National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

NL-339 (r. 88/04)

Canada

IDENTIFYING PROGNOSTIC FACTORS OF STAGE I MALIGNANT MELANOMA

by

Gemai Chen

B.Sc., Southwestern Jiaotong University, China, 1981

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics and Statistics

© Gemai Chen 1987

SIMON FRASER UNIVERSITY

December 1987

All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-42565-2

APPROVAL

Name: Gemai Chen

Degree: Master of Science

Title of project:

Identifying Prognostic Factors of Stage I malignant Melanoma

Examining Committee:

Chairman: Dr. G. Bojadziev

Dr. M.A. Stephens
Senior Supervisor

Dr. T. Swartz

Dr. R. Routledge

Dr. P. Kim
External Examiner

Date Approved: November 19, 1987

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Identifying Prognostic Factors of Stage I Malignant Melanoma

Author:

(signature)

GEMAI CHEN

(name)

Nov. 24, 1987

(date)

ABSTRACT

The Cox proportional hazards regression model is employed to analyze a data set from a programme named The Western Canada Melanoma Study. Of the thirteen potential prognostic factors covered, melanoma tumour depth, sex and age are identified as important prognostic factors. The prognosis of thin melanoma and the BANS concept are also studied in light of the data, followed by a review and comparison of tumour depth and Clark's levels of tumour invasion, an investigation of a new prognostic index, a systematic search for interactions between the potential prognostic factors and a series of predictions of five-year survival of the melanoma patients.

Statistical techniques related to the Cox model are also provided from an intuitive and practical point of view together with comments on the theory behind the techniques.

ACKNOWLEDGEMENTS

My sincerest thanks are to my committee members Dr. M.A. Stephens, Dr. T. Swartz, Dr. R. Routledge and Dr. P. Kim for their timely help and encouragement at the various stages of completing this project. I am especially grateful to my senior supervisor, Dr. M.A. Stephens who formally introduced me into the field of statistics and has been guiding me into fruitful directions ever since. I would like to thank also the Mathematics and Statistics Department, from which I got all the necessary support to finish my master's degree. The professors and the students I studied with were so helpful that without them I would have not been able to make so many progresses. In particular, I am very grateful to John Spinelli, who helped me through one of the most difficult times when I was writing this project.

The Cancer Control Agency of British Columbia kindly provided the data, and Bill and Rick took pains to clean the data for me, to them I want to show my deep appreciation here.

DEDICATION

To my family

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Skin diseases are fairly common, but in general are not lethal. However, there are some cutaneous diseases which are fatal. Malignant melanoma, for example, has been a killer of thousands of people. According to aetiologists, this is a kind of cancer developed from melanocytes, which are derived from neural crest cells that migrate to the skin, eye, central nervous system, and occasionally elsewhere during fetal life.

Malignant melanoma represents about 2% of all cancers. Over recent years the incidence of melanoma has been rising rapidly and steadily. In the United States, the incidence has increased by 4.5% annually, and in British Columbia 7.2%. Here are some more facts about this cancer:

1. The vast majority of melanoma seem to arise from pre-existing benign nevi;
2. There is an increase in melanoma in body sites that are exposed to the sun, and the greater the exposure, the higher the incidence;
3. Melanomas occur in all races, but are rare in blacks, orientals, and others with dark skin, and are more common in fair-skinned races and individuals;
4. The incidence of melanoma is latitude-dependent, e.g., the incidence among white people in Atlanta is almost twice that of Detroit;

5. There are minor differences in occurrence between males and females, although females have a better prognosis for unknown but possibly hormonal reasons;
6. Almost no melanomas occur before puberty, and the peak incidence occurs in the fifth to seventh decades;
7. A small proportion of patients with melanoma have a hereditary component;
8. Administrative and professional workers were found to have a higher incidence than construction workers and farmers ([11],[32],[45],[46],[58]).

Medical experts are still observing this disease. On the one hand, it has long been recognized that certain characteristics of the patients and of their diseases may strongly influence their survival. The fact that some patients with the disease die while others survive suggests that different factors affect prognosis. On the other hand, observations have been accumulated and analyzed to give useful assistance to conquer this deadly disease finally. For instance, point 2 in the above list emphasizes the importance of watching for any changes in existing lesions, and point 8 suggests that intermittent intense sun exposure may be a harmful pattern.

