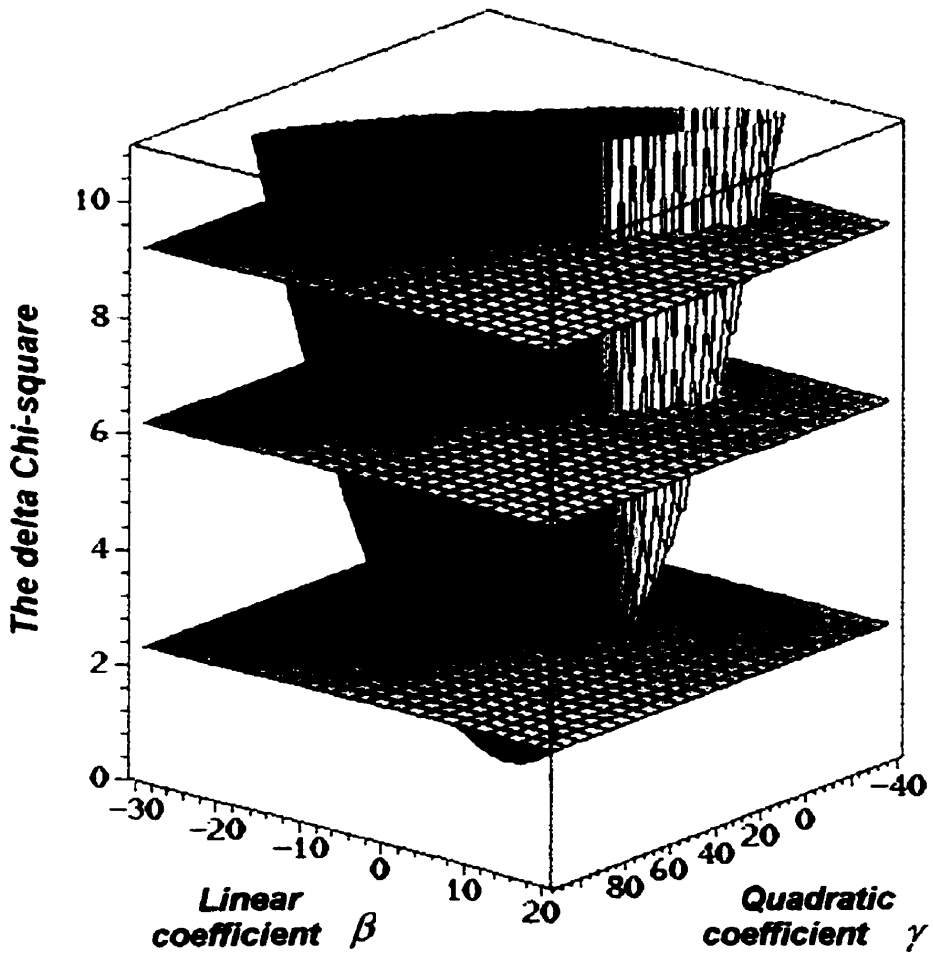

Figure 5.3 The delta Chi-square paraboloid for [Shim 92] leukemia data (Table 2.4). The data set is fit by a linear-quadratic function in the linear coefficient β, and quadratic coefficient γ. The intersections of the horizontal planes and the paraboloid correspond to 68.3%, 95%, and 99% joint confidence regions for parameters β and γ. These intersections have the form of ellipses (Figure 5.4).

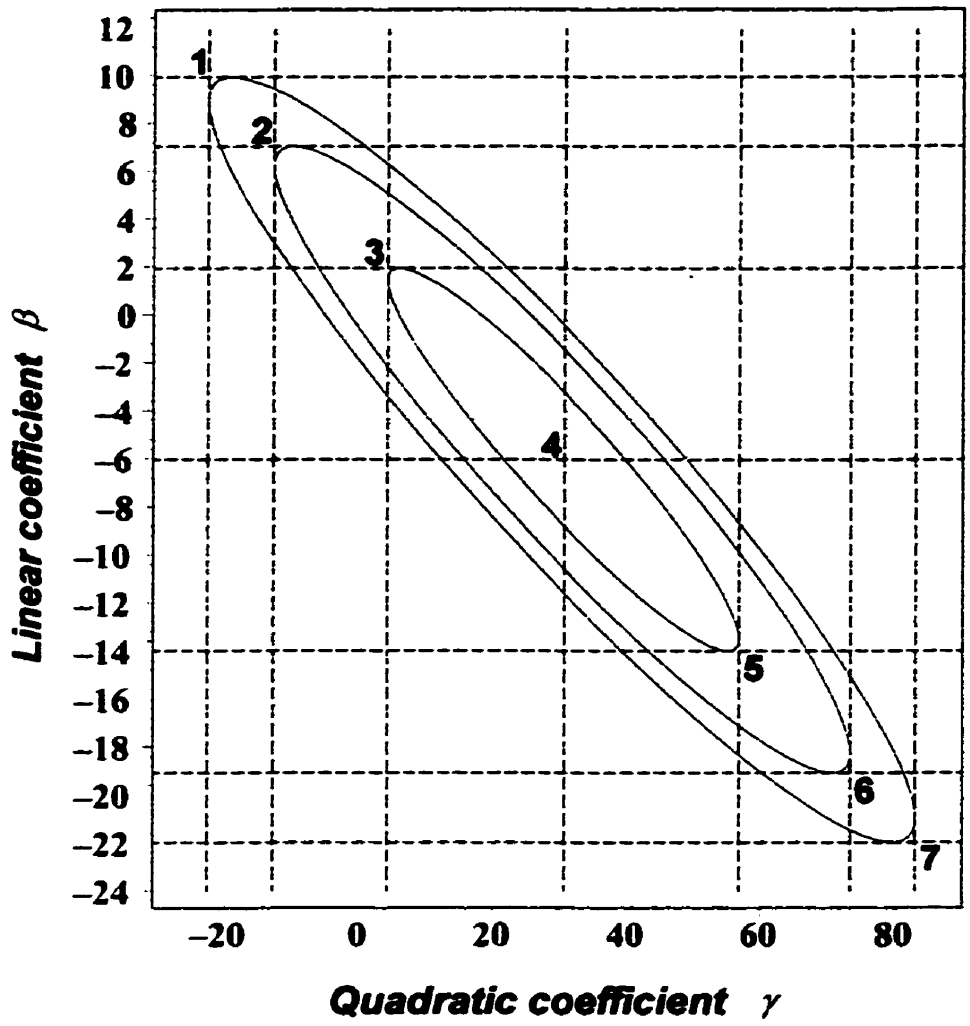

Figure 5.4 The delta Chi-square ellipses plotted as function of linear coefficient *b* and quadratic coefficient *c*. The ellipses are obtained by projecting the intersections of the delta Chi-square paraboloid (Figure 5.3) and the constant delta Chi-square planes that are equal to the 2.30, 6.17, and 9.21. The areas that are enclosed by ellipses are 68.3%, 95%, and 99% confidence regions for parameters β and γ jointly. ([Shim 92] leukemia data, Table 2.4, a linear-quadratic fit.)

Table 5.2 The estimates b and c obtained by projecting Chi-square ellipses onto corresponding axes, *ZEP*, Chi-square, and reduced Chi-square for the leukemia data sets fitted by a linear-quadratic dose response function.

Data Set (Leukemia)	Projections (j)	Linear coefficient (b_j)	Quadratic coefficient (c_j)	$ZEP\,(-\frac{b_j}{c_j})$ (Sv)	X_0^2	X_{ored}^2
[Pier 96] Table 2.2	1	9.51	-2.96	3.21	12.8	
	2	8.30	-1.78	4.66	9.73	
	3	6.17	0.312	-20.4	5.86	
	4	**2.83**	**3.56**	**-0.795**	**3.56**	**1.19**
	5	-0.51	6.83	0.074	5.86	
	6	-2.63	8.91	0.296	9.73	
	7	-3.85	10.1	0.381	12.8	
[Shim 92] Table 2.4	1	9.93	-21.9	0.453	10.4	
	2	7.03	-12.4	0.568	7.34	
	3	1.95	4.36	-0.447	3.47	
	4	**-6.02**	**30.6**	**0.197**	**1.17**	**0.39**
	5	-14.0	56.8	0.246	3.47	
	6	-19.1	73.5	0.259	7.34	
	7	-22.0	83.1	0.265	10.4	

5.3 *ZEP* analysis of Monte Carlo simulated leukemia data

The common formulae for combing uncertainties can not be used for the same reason as for the linear fit (large estimator relative uncertainties, see Table 5.1). In order to estimate confidence intervals of *ZEP*, a Monte Carlo simulation was done as described in section 3.3. The simulation was repeated one thousand times, and the simulated points were fitted with linear-quadratic functions as shown in section 3.3.3. This procedure produces one thousand parabolas which are forced through the coordinate origin. For each of the parabolas intersection with the dose axis (*ZEP* value) was calculated using equation 3.19. Estimators, the linear coefficient b and the quadratic coefficient c, for the linear-quadratic model were calculated using equations 3.8 and 3.9. Table 5.3 contains results for estimators b, c and *ZEP* values. The fourth column contains the mean values, next columns contain median, trimed mean values for the simulated estimators, their standard deviations, and their standard deviations of the mean. The minimum simulated values and the maximum simulated values of the estimators are in columns 9, and 10. The next two columns contain the first and third quartiles for the simulated estimators. The last column of Table 5.3 contains 95% confidence intervals for the median values of linear coefficient b, quadratic coefficient c, and *ZEP* values. Appendix A-3 explains how these values were obtained. The standard deviation of mean minimal and maximal simulated values are not of particular physical interest, but they are listed in order to characterize simulated distributions.

The median values of *ZEP* are -0.675 Sv for [Pier 96] leukemia simulated data, and 0.203 Sv for [Shim 92] leukemia simulated data. The confidence interval of simulated data

Table 5.3 Leukemia simulation: Statistical analysis of simulated estimators *(b, c)*, and *ZEP* values (in Sv). The linear-quadratic model. Total number of simulations (N) is one thousand.

Column #	1	2	3	4	5	6	7	8	9	10	11
Leukemia data set	Estimators	Mean	Median	Trim. Mean	s	s_{mean}	Min.	Max.	Q_1	Q_3	**95%-SI. Intervals for *ZEP***
[Pier 96] Table 2.2	b	2.78	2.81	2.79	2.18	0.07	-3.17	9.52	1.23	4.31	
	c	3.58	3.56	3.58	2.12	0.07	-2.72	11.2	2.23	4.98	
	ZEP	**-0.936**	**-0.675**	**-1.04**	**24.7**	**0.781**	**-119**	**726**	**-1.54**	**-0.192**	**-0.752--0.624**
[Shim 92] Table 2.4	b	-5.87	-5.96	-5.85	5.43	0.17	-23.7	15.4	-9.36	-2.29	
	c	30.0	29.9	30.0	17.9	0.57	-31.9	96.6	18.1	42.3	
	ZEP	**0.249**	**0.203**	**0.182**	**2.14**	**0.068**	**-4.63**	**65.4**	**0.136**	**0.245**	**0.197-0.206**

for median value (95%) of *ZEP* is between - 0.752 Sv and - 0.624 Sv for [Pier 96] simulated data set, and between 0.197 Sv and 0.206 Sv for [Shim 92] simulated data set. These different results can be explained by the influence of the dose points at 0.75 Sv, and 1.5 Sv on fit for [Pier 96] leukemia data set (Figure 5.1).

Table 5.4 contains the number of simulations in each "sign" group for each simulated data set. The "sign" refers to sign of estimators *b* and *c*, as shown in column 2. The distribution of simulations for [Pier 96] simulated data set is different from [Shim 92] simulated data set: 85.6% of simulations correspond to the no-threshold model, 9.9% of simulations correspond to the hormesis model, and 4.5% of simulations are with $b > 0$ and $c < 0$. The greatest number of simulations for the [Shim 92] data corresponds to the hormesis model (85.6% simulation). The 9.8% of simulations correspond to the no-threshold model, 4.5% of simulations has $b>0$ and $c<0$.

Table 5.4 Number of simulations classified according to signs of estimators *b*, and *c*. For each simulated data set the total number of simulations is one thousand.

Model	Sign of estimators	Number of simulations in each sign group	
		[Pier 96] leukemia	[Shim 92] leukemia
No-threshold	b>0; c>0	856	98
Hormesis	b<0; c>0	99	856
Not of interest	b>0; c<0	45	45
Not of interest	b<0; c<0	0	1
Total simulations		1000	1000

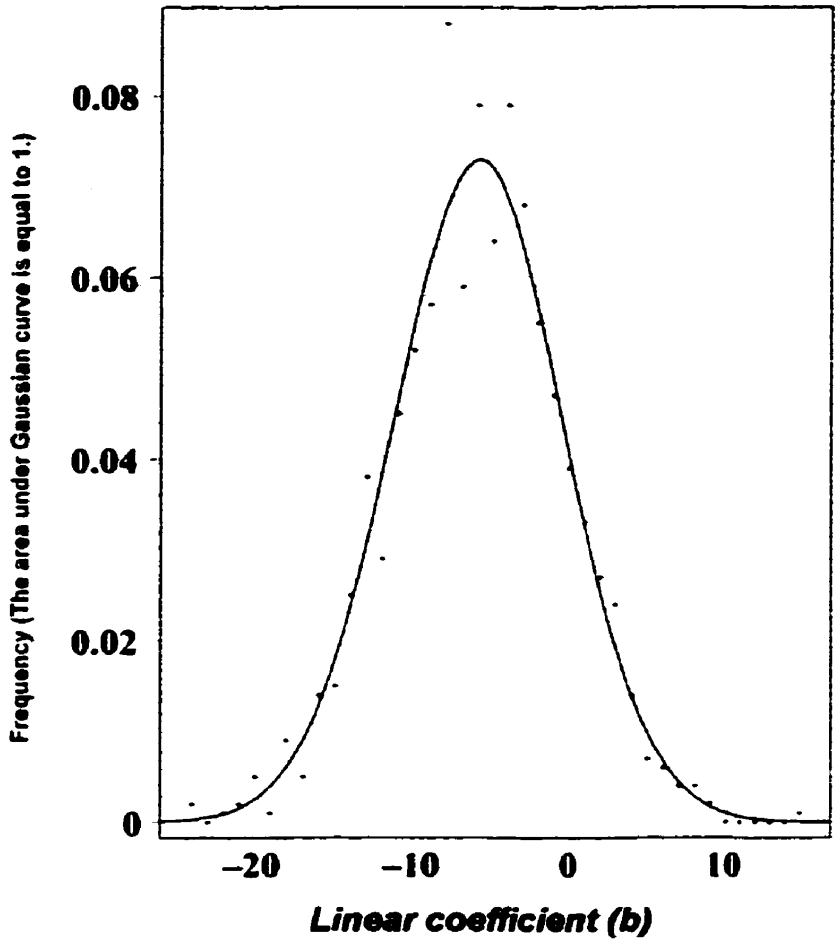

Figure 5.5 The distribution of the simulated linear coefficients *b* (for the linear-quadratic fit) and corresponding Gaussian fit. The mean value and the standard deviation of Gaussian function are listed in Table 5.3. The areas under Gaussian curve and histogram are normalized to one, ([Shim 92] leukemia data).

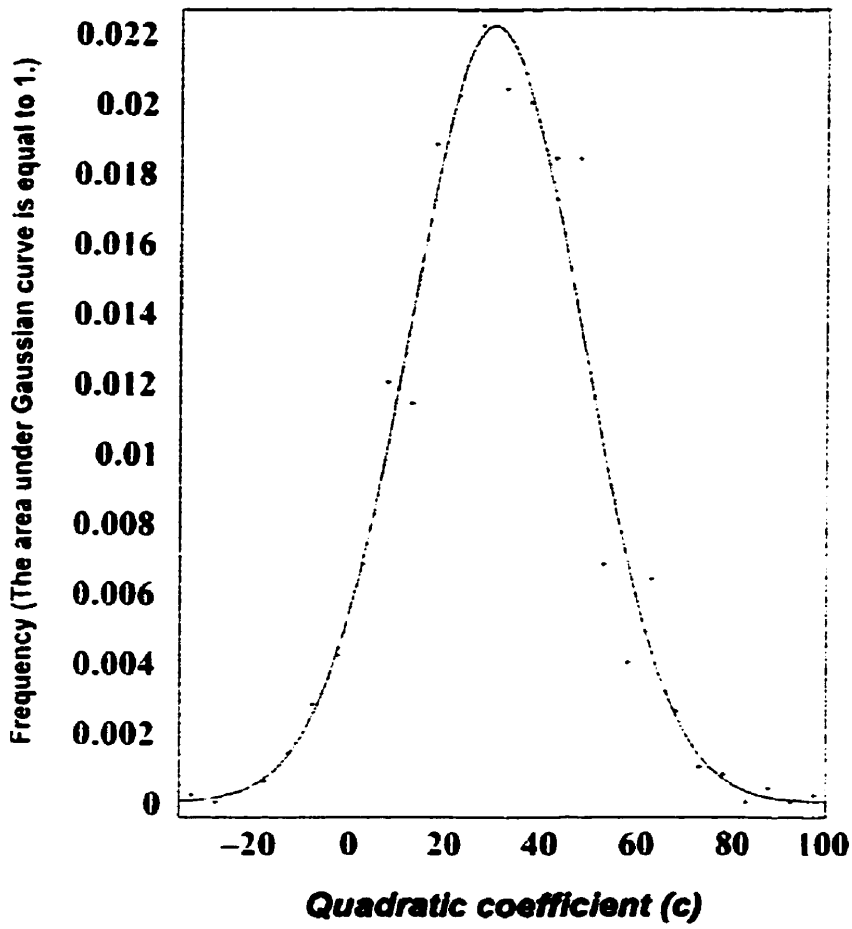

Figure 5.6 The distribution of the simulated quadratic coefficients *c* (for the linear-quadratic fit) and corresponding Gaussian fit. The mean value and the standard deviation of Gaussian function are listed in Table 5.3. The areas under Gaussian curve and histogram are normalized to one, ([Shim 92] leukemia data).

Figures 5.5 and 5.6 present distributions for estimators b, and c fitted by Gaussian functions with mean values, and standard deviations taken from Table 5.3. The area under Gaussian function is normalized to one. The [Shim 92] simulated data set was used for these two graphs. Histograms on Figures 5.7 and 5.8 represent distributions of *ZEP* values obtained by simulation. The frequency axis on both histograms is logarithmic. The histograms 5.7a were plotted in order to get a better resolution of central part of histograms 5.7. In this histogram the lowest twenty five points and highest fifty points are omitted. The histogram 5.8a includes *ZEP* values produced by parabolas with $b<0$ and $c>0$ only (hormesis, total 856 simulations). For the [Pier 96] leukemia data this kind of histogram was not plotted because total number simulations with $b<0$ and $c>0$ is only 99 (9.9% of total number of simulations , Table 5.4).

Maximum *ZEP* value produced by parabolas with $b<0$ and $c>0$ for [Shim 92] leukemia data set is 1.27 Sv, and 95% of simulations with $b<0$ and $c>0$ is up to 0.293 Sv.

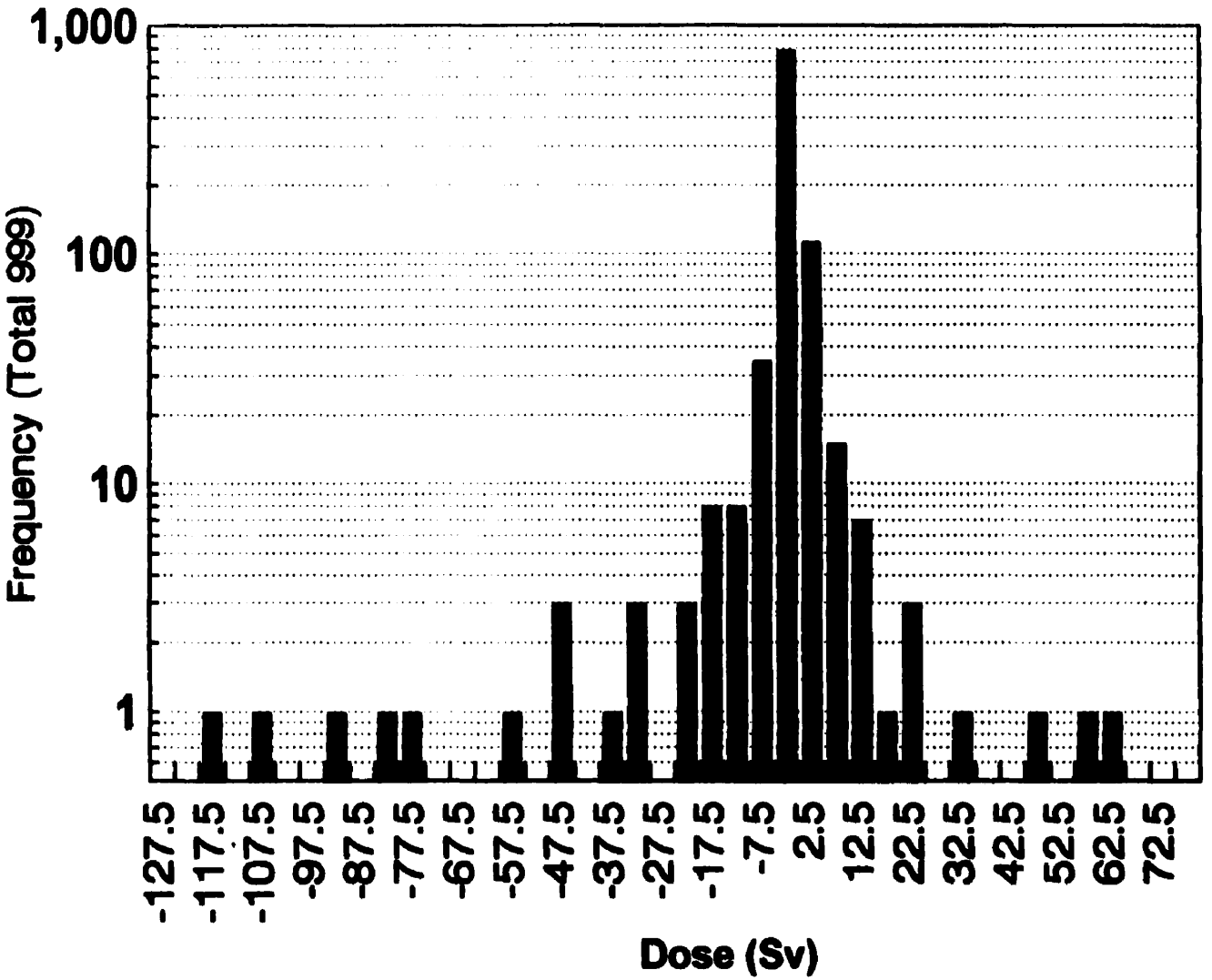

Figure 5.7 Histogram of ZEP values for the linear-quadratic fit of the [Pier 96] leukemia data, (Table 2.2). ZEP value 726 is omited.

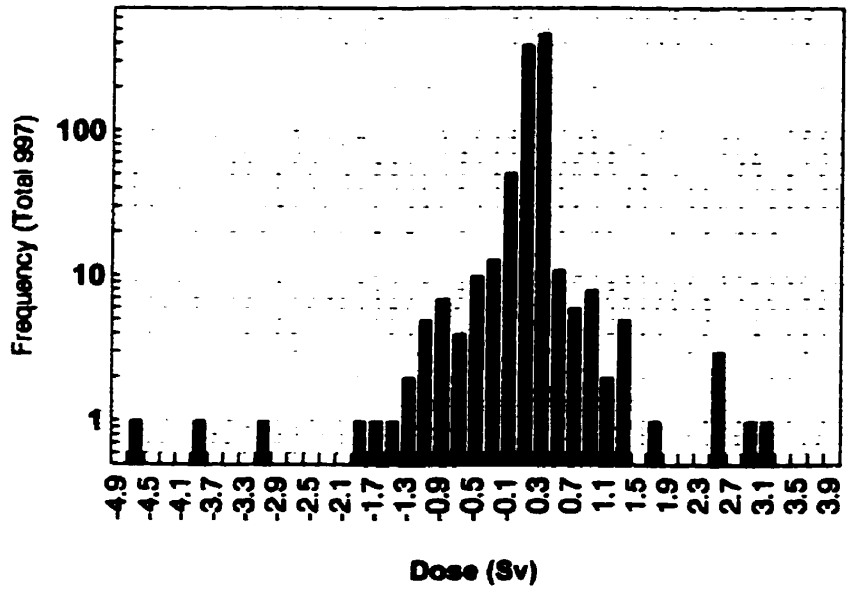

Figure 5.8 Histogram ZEP values for the linear-quadratic fit of the [Shim 92] leukemia data, (Table 2.4). ZEP values 9.05, 9.67, and 65.4 are omited. The frequency axis is logarithmic.

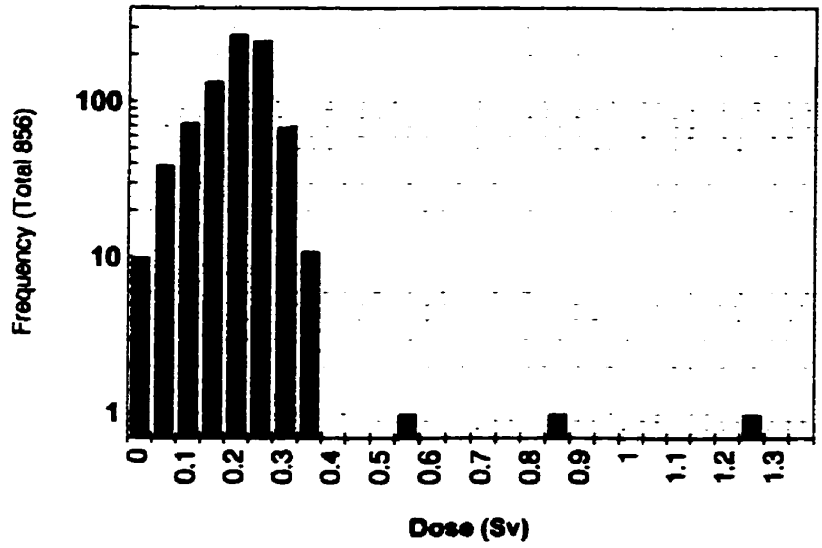

Figure 5.8a Histogram of ZEP values produced by parabolas with b<0 and c>0 (hormesis, see Table 5.4) for [Shim 92] leukemia data set.

Chapter VI

Results for the case when dose uncertainties are included

In this section dose uncertainties are included in calculations using equation 3.20. For a linear fit the Chi-square function has a form given by equation 3.21, and for a linear-quadratic fit a form is given by equation 3.22. Minimizing Chi-square in equations 3.21, and 3.22 was done by numerically solving system of equations 3.23 using a Maple program written for this purpose. Dose uncertainties, as discussed in section 2, were assumed to be 25% of the width of each dose interval. In order to distinguish estimators of fit obtained in this section from corresponding estimators obtained in sections 4 and 5, estimators of fit in this sections are labeled with index 1. To clarify, intercept and slope are labeled with a_1 and b_1. Linear and quadratic coefficients for the linear-quadratic fit are labeled with b_1 and c_1.

The above dose uncertainties are only for illustration purposes. Wider discussion of dosimetry for the atomic bomb survivors can be found in [Beir V 90, p. 190, Pier 90, Spos 91].

6.1 The linear fit

Values of intercept a_1, and slope b_1 for the linear fit were obtained by minimizing the Chi-square and are presented in Table 6.1. The threshold values in this table were calculated using equation 3.17. The last two columns of this table present the Chi-square, and its reduced Chi-square values.

By comparing the values of estimators a and a_1, and b and b_1 in tables 4.1 and 6.1, it can be easily observed that dose uncertainties cause a small changes of intercept a

and slope b, (see Table 6.1 columns Δa, Δb). Changes in thresholds ΔT are also small.

Table 6.1 Intercept a_1, slope b_1, threshold T_1, corresponding Chi-square values for the linear fit taking into account errors in doses. Table also includes differences ($\Delta a = a_1 - a$, $\Delta b = b_1 - b$, $\Delta T = T_1 - T$, and $\Delta X_o^2 = X_{o,1}^2 - X_o^2$) from corresponding values in Table 4.2. Dose errors were assumed to be equal to 25% of the dose interval width.

Set of data		a_1	Δa	b_1	Δb	T_1(Sv)	ΔT	$X_{o,1}^2$	ΔX_o^2
[Pier 96] solid cancer		1.85	-0.14	33.4	0.8	-0.1	0	1.67	-0.39
[Shim 92]	Stomach	-0.35	0	14.1	0.2	0.03	0	1.88	0
	Lung	0.64	0	5.86	0	-0.11	0	1.89	0
	All solid	0.49	-0.1	37.6	1	0	0	7.03	-0.1
[Pier 96] leukemia		-0.77	0	7.67	0	0.1	0	0.8	-0.13
[Shim 92] leukemia		-0.58	0.01	5.59	-0.1	0.104	0	0.88	0

Figures 6.1-6.6 present the excess in mortality rates plotted as a function of the received dose with included dose uncertainties. In these figures best fit lines from Figures 4.1-4.6 are also plotted (thin lines) in order to compare the differences when dose uncertainties are included (heavy lines).