In general, characteristics that influence prognosis (of a disease) are called <u>covariates</u>, or <u>prognostic factors</u>. In cancer research, there are several reasons for studying prognostic factors ([47]). The first reason is to understand how the disease behaves. Is the prognosis similar in men and in women?

Is age an important prognostic factor? Does the extent of a tumour or its histological grade materially influence the outcome? Are the values of certain laboratory tests and the results of physical examination significantly correlated with length of survival? Questions like these and the answers to them are certainly important for understanding the disease.

The second reason is the need to predict survival for groups of patients and to choose treatment accordingly. After several factors are found important separately, the way they act together needs to be understood and the results then can be used to aid doctors to classify patients accurately and to choose treatment correctly.

The third reason is to help researchers design experiments to study new treatments. Knowledge about prognostic factors of a cancer is the basis of any new experiment. What factors should be controlled and what factors should be stratified ? Whether or not one can answer these questions depends on how much one knows about the factors involved.

In this project, a data set on stage I malignant melanoma is analyzed. The main objective is to identify the important prognostic factors for this skin cancer. In chapter 2, a description of the data is given, and in chapter 3, statistical analyses are performed and important prognostic factors identified. In chapter 4, a summary is provided and some areas for further analyses are suggested.

CHAPTER II

DATA

The data to be analyzed are based on reports of all newly diagnosed, histologically confirmed cases of malignant melanoma seen in Western Canada (British Columbia, Alberta, Saskatchewan, Manitoba) from 1 April 1979 to 31 March 1981. The reports were obtained through the cancer registries of the above provinces, and the whole programme is named The Western Canada Melanoma Study ([33]).

Patients were interviewed in their homes by trained interviewers using a standardized questionnaire. Information was obtained on a number of variables, including host pigmentation and reaction to sunlight, residence, occupational history, recreational activities with specific reference to sunlight exposure, medical history, chronic drug use, family history, diet, smoking and alcohol consumption; and for women, reproductive history and use of oral contraceptives and menopausal oestrogens.

A standardized abstract of the medical record was made for each patient and included data on symptoms, treatment and recurrences. Pathological slides were reviewed by a pathologist, Dr. A. J. Worth, at the Cancer Control Agency of British Columbia. Pathological slides were not available for 20% of the patients, and in these cases the original pathology report was used.

In the 2-year period of data intake, 904 patients with primary cutaneous melanomatous lesions were registered. The patients were followed from their registration date, and the times at death and drop-out were recorded. To allow a five-year follow-up time for most of the patients, the termination date for analysis was December, 1986. Most patients were still alive by this date, hence their survival times were censored.

Because of some technical difficulties in a study like this one, not all patients included in the study have complete and correct records on all the variables on the raw data base. Therefore, a decision was made that a pre-analysis be carried out first. 430 patients with complete information on 13 important covariates form the data set for this project.

Of the 430 patients, 78 died, 18 dropped out and 334 had their survival time censored. The 13 covariates are defined below.

Table 2.1 Definitions of the 13 covariates.

---

1. Sex: male and female;
2. Site of tumour location: head and neck, trunk, upper limbs, lower limbs;
3. Depth of tumour invasion: the depth in millimeters of a melanoma measured vertically from the top of the granular layer to the base of the tumour, or the maximum invasion dimension;

4. Clark's levels of invasion:

level I: all melanoma cells restricted to the epidermis;

level II: melanoma cells penetrating the papillary;

level III: melanoma cells filling the papillary dermis;

level IV: melanoma cells extending to the retricular dermis;

level V: melanoma cells invading the subcutaneous tissue.

5. Mitoses:

level I: fewer than 1 per 5 high power microscopic fields;

level II: between 1 per 5 high power fields and 1 per each high power field;

level III: greater than 1 per high power field.

6. Histological cell type: (or growth pattern)

lentigo: the melanoma lesion grows radially for a long period, and may or may not eventually penetrate;

superficial spreading: the melanoma lesion has a disorderly appearance in colour and outline, and tends to grow horizontally before vertically. Also it tends to ulcerate and bleed with growth;

nodular: the melanoma lesion initially grows vertically without a visible pre-existing radial growth phase.