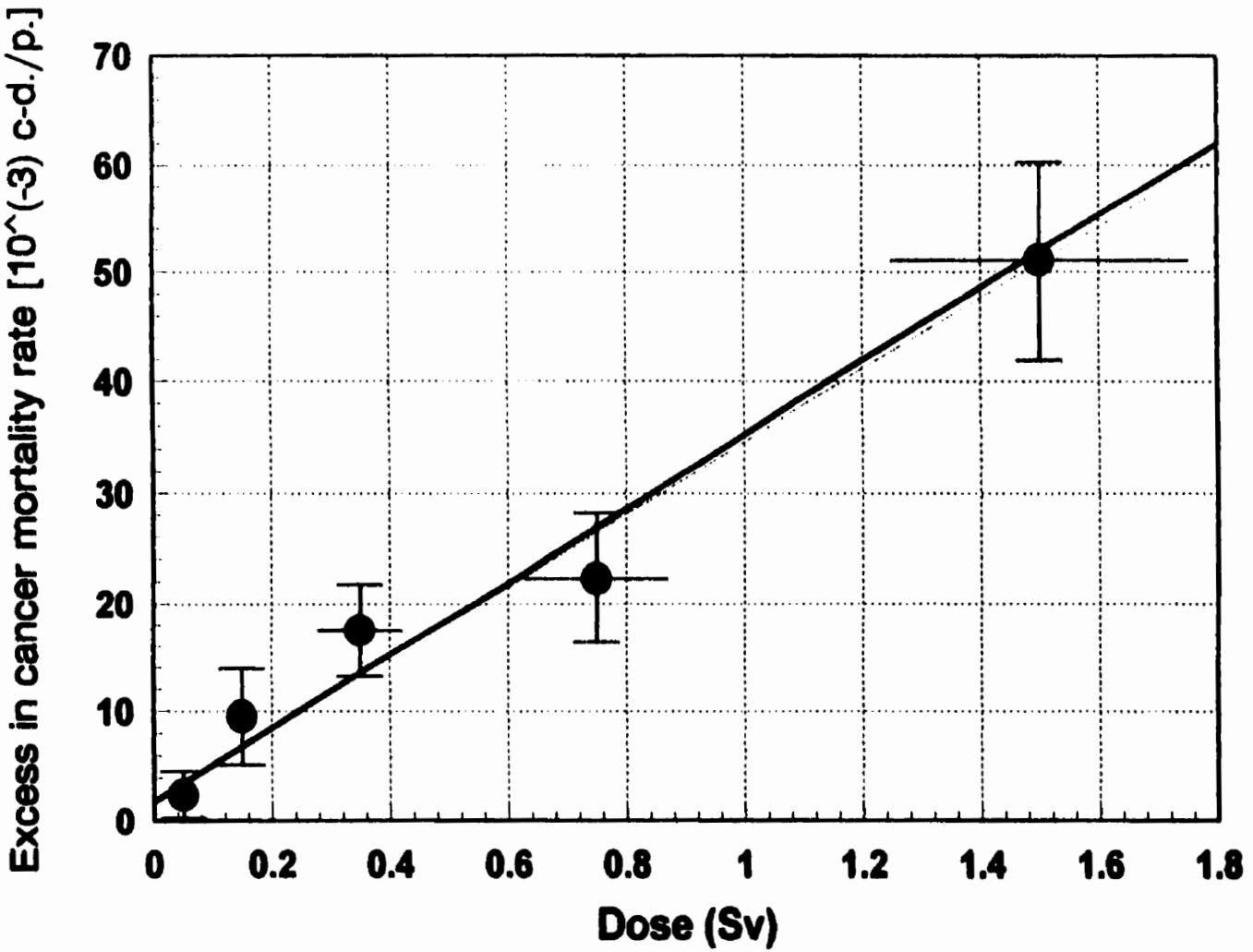

Figure 6.1 Solid cancer death excess is plotted as a function of received dose with dose uncertainties equal to +/- 25 % of the dose interval width (heavy line).
The light line is the best fit line when dose errors are not taken in account.
([Pier 96] solid cancer data, Table 2.1, Figure 4.1)

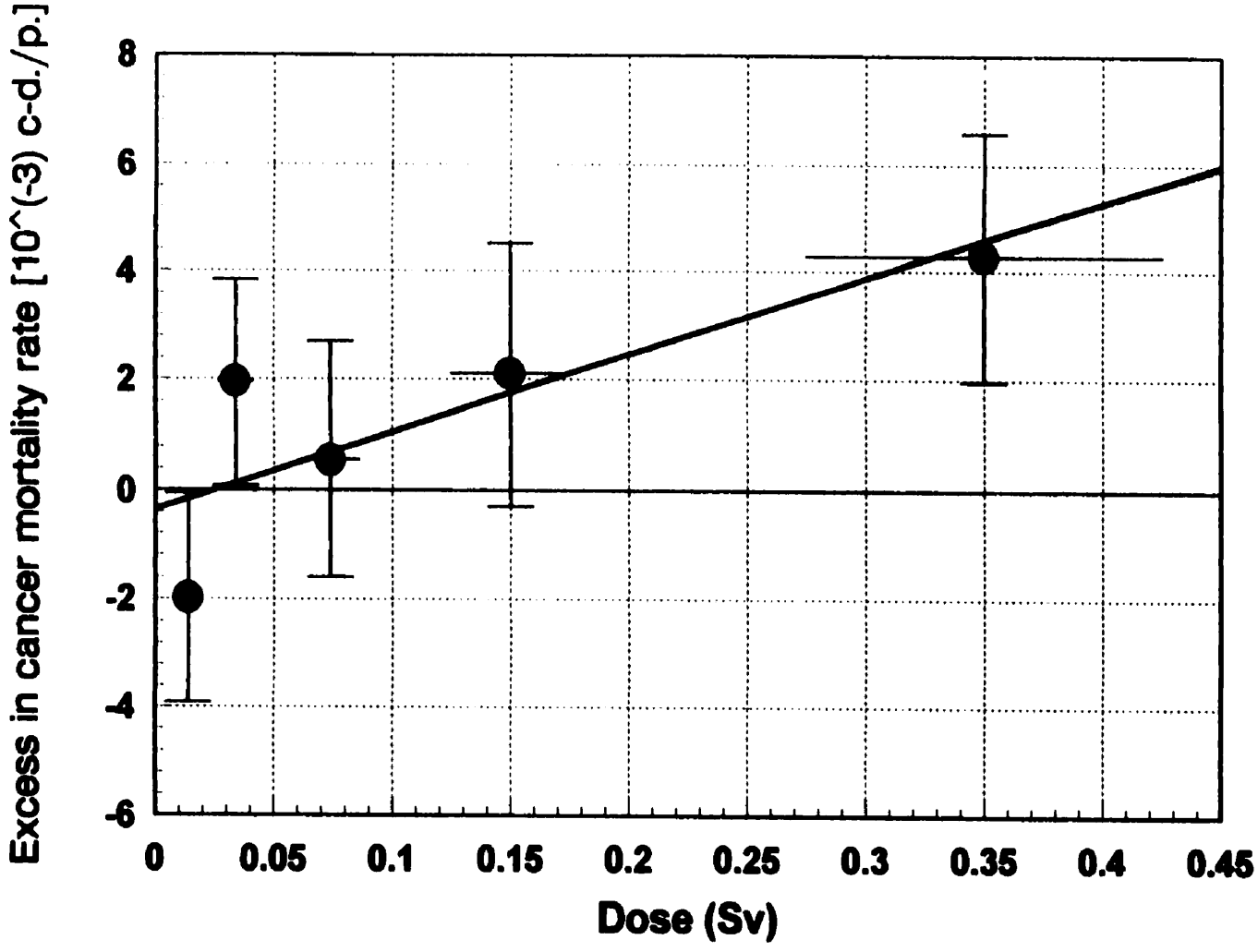

Figure 6.2 The excess in stomach cancer rate is plotted as a function of received dose with dose uncertainties equal to +/- 25 % of the dose interval width (heavy line).
The line is the best fit line when dose errors are not taken in account.
([Shim 92] stomach cancer, Table 2.3, Figure 4.2)

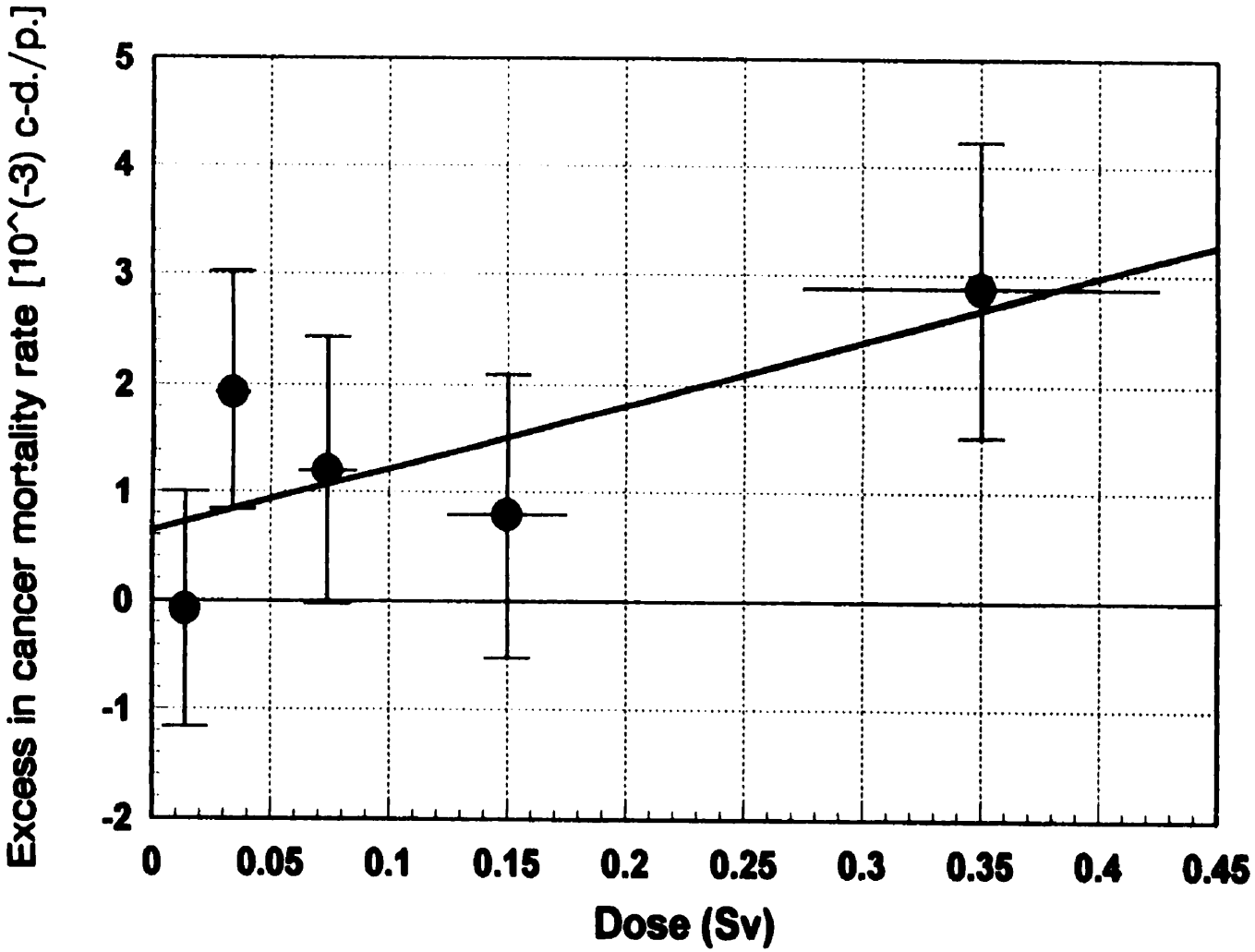

Figure 6.3 The excess in lung cancer rate is plotted as a function of received dose with dose uncertainties equal to +/- 25 % of the dose interval width (heavy line).
The light line (the best fit line when dose errors are not taken in account) is overwritten with the heavy line. ([Shim 92] lung cancer data, Table 2.3, Figure 4.3)

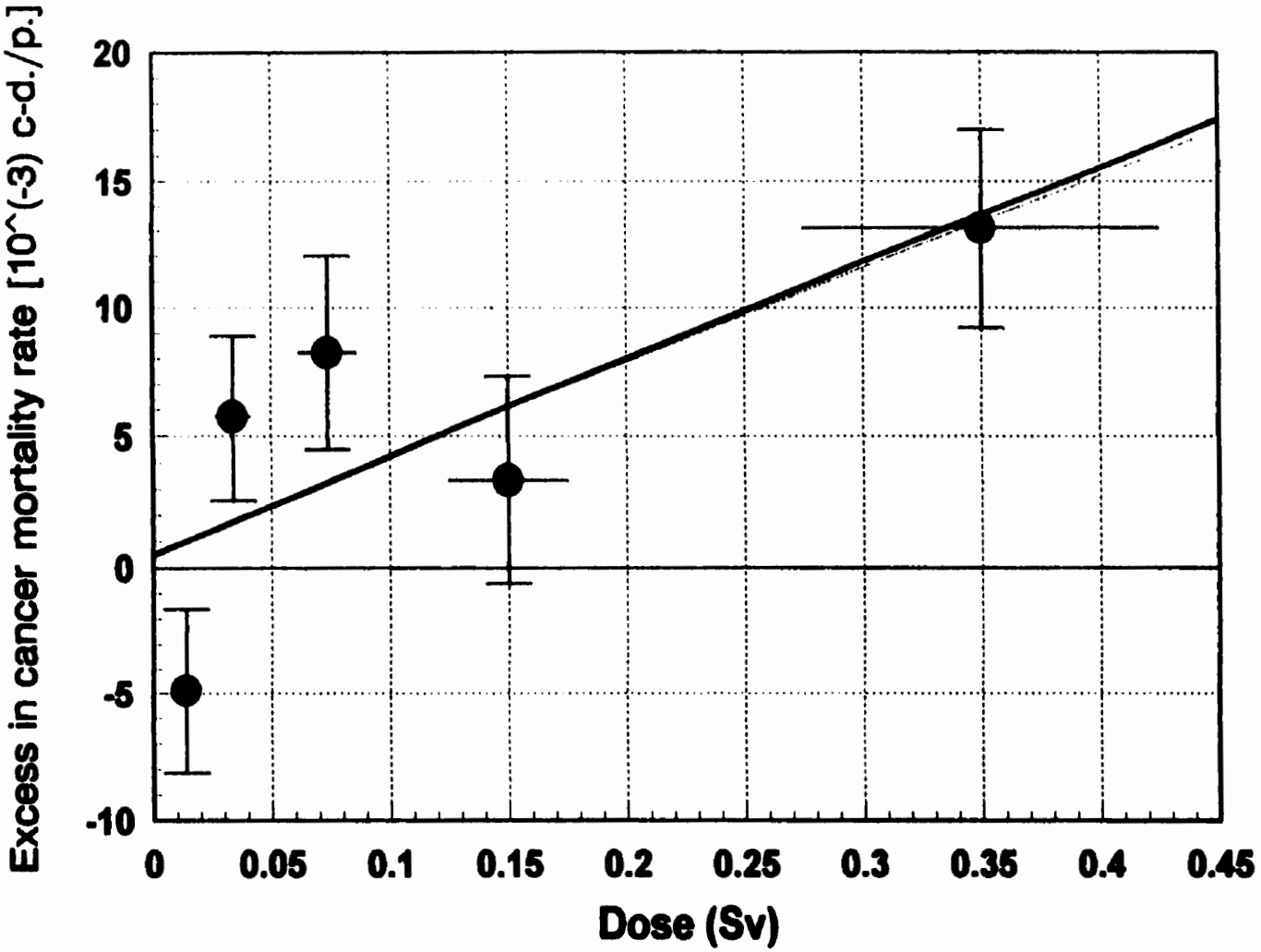

Figure 6.4 The excess in cancer mortality rate is plotted as a function of received dose with dose uncertainties equal to +/- 25 % of the dose interval width (heavy line).
The light line is the best fit line when dose errors are not taken in account.
([Shim 92] all solid cancer data, Table 2.3, Figure 4.4)

6.2 The linear quadratic fit for leukemia data

Results of the linear-quadratic fit with dose uncertainties taken into account are presented in Table 6.2. The dose uncertainties are the same as the ones presented in section 6.1. The ZEP_1 values were calculated using equation 3.19.

By comparing the values of estimators b and b_1, and c and c_1 in Tables 5.2 and 6.2 it can be easily observed that dose uncertainties cause an increase of values for estimator c, and decrease of estimator b. These changes are greater than for linear fit. This can be seen in Figures 6.5 and 6.6.

Table 6.2 Estimators b_1 and c_1, ZEP_1, corresponding Chi-square values for the linear-quadratic fit taking into account errors in doses. Table includes and differences ($\Delta b = b_1 - b$, $\Delta c = c_1 - c$, $\Delta ZEP = ZEP_1 - ZEP$, and $\Delta X_o^2 = X_{o,1}^2 - X_o^2$) from corresponding values in Table 5.2. Dose errors were assumed to be equal to 25% of the dose interval width.

Leukemia data set	b_1	Δb	c_1	Δc	ZEP_1 (Sv)	ΔZEP	$X_{o,1}^2$	ΔX_o^2
[Pier 96]	2.15	-0.68	4.67	1.11	-0.46	0.335	3.18	-0.38
[Shim 92]	-6.6	-0.23	34.6	4	0.19	0	1.13	-0.04

Figures 6.5-6.6 present the excess in leukemia mortality rates plotted as a function of received dose (heavy lines) with the dose uncertainties for both leukemia data sets. The linear and the linear-quadratic fits are presented. In these figures corresponding best fits obtained in section 5 (dose uncertainties not included) are also shown (light lines).