7. Differentiation:

level I: absent to mild lymphocytic infiltration around lesions;

level II: moderate lymphocytic infiltration represented by multiple foci of lymphocytes at the edge and beneath the lesion;

level III: marked lymphocytic infiltration where lymphocytes were confluent occasionally forming bandlike features.

8. Pigmentation:

level I: melanoma cells or macrophages contained no melanin pigment to their cytoplasm;

level II: melanoma cells or macrophages contained minimum melanin pigment to their cytoplasm;

level III: melanoma cells or macrophages contained moderate melanin pigment to their cytoplasm;

level V: melanoma cells or macrophages contained marked melanin pigment to their cytoplasm.

9. Ulceration:

absent: an interruption of the surface epithelium involved by the tumour is not seen;

present: an interruption of the surface epithelium involved by the tumour is seen.

10. Perilymphatic inflammation:

absent: absence of indicator of deep inflammatory response;

present: presence of deep inflammatory response indicating tumour invasion;

11. Regression:

absent: when no signs of increased vascularity with

scattered melaninladen macrophages in the dermis and no signs of fibrosis are seen;

present: when one or both types of the above signs are seen.

12. Age: age = diagnosis year - birth year.

13. Index: index = (depth of tumour invasion) times (number of mitoses).

For the purpose of statistical analysis, the above covariates are coded and shortened names given to them.

Table 2.2 Coding information.

Covariates	levels	Codes	Names	Base-level
Sex	male	1	sex1	sex1
	female	2	sex2	
Site	head and neck	1	site1	
	trunk	2	site2	site2
	upper limbs	3	site3	
	lower limbs	4	site4	
Depth	continuous	/	depth	/
Clark1	level II	2	clark2	clark2
	level III	3	clark3	
	level IV	4	clark4	
	level V	5	clark5	
Mito	level I	1	mito1	mito1
	level II	2	mito2	

	level III	3	mito3	
Cell	lentigo	1	cell1	
	superficial spreading	2	cell2	cell2
	nodular	3	cell3	
Diff	level I	1	diff1	
	level II	2	diff2	diff2
	level III	3	diff3	
Pigm	level 0	0	pigm0	
	level I	1	pigm1	
	level II	2	pigm2	pigm2
	level III	3	pigm3	
Ulce	absent	0	ulce0	ulce0
	present	1	ulce1	
Lymp	absent	0	lymp0	lymp0
	present	1	lymp1	
Regr	absent	0	regr0	regr0
	present	1	regr1	
Age	continuous	/	age	/
Index	continuous	/	index	/

As is well known, if a categorical covariate has k+1 levels, only k dummy variables are needed to represent it. The level without a specific dummy variable matched to it is referred to as the base level. For instance, to analyze the effect of site on survival, only site1, site3 and site4 are included into the model since site2 has been chosen as the base level. The

criterion to choose a level as base level is to allow it to contain enough cases so that comparison referred to it is practically meaningful. For example, there are 375 patients whose tumours did not ulcerate, and only 55 patients whose tumours had ulcerated, so the patients without ulceration are chosen as the base level.

Finally, the clinical staging system of melanoma used by doctors is described briefly here. Stage I melanomas are local disease only, with the primary melanoma present, previously excised, or locally recurrent. Stage II melanomas consist of the primary lesion and palpable regional lymphnodes, and stage III melanomas indicate widespread disease.

The patients covered by the data are all stage I patients.

More will be said about the basic features of the data in the next chapter.

CHAPTER III

STATISTICAL ANALYSIS

3.1 Descriptive Analysis

The data set contains 13 covariates as described in chapter 2. In this section, those covariates will be analyzed descriptively to gain some qualitative / quantitative insight into the problem at hand.

The three continuous covariates, i.e., depth, age and index, have the following characteristics.