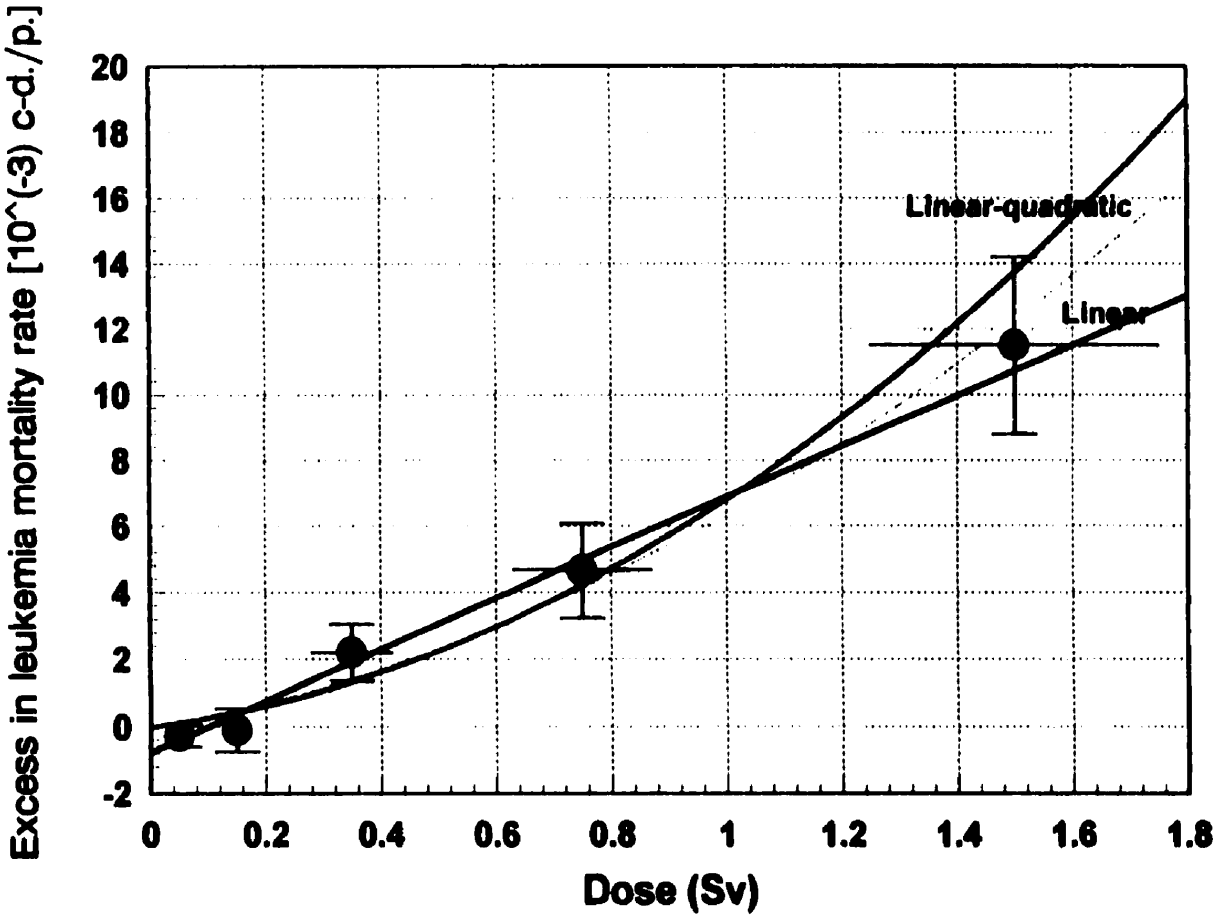

Figure 6.5 The excess in leukemia mortality rate is plotted as a function of received dose with the dose uncertainties equal to +/- 25 % of the dose interval width (heavy lines). The light lines are the best fit lines when dose errors are not taken in account. ([Pier 96] leukemia data, Ttable 2.2, Figures 4.5 and 5.1)

Linear, and linear-quadratic fits are shown.

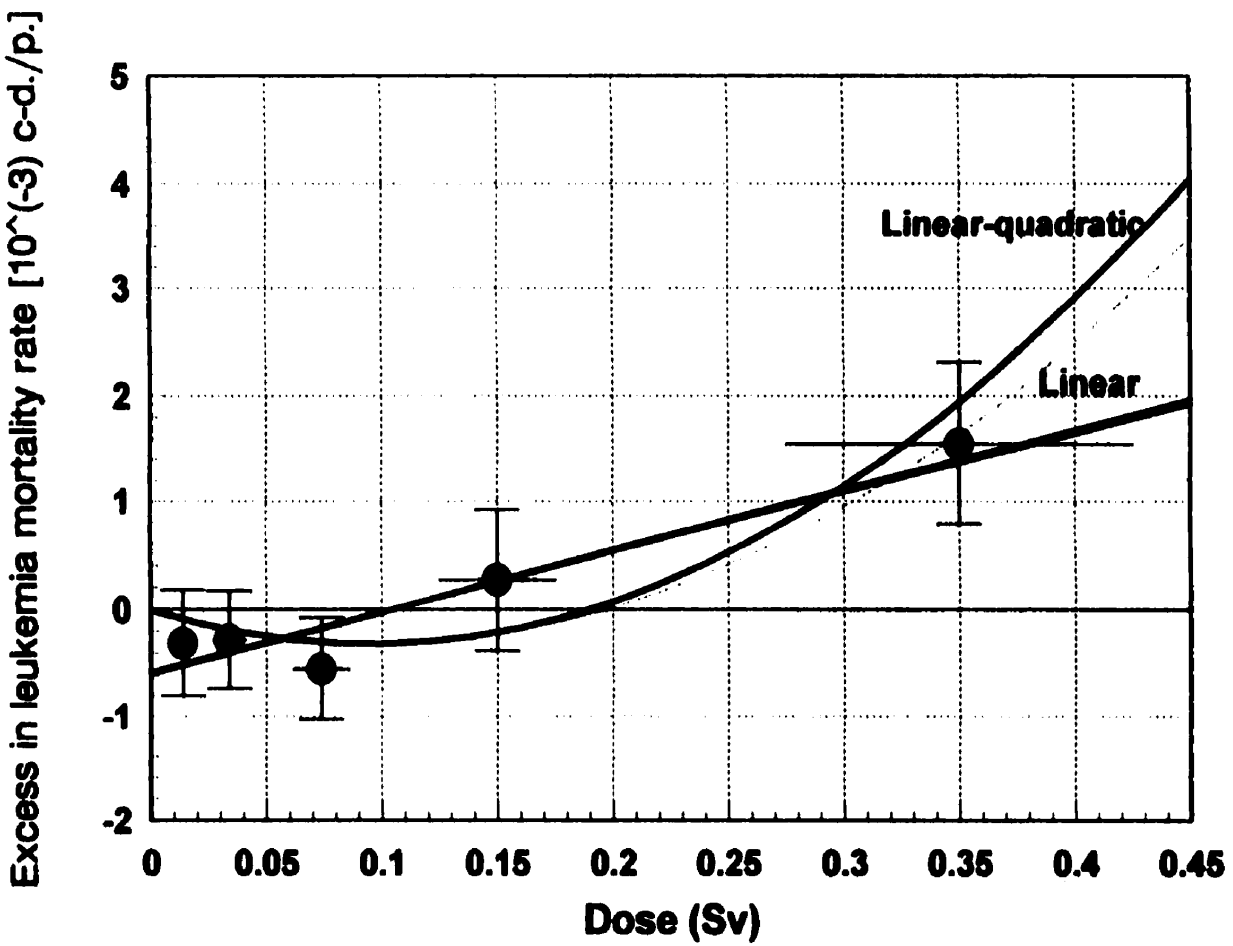

Figure 6.6 The excess in leukemia mortality rate is plotted as a function of received dose with the dose uncertainties equal to +/- 25 % of the dose interval width (heavy lines). The light lines are best fit lines when dose errors are not taken in account. ([Shim 92] leukemia data, Table 2.4, Figures 4.6 and 5.2)

Linear, and linear-quadratic fits are shown.

Chapter VII

Analysis of cancer and leukemia mortality rates

As we have stated in Chapter II, subtraction of cancer (leukemia) mortality rates of control group in expression for excess mortality rates (equation 2.1) introduces correlation among excess mortality rates. This correlation problem can be avoided by fitting mortality rates instead of excess mortality rates. In this chapter, mortality rates as a function of received dose are fitted with linear, "threshold" and quadratic form of dose response function. Comparison of "threshold" and quadratic fits with the linear fit was performed using F test.

7.1 Models for cancer (leukemia) mortality rates

Mortality rates and their uncertainties were calculated using equations 1.1 and 2.2 respectively. Mortality rates (MR) as a function of received dose (d) were first fitted with linear function in the standard manner, i.e.

$$MR = p_1 + p_2 \cdot d \,. \tag{7.1}$$

Second, the following functional form was used to describe threshold effect

$$MR = p_1 + \frac{p_2}{2} \cdot (d - d_0) + \frac{p_2}{2} \cdot |d - d_0| \,. \tag{7.2}$$

The parameter d_0 has a meaning of threshold. For the d_0 equal to zero, equation 7.2 becomes ordinary linear fit 7.1. For dose values greater than threshold value d_0 function 7.2 is a straight line with slope equal to p_2 and y-intercept equal to $p_1 - p_2 \cdot d_0$. For dose values smaller than threshold value d_0, the second and third terms in equation 7.2

are cancelled and the dose response function 7.2 is a constant equal to p_1.

The third functional form is ordinary quadratic fit in the form

$$MR = p_1 + p_2 \cdot d + p_3 \cdot d^2 \quad . \tag{7.3}$$

The threshold and the quadratic fits were compered to linear fit using F test of additional term.

Test of additional term

The ratio of two independent reduced Chi-square distribution $X^2_{\nu_1}$ and $X^2_{\nu_2}$ is distributed according F distribution [Bevi 92, p. 205].

$$P_f(f;\nu_1,\nu_2) = \frac{X^2_{\nu_1}}{X^2_{\nu_2}} \tag{7.4}$$

The F distribution can be used to measure an improvement of fit with m parameters caused by fit with additional parameter $(m+1)$ [Bevi 92, p. 209]. Let us illustrate this using the linear and threshold models. The threshold model has one additional parameter (d_0) more than the linear model. The test of additional term, tests whether or not that additional parameter (term) (d_0) significantly improves fit. This test is define using F_X ratio.

For the linear and threshold fits F_X ratio is

$$F_X = \frac{X^2_{lin} - X^2_{thr}}{X^2_{thr}/(6-2-1)} = \frac{1}{3}\cdot\left(\frac{X^2_{lin}}{X^2_{thr}} - 1\right). \tag{7.5}$$

The decision rule for $p = 0.05$ is:

> If $F_X < F'_{\alpha}$, the threshold model is not a statistically significant improvment over the linear model.
> If $F_X > F'_{\alpha}$, the threshold model fits significantly better than the linear model, where critical value
> $F'_{\alpha} = F(\nu_1 = 1, \nu_2 = 3) = 10.1$, [Bevi 92, Table C5, p. 262].

This form of statistical test is also valid for comparison the linear and quadratic models because quadratic fit (7.3) has one additional parameter more than linear model.

7.2 Results of analysis

Mortality rates were fitted in the forms 7.1-7.3 using the "Origin" computer program from Microcal. Results of fit are presented in Figures 7.1-7.6. Figures 7.1, 7.3, 4.4 contain linear (equation 7.1) and quadratic (equation 7.3) fits. For data sets (the [Pier 96] solid cancer, [Shim 92] lung cancer and [Shim 92] all solid cancers) presented in these figures threshold fit (equation 7.2) is identical to the linear fit (i.e. parameter d_0 =0). The [Shim92] stomach cancer and both leukemia data sets have threshold parameter d_0 different from zero (Figures 7.2, 7.5 and 7.6).

During minimization, the global minimum of Chi-square function has to be determined and other local minima must be avoided. For a poor choice of initial estimates of p_1, p_2, d_0, the non-linear fitting algorithm may find a local minimum that is not the global minimum. For example for the [Shim 92] leukemia data, the non-linear fitting algorithm converged to the local minimum at $p_1 = 0.00167$, $p_2 = 0.00923$, $d_0 = 0.16831$ (shown in Figure 7.7), which differs from the global minimum shown in Figure 7.6.

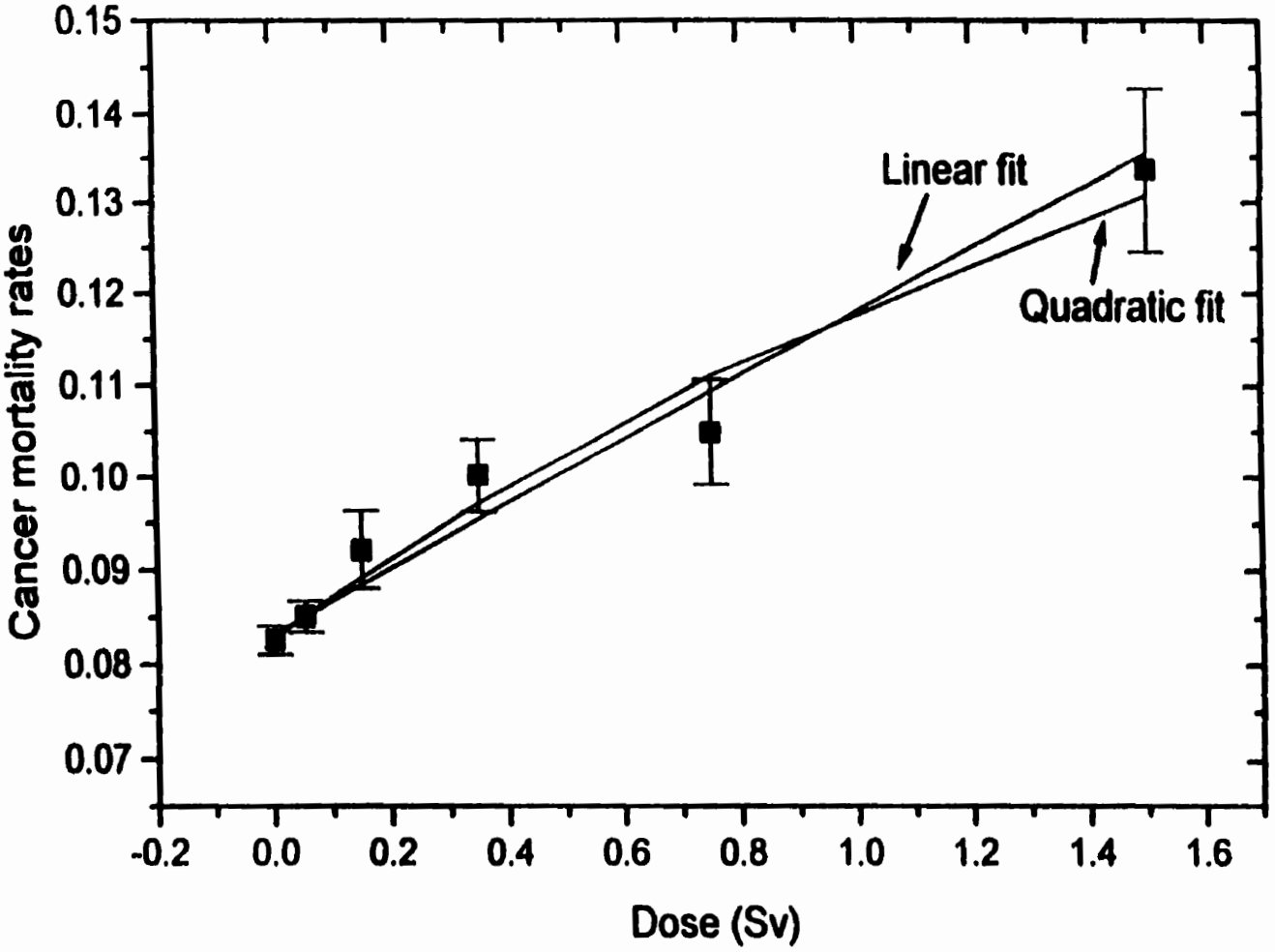

Figure 7.1 Linear and quadratic fits for [Pier 96] solid cancer mortality rates. Threshold fit is identical to linear fit (d_0=0, see Tables 7.1 and 7.2).