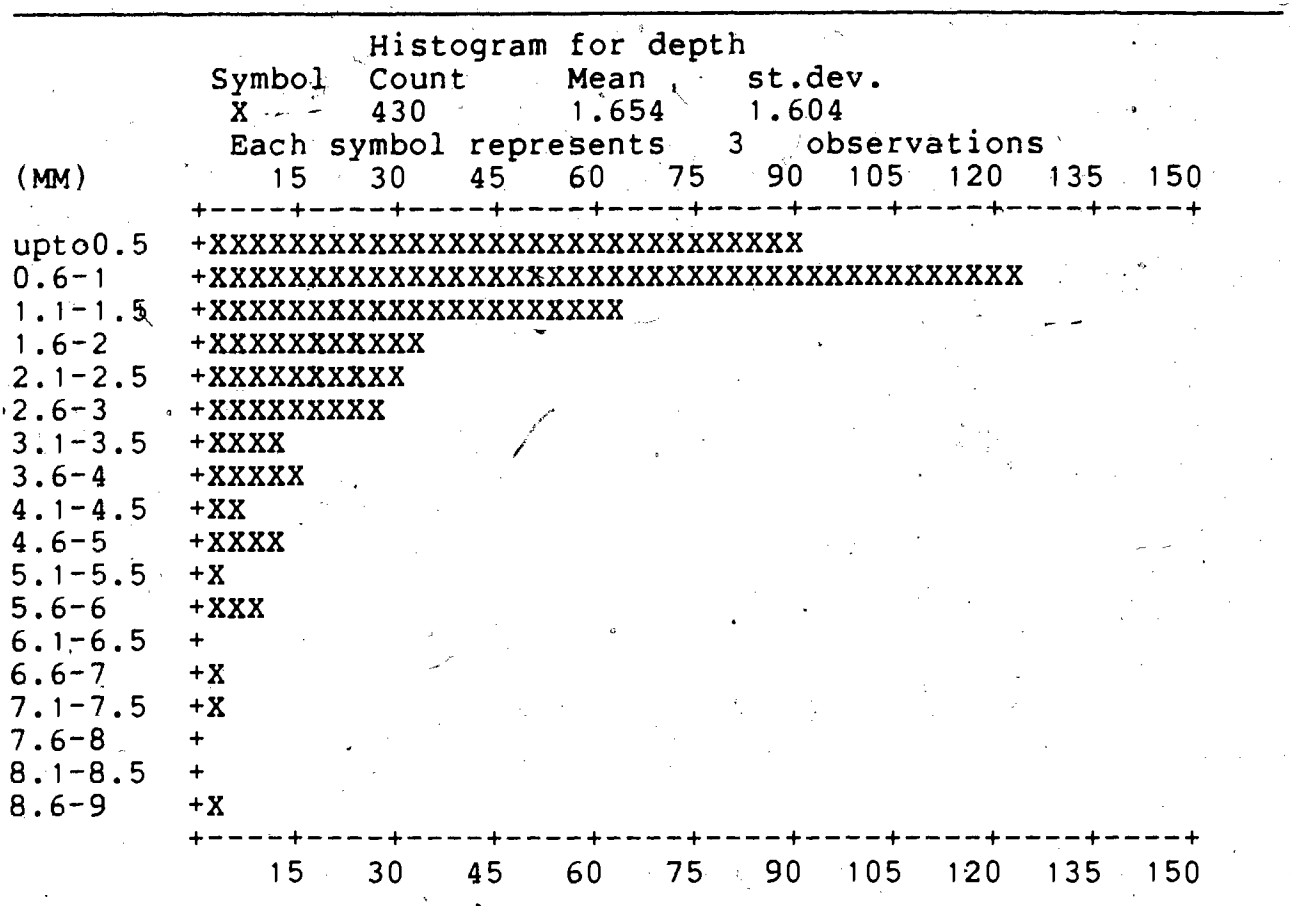
Table 3.1.1 Descriptive analysis of depth, age and index. (Units are mm for Depth, year for Age and mm by number of mitoses for Index.)

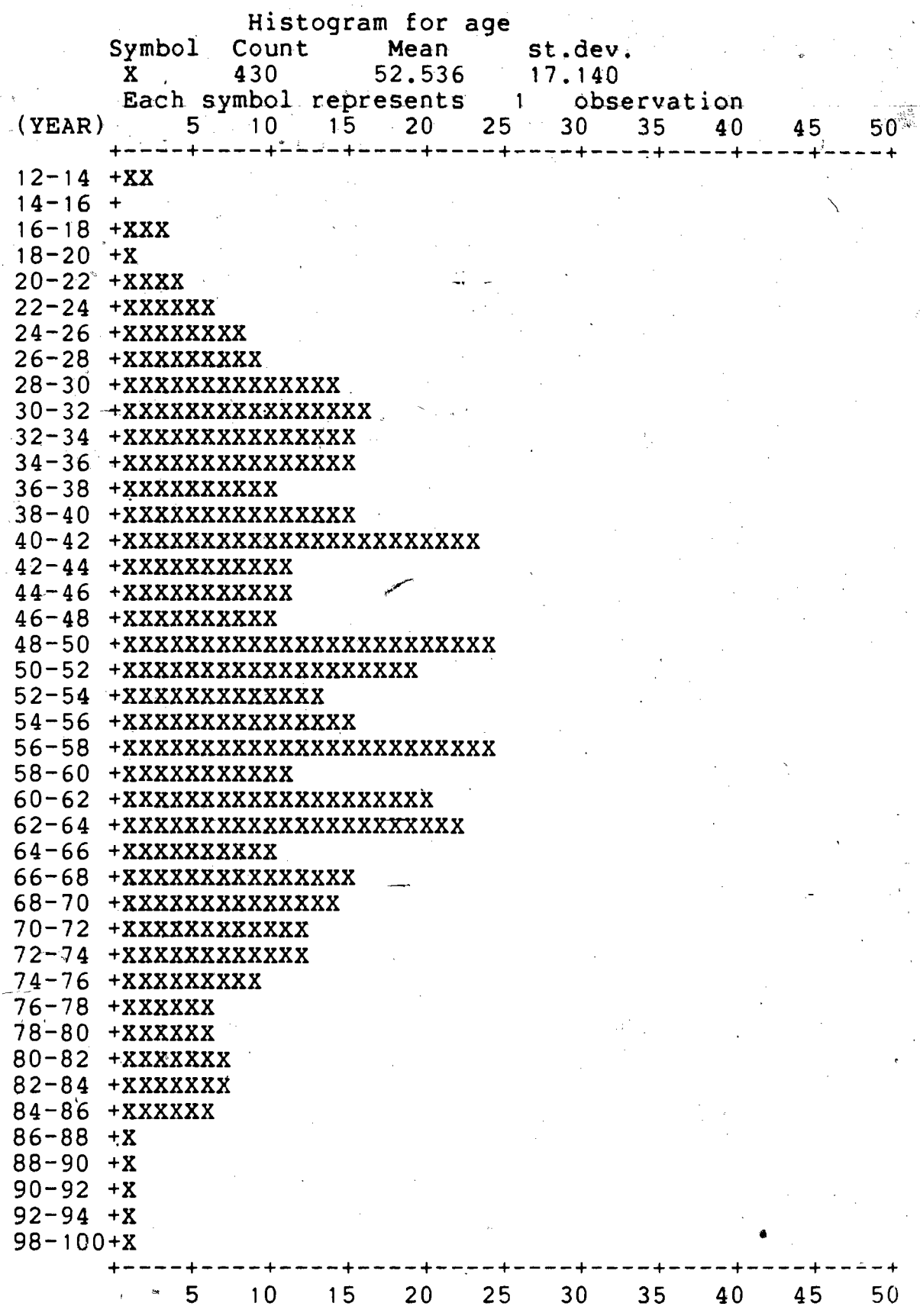
	Depth	Age	Index
minimum	0.05	12	0.05
maximum	9.00	100	27.00
range	8.95	88	26.95
mode	*	*	0.50
median	1.03	52	1.41
mean	1.65	52.54	3.52
st.dev.	1.60	17.14	4.72
skewness	1.86	0.08	2.27
kurtosis	3.60	-0.71	5.55

NOTE: * means that the corresponding mode is not unique.

The histograms of these three covariates are shown in Figure 3.1.1

Figure 3.1.1 Histograms of depth, age and index.