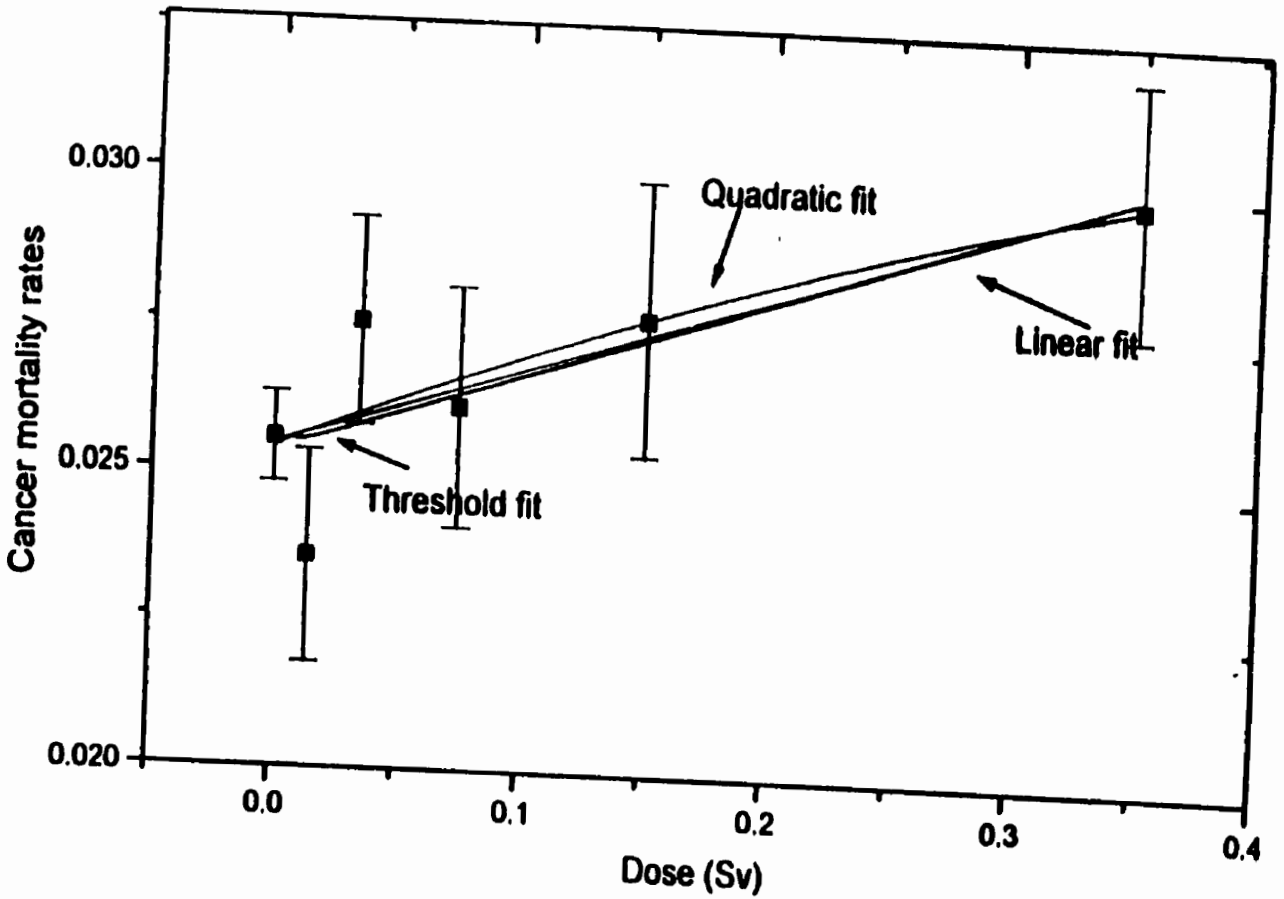

Figure 7.2 Cancer mortality rates plotted as a function of received dose. The [Shim 92] stomach cancer data set is shown. The threshold $d_0 = 0.014$ Sv (see Table 7.2).

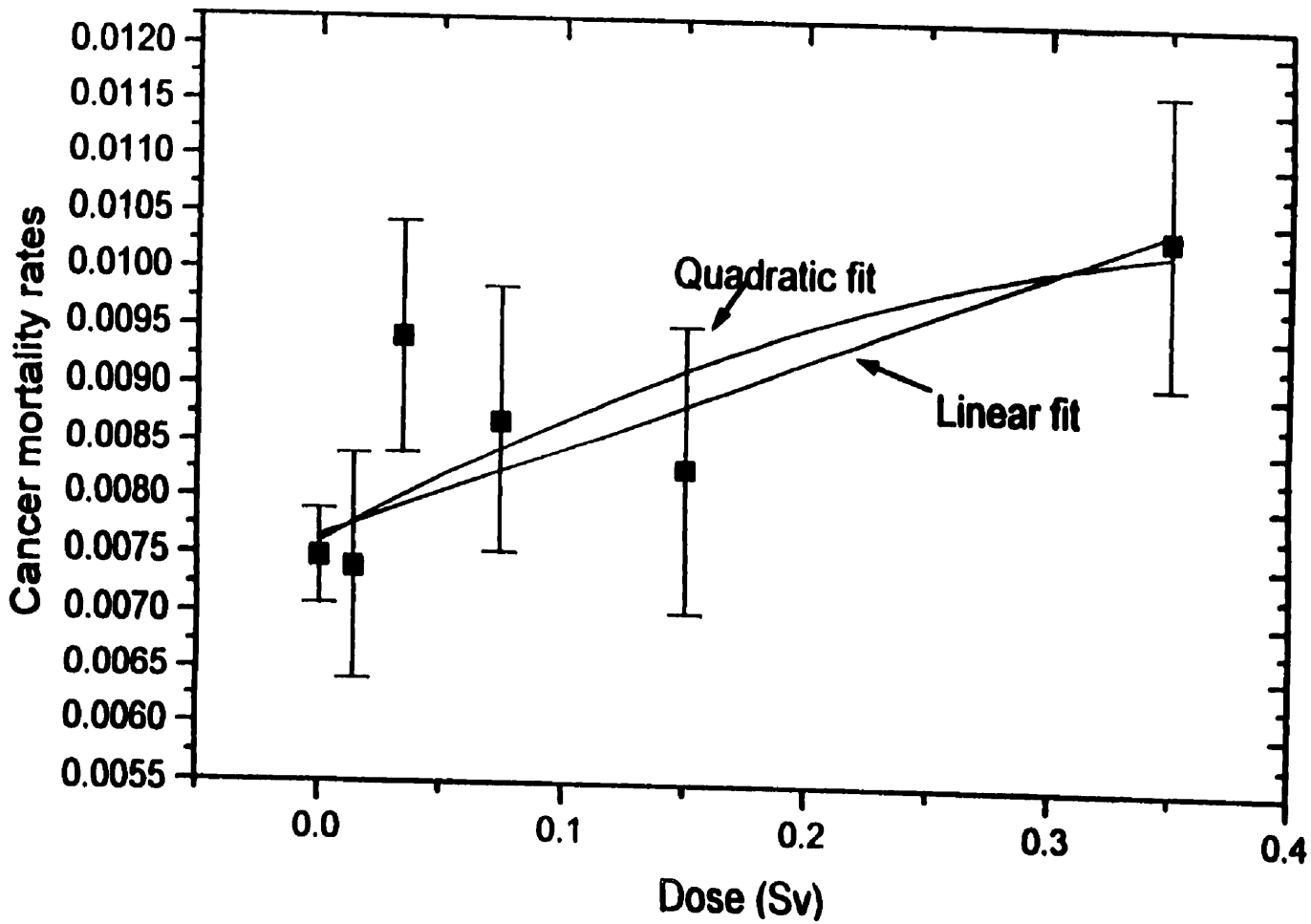

Figure 7.3 Linear and quadratic fits for [Shim 92] lung cancer mortality rates. Threshold fit is identical to linear fit (d_0=0, see Tables 7.1 and 7.2).

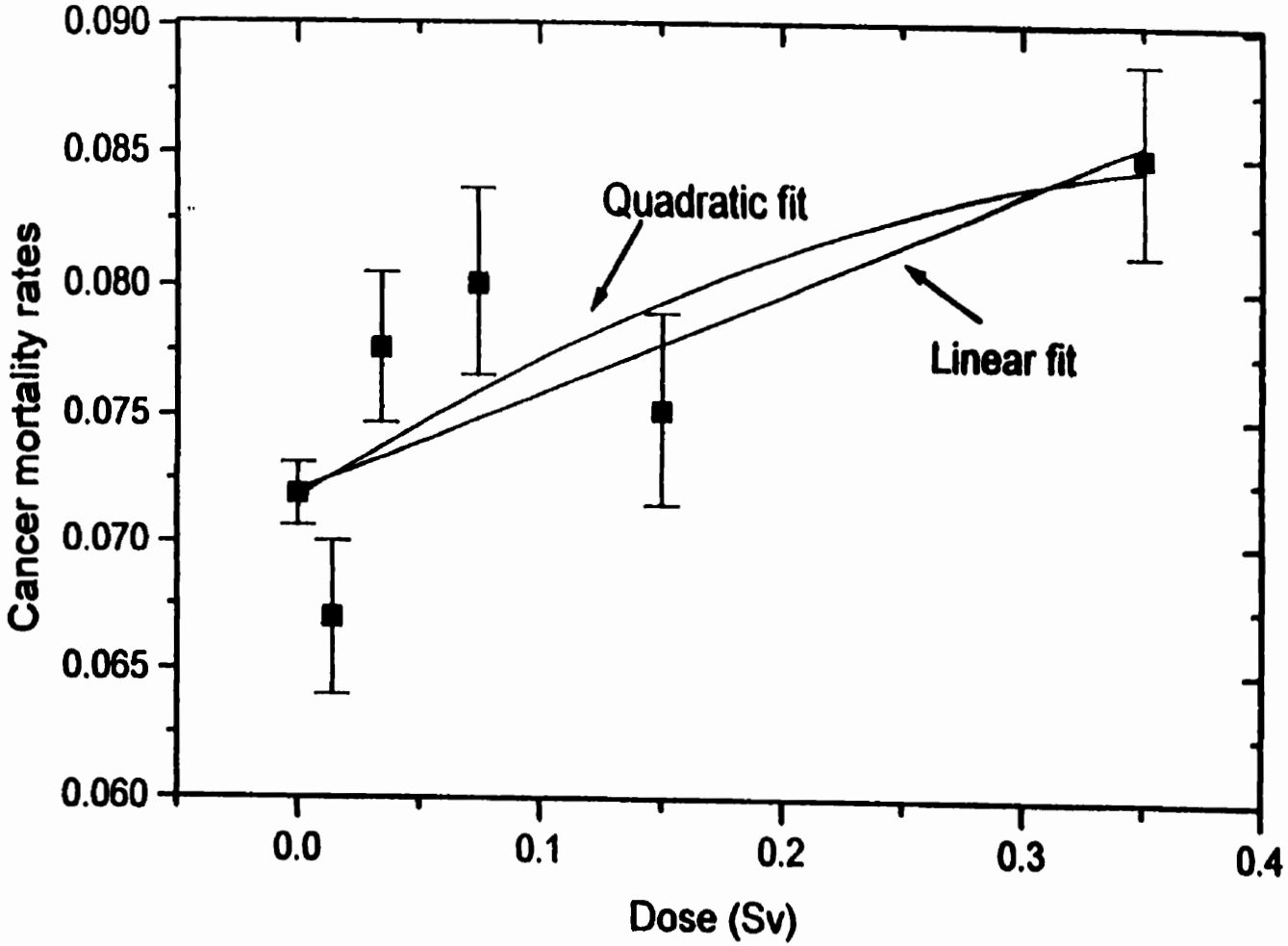

Figure 7.4 Linear and quadratic fits for [Shim 92]all solid cancer mortality rates. Threshold fit is identical to linear fit (d_0=0, see Tables 7.1 and 7.2).

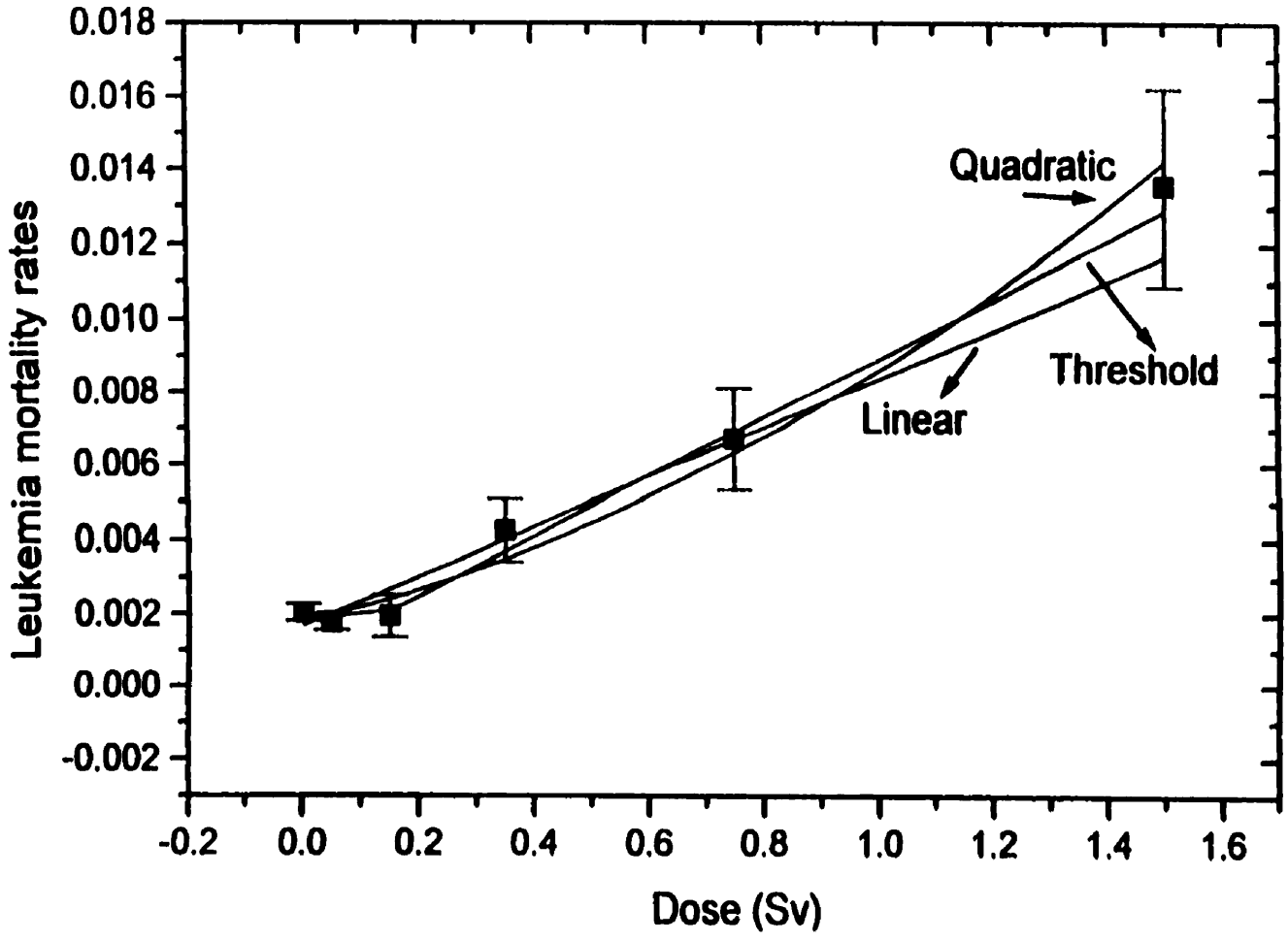

Figure 7.5 Leukemia mortality rates plotted as a function of received dose.
The [Pier 96] leukemia data set is shown.
The threshold $d_0 = 0.127$ Sv (see Table 7.2).