Histogram for age
Symbol Count Mean st.dev.
X 430 52.536 17.140
Each symbol represents 1 observation
(YEAR) 5 10 15 20 25 30 35 40 45 50
12-14 +XX
14-16 +
16-18 +XXX
18-20 +X
20-22 +XXXX
22-24 +XXXXXX
24-26 +XXXXXXXX
26-28 +XXXXXXXXX
28-30 +XXXXXXXXXXXXXX
30-32 +XXXXXXXXXXXXXXXX
32-34 +XXXXXXXXXXXXXXX
34-36 +XXXXXXXXXXXXXXX
36-38 +XXXXXXXXXX
38-40 +XXXXXXXXXXXXXXX
40-42 +XXXXXXXXXXXXXXXXXXXXXXXX
42-44 +XXXXXXXXXXX
44-46 +XXXXXXXXXXX
46-48 +XXXXXXXXXX
48-50 +XXXXXXXXXXXXXXXXXXXXXXXX
50-52 +XXXXXXXXXXXXXXXXXX
52-54 +XXXXXXXXXXXXX
54-56 +XXXXXXXXXXXXXXX
56-58 +XXXXXXXXXXXXXXXXXXXXXXX
58-60 +XXXXXXXXXXX
60-62 +XXXXXXXXXXXXXXXXXXX
62-64 +XXXXXXXXXXXXXXXXXXXXXX
64-66 +XXXXXXXXXX
66-68 +XXXXXXXXXXXXXX
68-70 +XXXXXXXXXXXXXX
70-72 +XXXXXXXXXXXX
72-74 +XXXXXXXXXXXX
74-76 +XXXXXXXXX
76-78 +XXXXXX
78-80 +XXXXXX
80-82 +XXXXXXX
82-84 +XXXXXXX
84-86 +XXXXXX
86-88 +X
88-90 +X
90-92 +X
92-94 +X
98-100+X
5 10 15 20 25 30 35 40 45 50

```
                 Histogram for index
          Symbol  Count       Mean      st.dev.
           X      430         3.515     4.721
           Each symbol represents  4   observations
              20   40    60    80   100   120   140   160   180   200
          +----+----+----+----+----+----+----+----+----+----+
upto1     +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
1.1-2     +XXXXXXXXXXXXXXXX
2.1-3     +XXXXXXXXXXXX
3.1-4     +XXXX
4.1-5     +XXXX
5.1-6     +XXXX
6.1-7     +XXXX
7.1-8     +XXX
8.1-9     +XXX
9.1-10    +XX
10.1-11   +X
11.1-12   +XX
12.1-13   +
13.1-14   +
14.1-15   +XX
15.1-16   +
16.1-17   +
17.1-18   +XX
18.1-19   +
19.1-20   +
20.1-21   +X
21.1-22   +
22.1-23   +
23.1-24   +
24.1-25   +
25.1-26   +
26.1-27   +X
          +----+----+----+----+----+----+----+----+----+----+
              20   40    60    80   100   120   140   160   180   200
```

Most patients are aged 30 years or older (90%), 5 patients are aged under 17 (1.2%), confirming the observations that the disease usually affects middle-aged and older people, and rarely occurs before puberty. For the depth, 66.7% patients had lesions at most 1.65 mm in thickness, and only 12% patients had ~~lesions~~ thicker than 3.50 mm (1.65 is the mean, 3.50 is the "cut point", that is, the point which divides a given range into intervals of

different survival patterns, usually cited in the literature), indicating that thin melanoma is more common in stage I disease than thick melanoma. That the mode does not exist uniquely for depth and age indicates that there is no heavily centered peak for depth or for age.

The ten categorical covariates, namely, sex, site, Clark's levels (clark1), mitoses (mito), celltype (cell), differentiation (diff), pigmentation (pigm), ulceration (ulce), perilymphatic inflammation (lymp), and regression (regr) are described in relation to survival status (that is, the patients are dead or of censored survival time), because survival is the major concern.

Table 3.1.2 Status by Sex

	MALE	FEMALE	TOTAL
DEAD	51	27	78
CENSORED	129	223	352
TOTAL	180	250	430
SEX SPECIFIC DEATH RATE	28%	11%	

Note that there are many more female patients than male patients, but the sex specific death rate is higher with the

males than with the females.

Table 3.1.3 Status by Site

	HDNK	TRUNK	UPLIM	LOLIM	TOTAL
DEAD	23	29	8	18	78
CENSORED	64	101	80	107	352
TOTAL	87	130	88	125	430
SITE SPECIFIC DEATH RATE	26%	22%	9%	14%	

More cases were observed on trunk and lower limbs, while head and neck has the highest site specific death rate.