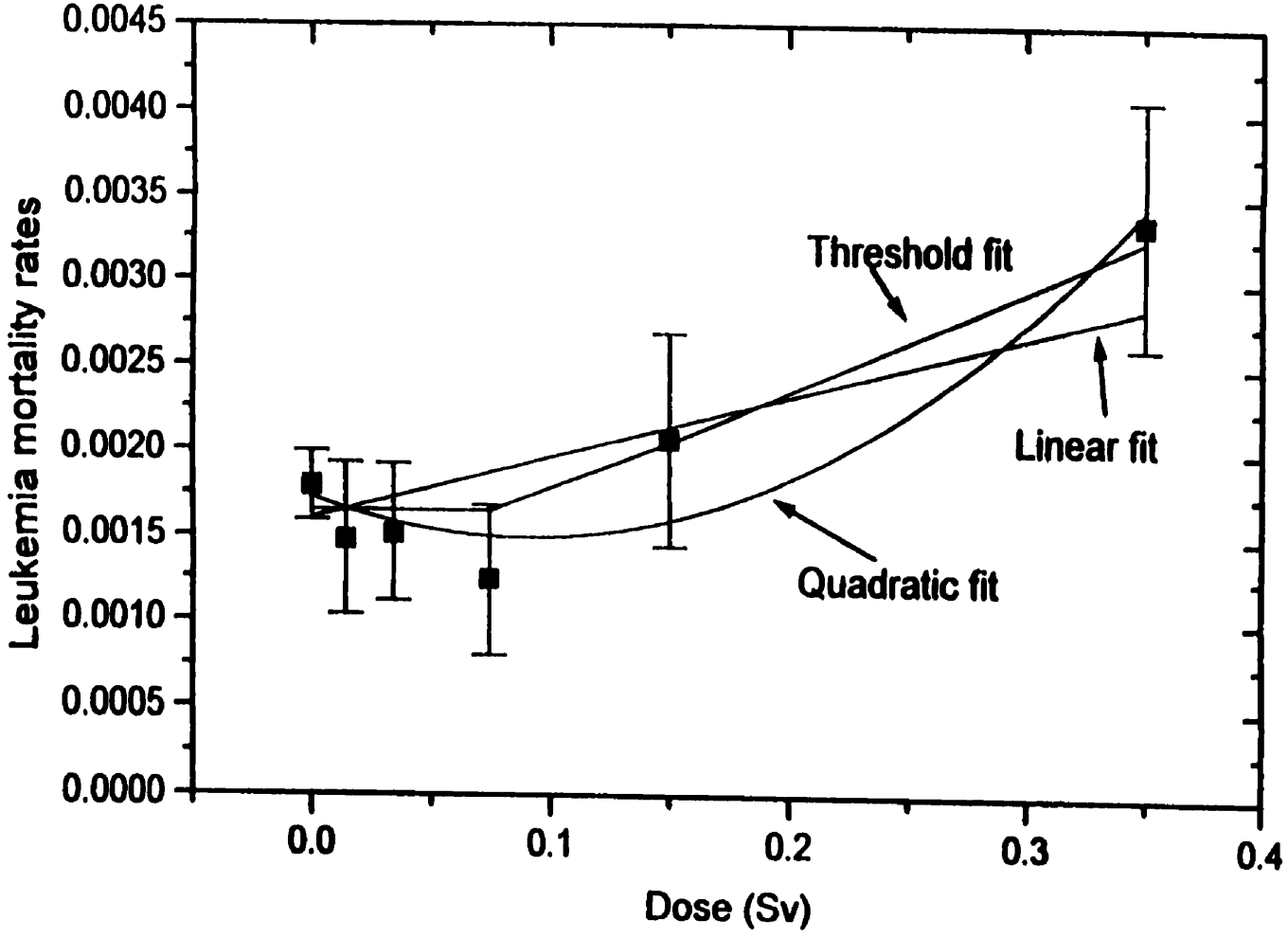

Figure 7.6 Leukemia mortality rates plotted as a function of received dose.
The [Shim 92] leukemia data set is shown.
The threshold $d_0 = 0.084$ Sv (see Table 7.2).

All fits were examined graphically and fits that did not look reasonable (as in Figure 7.7) were rejected on the assumption that fitting parameters were a local minima, but not the global could be minimum. Another indication of a minimum that was not a global minimum unrealistic parameter uncertainties. For example uncertainties for the parameters in Figure 7.7 were $\Delta p_1 = 0.00013$, $\Delta p_2 = 277403$, $\Delta d_0 = 5460617$.

Figures 7.8 shows the Chi-square for [Shim92] leukemia data set as a function of parameter d_0. For each value of d_0, p_1 and p_2 were optimized before calculating X^2. For values of parameter d_0 above 0.15 the Chi-square function has a constant value because p_2 can be chosen such that the fit goes through the data point for dose 0.35 Sv for any value of d_0 greater than 0.15. If we started minimization with d_0 greater than 0.15 the program can not find the global minimum. The problem of minimization is described in more details in chapter 10 of [Pres 92].

Tables 7.1 and 7.2 summarize results of all fits (linear, threshold, quadratic). Table 7.1 shows parameters of fit and their uncertainties. Stomach cancer [Pier 96] and [Shim 92] leukemia data sets have threshold parameter d_0 different from zero (0.014 Sv, 0.127 Sv and 0.084 Sv respectively). Three other data sets ([Pier 96] solid cancer, [Shim 92] lung and all solid cancers) data have threshold parameter d_0 practically equal to zero (threshold fit is equivalent to the linear). Threshold fits of these three data sets have threshold uncertainties equal to 0.025 Sv, 0.056 Sv and 0.107 Sv respectively.

Table 7.2 Contains Chi-square and reduced Chi-square values for each fit. The Chi-values were used to compute F values in columns six and ten. Columns six and ten contain F_X values for threshold and quadratic fits respectively. All these values are smaller

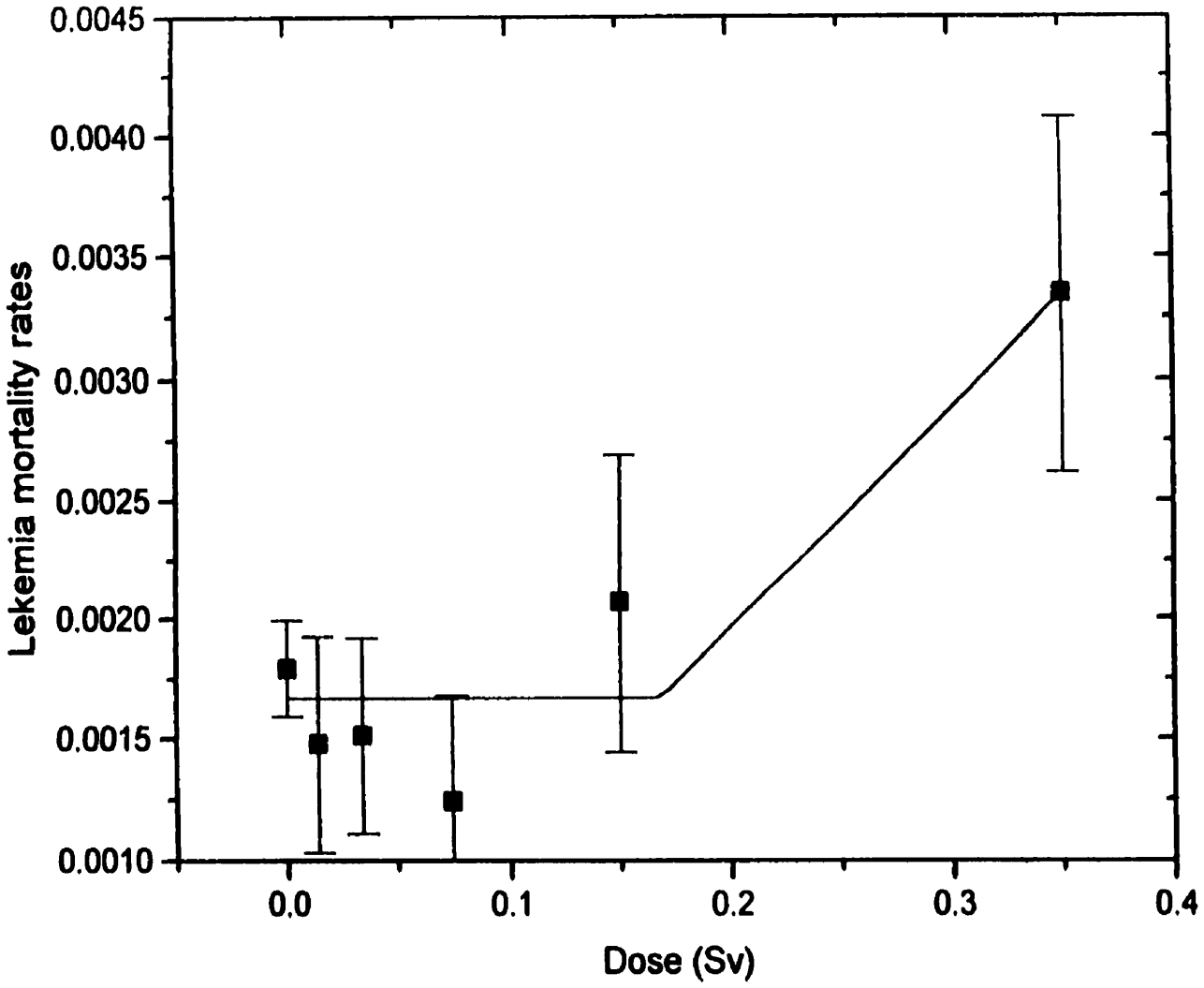

Figure 7.7 Example of poor fit for [Shim 92] leukemia data, obtained by fiding a local, not the global minimum. This fit was rejected.

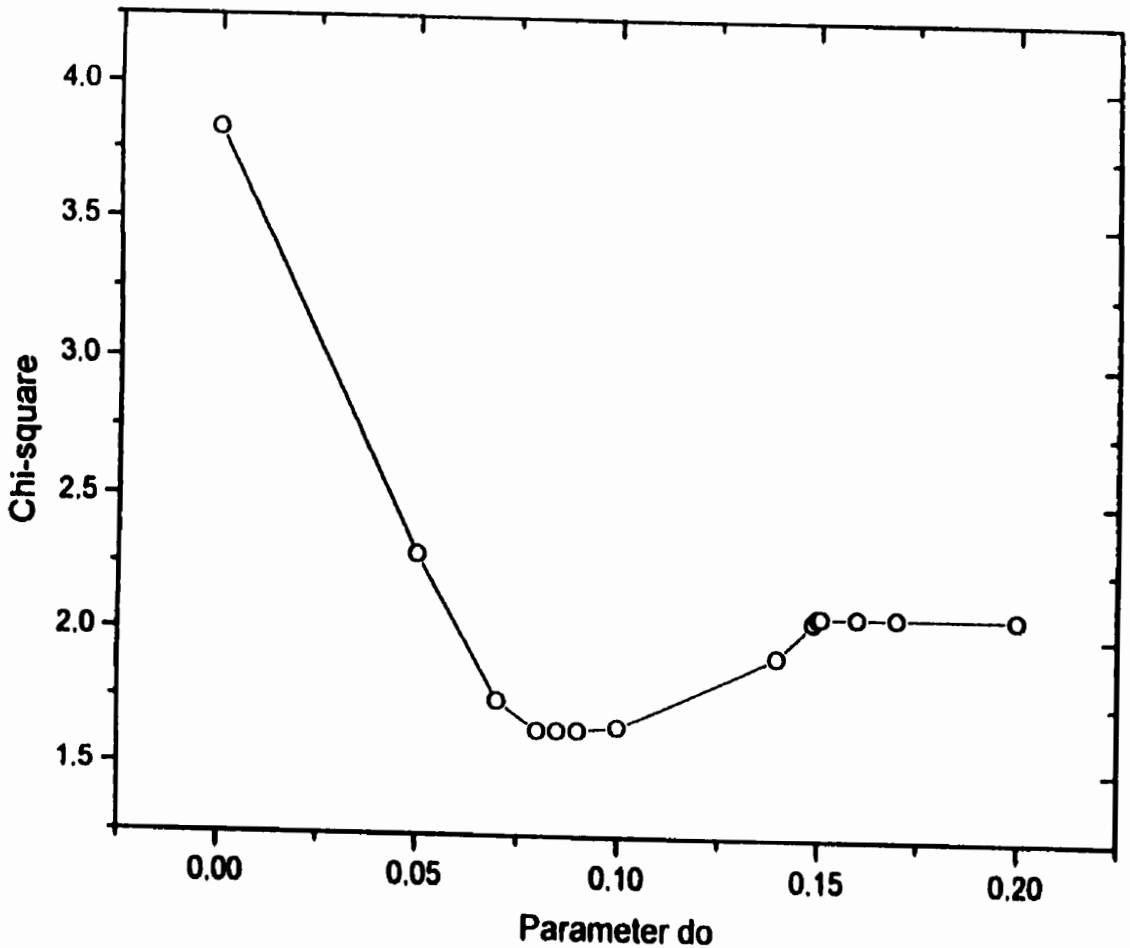

Figure 7.8 Chi-square for the threshold fit (equation 7.3) of [Shim 92] leukemia data set, plotted as a function of parameter do.

Table 7.1 Table presents parameters of fit for cancer and leukemia mortality rates. The threshold values for [Pier 96] solid cancer data, [Shim 92] lung and all solid cancer data sets are zero (the threshold fit corresponds to the linear fit).

Fit	Linear fit (10^{-3})				Threshold fit (10^{-3})						Quadratic fit (10^{-3})					
Data set	p_1	p_2	Δp_1	Δp_2	p_1	p_2	d_0	Δp_1	Δp_2	Δd_0	p_1	p_2	p_3	Δp_1	Δp_2	Δp_3
[Pier 96] solid c.	83	35	0.95	4.1	83	35	0.00	1.1	5.6	25	83	43	-7.0	1.1	11	8.6
[Shim 92] stomach c.	25	13	0.40	4.7	25	13	14	6.2	7.2	100	25	17	-12	6.0	17	58
[Shim 92] lung c.	7.7	7.9	0.30	3.1	7.5	5.9	0.00	0.46	4.5	56	7.6	12	-14	0.36	12	36
[Shim 92] all solid c.	72	38	1.5	15	72	37	0.00	2.1	21	107	72	62	-73	1.9	62	180
[Pier 96] leukemia	1.7	6.7	0.20	1.3	1.9	8.5	127	0.10	1.0	41	1.8	3.8	2.9	0.21	2.6	2.4
[Shim 92] leukemia	1.6	3.6	0.16	1.9	1.7	6.4	84	0.11	3.5	101	1.7	5.2	29	0.12	3.7	11

Table 7.2 Table includes Chi-square and reduced Chi-square values for linear, threshold and quadratic fits. Column six contain F test values for the threshold fit (see text). Column ten contain F test values for the quadratic fit. All test F values are smaller than required critical values, what indicates that threshold and quadratic fits do not significantly fit data better than the linear fit.