Table 3.1.4 Status by Clark's Levels

	II	III	IV	V	TOTAL
DEAD	12	15	44	7	78
CENSORED	133	90	123	6	334
TOTAL	145	105	167	13	430
LEVEL SPECIFIC DEATH RATE	8%	14%	26%	54%	

Although Clark's levels measure the same thing as depth does, from a biological point of view, the picture here is clearer than that described by depth alone. For example, one can

see a steady increase in level specific death rates from level II to level V.

Table 3.1.5 Status by Mitoses

	I	II	III	TOTAL
DEAD	29	18	31	78
CENSORED	216	63	73	352
TOTAL	245	81	104	430
LEVEL SPECIFIC DEATH RATE	12%	22%	30%	

In agreement with intuition, the rates of death go up as the numbers of mitoses go up.

Table 3.1.6 Status by Cell Type

	LENTIGO	SUPERFICIAL SPREADING	NODULAR	TOTAL
DEAD	4	43	31	78
CENSORED	29	242	63	334
TOTAL	35	298	97	430
TYPE SPECIFIC DEATH RATE	11%	14%	32%	

Most patients had superficial spreading melanoma, fewer patients had nodular melanoma, and even fewer patients had

lentigo melanoma. However, the type specific death rate is the highest with the nodular type melanoma.

Table 3.1.7 Status by Differentiation

	I	II	III	TOTAL
DEAD	11	62	5	78
CENSORED	33	229	90	352
TOTAL	44	291	95	430
LEVEL SPECIFIC DEATH RATE	25%	21%	5%	

The observation is in accordance with biological knowledge, that is, the death rate decreases as the level of differentiation increases.

Table 3.1.8 Status by Pigmentation

	0	I	II	III	TOTAL
DEAD	3	20	38	17	78
CENSORED	14	114	131	93	352
TOTAL	17	134	169	110	430
LEVEL SPECIFIC DEATH RATE	18%	4%	22%	15%	

There is no clear explanation with this table how the degree of pigmentation affects survival.

The next three tables describe the last three categorical covariates.

Table 3.1.9 Status by Ulceration

	NO	YES	TOTAL
DEAD	58	20	78
CENSORED	317	35	352
TOTAL	375	55	430
ULCE. SPECIFIC DEATH RATE	15%	36%	

Table 3.1.10 Status by Perilymphatic Inflammation

	NO	YES	TOTAL
DEAD	49	29	78
CENSORED	274	78	352
TOTAL	323	107	430
LYMP. SPECIFIC DEATH RATE	15%	36%	

Table 3.1.11 Status by Regression

	NO	YES	TOTAL
DEAD	63	15	78
CENSORED	253	99	352
TOTAL	316	114	430

REGR. SPECIFIC		
DEATH RATE	20%	13%

Table 3.1.9 and 3.1.10 indicate that ulceration and perilymphatic inflammation are signs of poor prognosis, while Table 3.1.11 indicates that regression is a sign of good prognosis.

Two more tables of interest are presented below. The cut points used for age and depth are based on those seen in the literature. ([3], [12] and [22]).

Table 3.1.12 Status by Age

	under 39	40-64	over 65	TOTAL
DEAD	16	10	52	78
CENSORED	90	81	181	352
TOTAL	106	91	233	430
AGE SPECIFIC				
DEATH RATE	15%	11%	22%	

Patients under 39 years old and patients aged between 40 and 64 have similar death rates, but for patients over 65 years old, the death rate increases sharply.

For depth, a similar picture to that for Clark's levels can be seen. That is, the thicker the invasion depth, the larger the

death rate.

Table 3.1.13 Status by Depth

	≤0.74	0.75-1.49	1.50-3.49	≥3.50	TOTAL
DEAD	12	15	28	23	78
CENSORED	151	93	77	31	352
TOTAL	163	108	105	54	430
DEPTH SPECIFIC DEATH RATE	7%	14%	27%	43%	

For another aspect of the data, it might be interesting to have a look at the registration pattern.

Since the data were collected in two years, it is difficult to identify a seasonal pattern, but winter does seem to be the season with the lowest number of cases reported (Figure 3.1.2).

Finally, the estimated survival curve for the whole data set is shown in Figure 3.1.3. The overall five-year survival (percentage) is 83.04%. The methods of obtaining estimated survival curves are given in section 3.3.