Data Set	Linear fit ($\nu = 4$)		Threshold fit ($\nu = 3$)			Quadratic fit ($\nu = 3$)		
	X^2_{red}	X^2	X^2_{red}	X^2	F_X	X^2_{red}	X^2	F_X
[Pier 96] solid c.	0.758	3.03	1.01	3.03	0.00	0.801	2.40	0.09
[Shim 92] stomach c.	0.573	2.29	0.750	2.25	0.01	0.753	2.26	0.004
[Shim 92] lung .	0.715	2.86	0.954	2.86	0.00	0.907	2.72	0.02
[Shim 92] all solid c.	2.08	8.32	2.80	8.40	0.00	2.63	7.89	0.02
[Pier 96] leukemia	1.31	5.22	0.372	1.12	1.22	1.16	3.48	0.17
[Shim 92] leukemia	0.957	3.830	0.545	1.63	0.45	0.412	1.24	0.70

than the corresponding critical value (10.1) of F distribution. This implies that adding one additional parameter in threshold and quadratic fits do not significantly improve these fits in a comparison to the linear fit. Thus, these data do not provide significant evidence for the threshold or quadratic models.

8. Conclusion

In this thesis reanalysis of the excess cancer and leukemia mortality rates for the atomic bomb survivors was carried out on a transformed version of the raw data; methods used included the standard Chi-square analysis and a Monte Carlo simulation. The simulation was used to obtain estimates of the threshold for a linear model, the *ZEP* for a linear-quadratic model and upper limits of their uncertainties. The linear model was applied to solid cancer and leukemia data, the linear-quadratic model to leukemia data only. First the standard Chi-square method was applied to the excess cancer (leukemia) mortality rates. The Cancer and leukemia data were fitted with a linear function in the form described in Chapter I. That form is capable of describing the threshold effect. The linear no-threshold model uses a straight line which is forced to go through the coordinate origin (the intercept to the dose axis is equal to zero). Depending on the signs of estimators *b*, and *c*, a linear-quadratic function can describe several models, as shown in section two. The linear-quadratic model has the ability to describe the hormesis effect which is of particular interest.

When statistical dose uncertainties (arbitrarily estimated as 25% of dose interval width) are included in the calculation, estimators of the fit are not too different from the case when these uncertainties were not included for the linear fit (see Figure 6.1-6.4, and Table 6.1). For the linear-quadratic fit the effect of including dose uncertainties is more apparent (see Figure 6.5 and 6.6, Table 6.2).

Table 8.1 summarizes results of the analysis of all data sets. Columns three to seven contain results obtain by least-squares fit. Columns four and five contain *p* values

for estimates *b* and *c*. The *p* values higher than 0.1 are usually taken as statistical non-significant. Columns eight to eleven contain results of Monte Carlo simulation for threshold. The column eight contains median threshold and *ZEP* values obtained by simulation. The nine eight contains upper limit of 95% thresholds produced by straight lines with positive slope *b* for linear fit (negative slopes are excluded, see sections 4.3), the column ten contains upper limit of 95% *ZEP* values produced by parabolas with b<0 and c>0, hormesis model for linear-quadratic fit, (see section 5.3). The last column eleven contains 95% confidence interval for median threshold value for the linear fit, or 95% confidence interval for median *ZEP* value for the linear-quadratic fit as obtained from thousand simulated experiments.

Stomach cancer data set and both leukemia data sets (linear fit) suggest possibility of existence of threshold namely, median threshold values (obtained by the least-squares fit and by simulation) are positive, 95% confidence intervals for threshold median are in positive limits, and of course the upper limits of 95% thresholds produced by lines with b>0 are positive. Upper limits of 95% thresholds produced by lines with b>0 are positive for all data sets. Only [Pier 96] leukemia data set has significant p_b value equal to 0.02.

For both leukemia data sets fitted with linear-quadratic function the p_c values are 0.20 (it is high value that suggests non-significant result for quadratic coefficient *c*). Only 9.9% of simulations for the [Pier 96] leukemia data set correspond to hormesis model (b<0, c>0). The [Shim 92] leukemia data set has 85.6% simulations that correspond to hormesis model. The median *ZEP* , 95% confidence interval for *ZEP* median and the upper limits of 95% *ZEP* values produced by parabolas with b<0 and c>0 are positive.

Table 8.1 Summary table, the last four columns contain results of Monte Carlo simulation. Threshold values are applicable for linear fit (see Chapter IV), *ZEP* for linear-quadratic fit (see Chapter V).

Form of function used to fit data	Set of data	The Chi-square method					M.C. Simulation (*T* and *ZEP*)			
		Goodness of fit			*T* (Sv)	*ZEP* (Sv)	Median (Sv)	Upper 95 *T* (b>0) (Sv)	Upper 95 *ZEP* (b<0, c>0) (Sv)	95% conf. interval for median (Sv)
		X^2_{red}	p_b	p_c						
Linear $y = a + b \cdot x$	[Pier 96] solid cancer	0.69	0.01	-	-0.061	-	-0.059	0.037	-	-0.066--0.053
	[Shim 92] stomach c.	0.63	0.2	-	0.024	-	0.026	0.128	-	0.021-0.032
	[Shim 92] lung cancer	0.63	0.3	-	-0.108	-	-0.074	0.054	-	-0.084--0.062
	[Shim 92] all solid cancers	2.37	0.1	-	-0.015	-	-0.011	0.061	-	-0.016--0.007
	[Pier 96] leukemia	0.31	0.02	-	0.099	-	0.099	0.154	-	0.096-0.101
	[Shim 92] leukemia	0.3	0.2	-	0.104	-	0.102	0.193	-	0.099-0.105
Linear-quadratic $y=b \cdot x+c \cdot x^2$	[Pier 96] leukemia	1.19		0.2	-	-0.795	-0.675	-	-	-0.752--0.264
	[Shim 92] leukemia	0.39		0.2	-	0.197	0.203	-	0.235	0.197-0.206

This suggests weak possibility of hormesis existence for the [Shim 92] leukemia data. This result should be interpreted with reserve because of high p_c value (0.20).

Analysis of cancer and leukemia mortality rates (Chapter VII) does not introduce correlation among dependent variables. Threshold parameter d_0 is different from zero for [Shim 92] stomach cancer data and both leukemia data sets. According to the F-test none of the six data sets provide significant evidence for threshold and quadratic models.

Appendices

A-1. The two sided *t* test concerning slope *b*

This test tests whether or not there is a linear relationship between fitted variable Y and variable d, using a linear regression model [Neter 90, p. 69]. The test is in the form of alternatives:

$$\begin{aligned} H_0 &: \; b = 0 \\ H_a &: \; b \neq 0 \end{aligned} \tag{A-1}$$

The first alternative is a null hypothesis H_0. The null hypothesis assumes that slope b is equal to zero. If the null hypothesis is true, then there is no linear relationship between variables. In order to test the null hypothesis the t value (ratio) is calculated:

$$t = \left| \frac{b}{s_b} \right| \tag{A-2}$$

.

The t value is used to establish the decision rule:

$$If \; t \leq t(1-\frac{\alpha}{2}; n-2), \quad H_o \; is \; valid \tag{A-3}$$

$$If \; t > t(1-\frac{\alpha}{2}; n-2), \quad H_a \; is \; valid \tag{A-4}$$

where α is a level of significance.

For example consider the [Pier 96] solid cancer data . In table 4.1 the value of slope b is 32.6, and the value of variance s_b^2 is 28.9. The standard deviation s_b is 5.36 . The t value is 6.06. In this example the number of degrees of freedom is 3. In the table of t distribution the value of 6.06 is between t (0.995, 3) = 5.841 and t (0.9975, 3) = 7.453 for three degrees of freedom [Neter 90 p.1128]. According to A-4 we can conclude that slope b is different from zero with confidence

$$\alpha / 2 = 1 - 0.995 = 0.005 \quad or \quad \alpha = 0.01. \tag{A-5}$$

The value of p smaller than 0.01, because $t = 6.06 > 5.841$. Thus, it can be concluded that slope b is different from zero with 99% of confidence.

A-2. The two sided t test concerning quadratic effect coefficient c

This test tests whether or not quadratic term c could be omitted from the model [Neter 90 p 326]. The test is in the form of alternatives:

$$\begin{aligned} H_0 &: \ c = 0 \\ H_a &: \ c \neq 0 . \end{aligned} \tag{A-6}$$

If H_0 is valid quadratic term c can be dropped from the model. The t value is calculated as

$$t = \left| \frac{c}{s_c} \right| . \tag{A-7}$$

The decision rule is established by A-3 and A-4 as for previous test.

For example consider the [Pier 96] leukemia data. In the table 5.1 value of quadratic term c is 3.56, and value of variance s_c^2 is 4.63. The standard deviation s_c is 2.15 . The t value is 1.65. In the table of t distribution the value 1.65 is between t (0.90, 3) = 1.638 and t (0.95, 3) = 2.353 for 3 degrees of freedom. According to A-4 we can conclude that quadratic coefficient c is different from zero with confidence

$$\alpha / 2 = 1 - 0.90 = 0.10 \quad or \quad \alpha = 0.20 \cdot \tag{A-8}$$

The value of p is smaller than 0.20, because $t = 1.65 > 1.638$. It means that quadratic term c is different from zero (can not be dropped from the model) with 80% of confidence. Usually required level of confidence to accept assumed hypothesis is 95% or more, or $p \leq 0.05$.

A-3 Determination of threshold and its statistics

As was described in section 3.31 each simulation with a linear fit, gives one thousand estimators for T (equation 3.17). A statistical analysis of those estimators was performed, then a histogram (frequency of T versus the dose) was plotted, the mean value, the trimmed mean value, the standard deviation, the standard error of the mean, the first and third quartiles, 95% confidence interval for the median value was calculated. The distribution of the estimated thresholds versus received dose is asymmetric. Because of that non-parametric, sign confidence interval for the median is used to obtain a confidence interval of the threshold. This part of the analysis is represented in section 4. Calculations were done using the statistical program Minitab. Minitab calculates those values as follows [Ryan 85]:

Denote T_i the i -th estimated threshold value ($i = 1..1000$ (N)).

Mean value is

$$T_{mean} = \frac{1}{N}\sum_{i=1}^{N} T_i . \qquad \text{(A-9)}$$

Median value shows a central point of data (50% are below and 50% above that point). The data are ordered in ascending order. If N is odd, median is the middle value. In case of $N = 1000$ (even) the median is calculated as the average value of two central points (500-th and 501-st).

Trimmed mean (e.g. 5% trimmed mean), is calculated by sorting the data in ascending order, then the smallest 5% and the largest 5% of the values are deleted, and the rest (90%) of the values are averaged in the standard manner that is using equation A -9.

Standard deviation is calculated in the standard maner

$$S_T = \sqrt{\frac{\sum_{i=1}^{N} \left(T_i - T_{mean}\right)^2}{N-1}} . \qquad \text{(A-10)}$$

Standard error of the mean is

$$S_{Emean} = \frac{S_T}{\sqrt{N}} . \qquad \text{(A-11)}$$

First and third quartile's divide the data in groups that contain the first 25% of data and the last 25% of data sorted in ascending order.

Sign test for median confidence interval does not assume a form of distribution. This test is used because simulated threshold distribution is asymmetric. This test uses the Binomial distribution to compute the required confidence level.

Binomial distribution can be used to compute that x values out of the total N observations are or are not in one of the two possible categories. For purposes of the calculation confidence interval x-th value can be in or out of required interval of confidence.

Binomial distribution has the form:

$$P(x) = \binom{N}{x} \cdot \pi^x \cdot (1-\pi)^{(N-x)}. \qquad \text{(A-12)}$$

In order to calculate the confidence level equal to p, first all data points have to be ordered. The position of each data point has a binomial probability according to equation A-12 (x is number of data points less than the median, N is the total number of data points, π =0.5). For p = 95% (95% of confidence) the cumulative probability of all points in the interval has to be 0.95. This means that the cumulative probability of all points out of the interval is 0.05. The cumulative binomial probability for x smallest values in the row is

$$P(0..x) = \sum_{r=0}^{x} \binom{N}{r} \cdot 0.5^r \cdot 0.5^{(N-r)}. \qquad \text{(A-13)}$$

Because for π = 0.5, the binomial distribution is symmetrical, the cumulative probability for the largest x values in the row is equal to the cumulative probability for the smallest x values in the row. In order to get 95% confidence interval, all of the smallest x values (cumulative probability 0.025), and all of the largest x values (the same cumulative probability 0.025) have to be left out of the interval. The endpoints of the smallest interval containing the rest of the points is 95% confidence interval.

For example if N =1000, the binomial cumulative probability for the smallest and largest 469 points (together) is 0.0463. The binomial cumulative probability for the smallest and largest 470 points (together) is 0.0537.

The interval whose endpoints are the 470-th and 531-st gives 95.37% confidence interval. The interval whose endpoints are the 471-th and 530-st gives 94.63% confidence interval. The exact 95% confidence interval for 1000 events can only be interpolated between 94.63% and 95.37% confidence.

Apendix B-1

Maple program for computation of estimates a, b, Chi-quare, reduced Chi-square values and Error matrix for linear fit (see Chapters III and IV).

```
> restart;

> N:=Integer_1:

> M:=array(1..N,1..3,[[d1,Y1,sY1],[d2,Y2,sY2],.....,[dN,YN,sYN]]);

> x[k]:=M[k,1]:

> y[k]:=M[k,2]:

> s[k]:=M[k,3]:

> A1:=sum(y[k]/s[k]^2,k=1..N):

> A2:=sum(x[k]/s[k]^2,k=1..N):

> A3:=sum((x[k]*y[k])/s[k]^2,k=1..N):

> A4:=sum(x[k]^2/s[k]^2,k=1..N):

> B1:=sum(1/s[k]^2,k=1..N):

> C:=B1*A4-A2^2:

> a:=(A1*A4-A2*A3)/C:

> b:=(B1*A3-A1*A2)/C:

> evalf(a);

> a1:=%:

> evalf(b);

> b1:=%:

> Sa:=sqrt((A4/C)):

>evalf(Sa);

> Sa1:=%:

> Sb:=sqrt((B1/C)):

> evalf(Sb);
```

```
> Sb1:=%:

> f[k]:=a+b*x[k];

> K1:=sum((y[k]-f[k])^2/(s[k])^2,k=1..N);

> K:=evalf(K1/3);

 > Sa2:=sum((s[k]^2/C^2)*((A4/s[k]^2)-(A2*x[k]/s[k]^2))*((A4/s[k]^2)-(A2*x[k]/s[k]^2)),k=1..5):

> evalf(%);

> sqrt(%);

> Sab2:=sum((s[k]^2/C^2)*((B1*x[k]/s[k]^2)-(A2/s[k]^2))*((A4/s[k]^2)-(A2*x[k]/s[k]^2)),k=1..5):evalf(%);

> Sb2:=sum((s[k]^2/C^2)*((B1*x[k]/s[k]^2)-(A2/s[k]^2))*((B1*x[k]/s[k]^2)-(A2/s[k]^2)),k=1..5);

> sqrt(%);

> Error_Matrix:=array(1..2,1..2,[[Sa2,Sab2],[Sab2,Sb2]]);
```

Apendix B-2

Maple program for computation estimates b, c, and Error matrix for a linear-quadratic fit (see Chapters III and V).

```
>restart;
> N:=Integer_1:
> M:=array(1..N,1..3,[[d1,Y1,sY1],[d2,Y2,sY2],.....,[dN,YN,sYN]]);
> x[k]:=M[k,1]:
> y[k]:=M[k,2]:
> s[k]:=M[k,3]:
> A21:=sum((x[k]*y[k])/s[k]^2,k=1..N):
> A23:=sum((x[k]^3/s[k]^2),k=1..N):
> A13:=sum(x[k]^2/s[k]^2,k=1..N):
> A31:=sum((y[k]*x[k]^2)/s[k]^2,k=1..N):
> A33:=sum(x[k]^4/s[k]^2,k=1..N):
> D11:=sum(x[k]/s[k]^2,k=1..N):
> C1:=array(1..2,1..2,[[A13,A23],[A23,A33]]):
> c:=linalg[det](C1):
> A2:=array(1..2,1..2,[[A21,A23],[A31,A33]]):
> a:=linalg[det](A2):
> a2:=evalf(a/c);
> A3:=array(1..2,1..2,[[A13,A21],[A23,A31]]):
> b:=linalg[det](A3):
> a3:=evalf(b/c);
> Y:=a2*x+a3*x^2;
> Sb2:=sum((s[k]^2/c^2)*((A33*x[k]/s[k]^2)-(A23*x[k]^2/s[k]^2))*((A33*x[k]/s[k]^2)-
(A23*x[k]^2/s[k]^2)),k=1..5);sqrt(%);
```

```
> Sbc2:=sum((s[k]^2/c^2)*((A33*x[k]/s[k]^2)-(A23*x[k]^2/s[k]^2))*((A13*x[k]^2/s[k]^2)-
(A23*x[k]/s[k]^2)),k=1..5);
> Sc2:=sum((s[k]^2/c^2)*((A13*x[k]^2/s[k]^2)-(A23*x[k]/s[k]^2))*((A13*x[k]^2/s[k]^2)-
(A23*x[k]/s[k]^2)),k=1..5);sqrt(%);
> Error_Matrix:=array(1..2,1..2,[[Sb2,Sbc2],[Sbc2,Sc2]]);
```

APENDIX B-3

Maple program for computing the "Delta-Chi-Square" paraboloid for linear fit (see Chapters III and IV).

```
> restart;
> d1:=valu_1:.....dn:=value_n:
> M:=array(1..N,1..3,[[d1,Y1,sY1],[d2,Y2,sY2],.....,[dN,YN,sYN]]);
> abest:=a0:
> bbest:=b0:
> plane1:=2.30:
> plane2:=6.17:
> plane3:=9.21:
> pogled:=0..11:
> orent:=[40,83]:
> N:=Integer_1:
> x1:=valu_x1: x2:=value_x2:
> y1:=value_y1: y2:=value_y2:
> x[k]:=M[k,1]:
> y[k]:=M[k,2]:
> s[k]:=M[k,3]:
>
> Ko:=sum((y[k]-abest-bbest*x[k])^2/(s[k])^2,k=1..N):
> f[k]:=a+b*x[k]:
> Ki:=sum((y[k]-f[k])^2/(s[k])^2,k=1..N):
> K:=Ki-Ko:
> a:=y:b:=x:
> z:=K:
```

```
> colorplot3d:=proc(f::algebraic, xrange::`=`, yrange::`=`) /* colr3d
command [Kofler 97 p.485]*/
>   local varx, vary, x, x0, x1, y, y0, y1, i, j, opts, gridopt,
>         dataf1, dataf2, datac1, datac2, c:
>   Digits:=6:
>   #
>   #     analyse parameters
>   #
>   varx:=op(1,xrange): vary:=op(1,yrange):
>   x0:=evalf(op(1, op(2,xrange))): x1:=evalf(op(2, op(2,xrange))):
>   y0:=evalf(op(1, op(2,yrange))): y1:=evalf(op(2, op(2,yrange))):
>   #
>   #     analyse options
>   #
>   gridopt:=[20, 20]:      # defaults
>   opts:=[args[4..nargs]]:
>   i:=1:
>   while i<=nops(opts) do:
>     if type(opts[i], identical(`grid`)=list) then
>       gridopt:=rhs(opts[i]):
>       opts:=subsop(i=NULL, opts):  # remove from options list
>     elif type(opts[i], identical(`color`)=function) or
>           type(opts[i], identical(`COLOR`)=function) then
>       c:=[ op(2..4, rhs(opts[i])) ]:
>       opts:=subsop(i=NULL, opts):  # remove from options list
>     else
>       i:=eval(i)+1:
>     fi:
>   od:
```

```
>   if c='c' then  # no color option
>     ERROR(`wrong or missing color function, use
color=COLOR(RGB,r,g,b)`):
>   fi:
>   #     loop to calculate graphic- and colordata
>
>   #
>   dataf1:=[]: datac1:=[]:
>   for i from 1 to gridopt[1] do:
>     x:=evalf(x0+(x1-x0)/(gridopt[1]-1)*(i-1)):
>     dataf2:=[]: datac2:=[]:
>     for j from 1 to gridopt[2] do:
>       y:=evalf(y0+(y1-y0)/(gridopt[2]-1)*(j-1)):
>       dataf2:=[ op(dataf2), evalf( subs(varx=x, vary=y, f)) ]:
>       datac2:=[ op(datac2), op(evalf( subs(varx=x, vary=y, c))) ]:
>     od:
>     dataf1:=[ op(dataf1), dataf2]:
>     datac1:=[ op(datac1), op(datac2)]:
>   od:
>   #    show plot, use remaining options
>   #
>   plots[display]( PLOT3D( GRID(x0..x1, y0..y1, dataf1,



>                                 COLOR(RGB, op(datac1))) ), op(opts) );
> end:
>
> greyscale:=(x)->COLOR(RGB,0.85,0.9,0.95):
```

```
> with(plots):g0:=colorplot3d(z, x=x1..x2, y=y1..y2,
color=greyscale(z*0.01+0.1),
>   grid=[40,40], view=pogled,
orientation=orent,axes=boxed,numpoints=10000,labels=[` `,` `,` `],
scaling=unconstrained, style=patch):
> with(plottools):with(plots):
> g1:=plot3d(plane1, x=x1..x2, y=y1..y2,axes=none,labels=[` `,` `,` `]):
> g2:=plot3d(plane2, x=x1..x2, y=y1..y2,axes=none,labels=[` `,` `,` `]):
> g3:=plot3d(plane3, x=x1..x2,y=y1..y2,axes=none,labels=[` `,` `,` `]):
> plotsetup(jpeg,
plotoutput=`FileName.jpg`,plotoptions=`portrait,width=600,height=700,nobor
der`);
display([g0,g1,g2,g3],axes=boxed,ambientlight=[0.7,0.9,0.8],font=[TIMES,BO
LD,15],axesfont=[TIMES,BOLD,15]);
```

APENDIX B-4:

Maple program for linear-quadratic fit of simulated set of data (see Chapters III and V).

```
/*N  is number of simulations; n number of points in each simulation.*/
> restart;
> readdata(`FileNAme.txt`,float,n):
> M0:=matrix(%):
> close(`FileNAme.txt`);
> p1:=value_1:
> p2:=value_2:
> pn:=value_n:
> N:=Integer_1:
> bbest:=value_b_best: cbest:=value_c_best:
> for z from 1 by 1 to N do
Mz:=array(1..n,1..3,[[d1,M0[z,1],p1],.....,[dn,M0[z,5],p5]]):
x[k]:=Mz[k,1]:
y[k]:=Mz[k,2]:
s[k]:=Mz[k,3]:
A21:=sum((x[k]*y[k])/s[k]^2,k=1..N):
A23:=sum((x[k]^3/s[k]^2),k=1..N):
A13:=sum(x[k]^2/s[k]^2,k=1..N):
A31:=sum((y[k]*x[k]^2)/s[k]^2,k=1..N):
A33:=sum(x[k]^4/s[k]^2,k=1..N):
C1:=array(1..2,1..2,[[A13,A23],[A23,A33]]);
c:=linalg[det](C1):
A2:=array(1..2,1..2,[[A21,A23],[A31,A33]]):
a:=linalg[det](A2):
```

```
a2:=evalf(a/c);
A3:=array(1..2,1..2,[[A13,A21],[A23,A31]]):
b:=linalg[det](A3):
a3:=evalf(b/c);
sln:=solve(a2*x+a3*x^2=0);
v:=sln[1]:
T:=sln[2]:
f[k]:=bbest*x[k]+cbest*(x[k])^2:
K1:=sum((y[k]-f[k])^2/(s[k])^2,k=1..N):
appendto(`FileName_Res.txt`):
array([z,a2,a3,T,K1,z]);
writeto(terminal);
> od;
```

Glossary

Absorbed dose (*D*)- the energy imparted to matter by ionizing radiation per unit mass. SI unit of absorbed dose is gray (Gy).

Background radiation - radiation that is part of natural environment (not caused by a human action). It is caused by natural radio-isotopes in environment and cosmic radiation.

Cancer (leukemia) incidence rate - number of cancer (leukemia) cases per person in observed population.

Cancer (leukemia) mortality rate - number of deaths due to cancer (leukemia) per person in observed population.

[*c-d./p.*] - means cancer (leukemia) deaths per person.

Cumulative dose - is a total dose received if someone was repetitiously exposed to radiation.

Dose equivalent (*DE*)- is a product of the absorbed dose and quality factor (see Quality factor) $DE=Q*D$. Si unit for dose equivalent is Sievert (Sv).

Dose ranges - Arbitrary ranges of received dose. Low dose is below 0.2 Sv, intermediate dose is between 0.2 and 2.5 Sv, high dose is above 2.5 Sv. [Brill 82].

Gray (Gy) - Si unit for absorbed dose. One Gray is 1 Joule of energy imparted to 1 kg of matter by ionizing radiation, (1Gy = 1J/kg).

Hormesis - beneficiary influence of some agent on health. For case of ionizing radiation it is hypothetical.

Ionizing radiation - any radiation that produces ionization in primary or secondary processes (*X* and gamma ray photons, charged and uncharged particles).

Linear no-threshold model - a model that assumes linear dependence between number of cancers and received dose of ionizing radiation. It assumes that straight line passes through the coordinate origin.

Linear threshold model - a model that assumes existence of a threshold dose below which ionizing radiation has no effect on health. Above the threshold linear dependence between number of cancer and received dose is assumed.

Maple - a computer algebra program, developed by Waterloo Maple, Inc.

Minitab - a computer statistical program developed by Minitab Inc.

Origin - a computer program for data analysis and technical graphics developed by Microcal Software, Inc.

Quality factor (*Q*) - is a multiplicative factor that express effectiveness of ionizing radiation on biological tissue. This factor is equal to one for x rays, gamma rays, and beta particles. Quality factor is equal to 20 for fast neutrons, alpha particles, and heavy particles. For fast neutrons some authors use quality factor equal to ten.

Sievert (Sv) - Si unit for dose equivalent. For ionizing radiation with quality factor one (x rays, gamma rays beta particles) 1Sv = 1Gy.

Threshold *(T)*- an assumed dose below which the effect of ionizing radiation on health does not exist.

ZEP - refers to Zero Equivalent Point. In the hormesis model this is the dose below which radiation has beneficial effect on a biological systems, and above it has harmful effect. At that point the effect of ionizing radiation is equal to zero (no effect).

REFERENCES:

(Beir V 90) Committee on the Biological Effects of Ionizing Radiation, "Health Effects of Exposure to Low Levels of Ionizing Radiation - BEIR V", National Academy Press, Washington, D.C. (1990).

(Bevi 92) P. R. Bevington, D. K. Robinson, "Data Reduction and Error Analysis for the Physical Sciences", Second Edition, McGraw-Hill, New York (1992).

(Bres80) N. Breslow, N. Day, "Statistical Methods in Cancer Research Volume I" International Agency for Research on Cancer, Lyon (1980).

(Bres 87) N. Breslow, N. Day, "Statistical Methods in Cancer Research Volume II" International Agency for Research on Cancer, Lyon (1987).

(Bril 82) B. Brill, "Low-level Radiation Effects: A Fact Book", Society of Nuclear Medicine, New York (1982).

(Cohe 97) Bernard L. Cohen, "Lung Cancer Rate vs. Mean Radon Level in U.S. Counties of Various Characteristics", Health Physics, Vol. 72 (No. 1), pp. 114-119 (1997).

(Epa 94) Radiation Advisory Committee of the EPA Science Advisory Board, "Estimating Radiogenic Cancer Risks", U.S. Environmental Protection Agency, Washington, DC. (1994).

(Este 94) Esteve at al., "Statistical Methods in Cancer Research Volume IV" International Agency for Research on Cancer, Lyon (1987).

(Glob 92) H. W. Ellasaesser, editor, "Global 2000 revisited", Paragon House, New York, 1992.

(Heid 97) W. Heidenreich, H. Paretzke, P. Jacob, "No evidence for increased tumour rates below 200 mSv in the atomic bomb survivors data", Radiation Environmental Biophysics 36 (1997), p. 205-207.

(Kofl 97) M. Kofler, "Maple an introduction and reference", Addison Wesley Longman Ltd. Harlow, England, (1997).

(Kondo 93) S. Kondo, "Health Effects of Low-level Radiation", Osaka, Japan and Medical Physics Publishing, Madison, WI, USA (1993).

(Losal 95) Los Alamos staff, "Radiation Protection and the Human Radiation Experiments", Los Alamos Science, No. 23 (1995).

(Lucke 91) T. D. Luckey, "Radiation Hormesis", Chemical Rubber Company Press, Inc., Boca Raton, Florida (1991).

(Lucke 92) T. D. Luckey, "Hormesis and nature with ionizing radiation", in (Glob 92), p. 189-251.

(NCRP 93) National Council on Radiation Protection and Measurements, "Limitation of Exposure to Ionizing Radiation", NCRP Report No. 116,Bethesda (1993).

(Neter 90) J. Neter, W. Wasserman, M. H. Kutner, "Applied Linear Statistical Models", IRWIN, Homewood, IL, Boston, MA (1990).

(Okam 92) K. Okamoto, "A Model to Explain Radiation Hormesis, its Underlying Mechanisms and Comparisons with Experimental Data", p. 449-456. in Suga 92).

(Pier 90) Donald A. Pierce, Daniel O. Stram, and Michael Vaeth, "Allowing for Random Errors in Radiation Dose Estimates for the Atomic Bomb Survivor Data," Radiation Research, Vol. 123, pp. 275-284 (1990).

(Pier 96) Donald A Pierce et all., "Studies of the Mortality of Atomic Bomb Survivors. Report 12, Part I. Cancer: 1950-1990." Rad. Research 146, pp. 1-27 (1996).

(Pres 92) William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery, "Numerical Recipes in Fortran - The Art of Scientific Computing", Second Edition, Cambridge University Press, Cambridge (1992).

(Pres 94) Dale L. Preston, et al., "Cancer Iincidence in Atomic Bomb Survivors. Part III: Leukemia, Lymphoma and Multiple Myeloma, 1950-1987," Radiation Research, Vol. 137, pp. S68-S97 (1994).

(Ryan 85) B. Ryan, B. Joiner, T. Ryan, "Minitab-Handbook", PWS-Kent Publishing Company, Boston (1985).

(Schi 95) Mario E. Schillaci, "Radiation and Risk: A Hard Look at the Data", pp.90-115, in (Losal 95).

(Shesk 97) D. J. Sheskin, "Handbook of Parametric and Nonparametric Statistical Procedures", CRC Press, Boca Raton, New York, London, Tokyo (1997).

(Shim 92) Y. Shimizu, H. Kato, W. J. Schull, and K. Mabuchi, "Dose-response analysis among atomic-bomb survivors exposed to low-level radiation" pp. 71-74, in (Suga 92).

(Spos 91) Richard Sposto, Daniel O. Stram, and Akio A. Awa, "An Estimate of the Magnitude of Random Errors in the DS86 Dosimetry from Data on Chromosome Aberrations and Severe Epilation," Radiation Research, Vol. 128, pp. 157-169; (1991).

(Suga 92) T. Sugahara, L. Sagan, T. Aoyama, editors, "Low Dose Irradiation and Biological Deafens Mechanisms", Elsevier Science Publishers B.V. (1992).

(UNSCEAR 94) Substantive reports of the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) to the General Assembly are to be found in Official Records of the General Assembly, 49th Session, Suppl. No. 46 (A/49/46) "Sources and Effects of Ionizing Radiation", (1994).

(Winkl 75) R. Winkler, W. Hays, "Statistics: Probability, Inference, and Decision", Holt, Rinehart and Winston NY, Chicago, etc. (1975).