3.2 Screening for Prognostic Factors

When the clinical experiment was designed, many potential prognostic factors were included. The ideal approach to determine the important prognostic factors is to treat

Figure 3.1.2 Registration Pattern

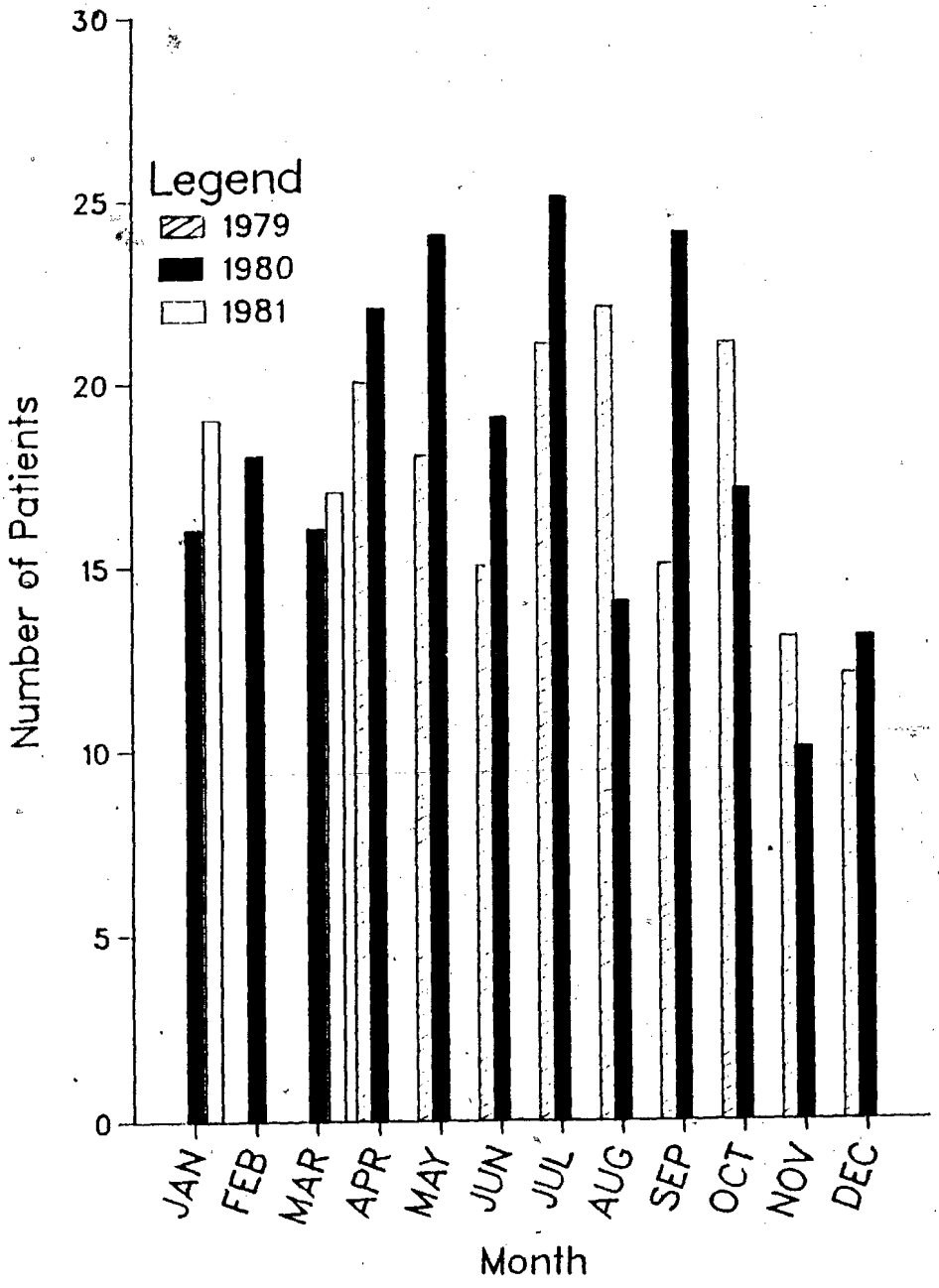
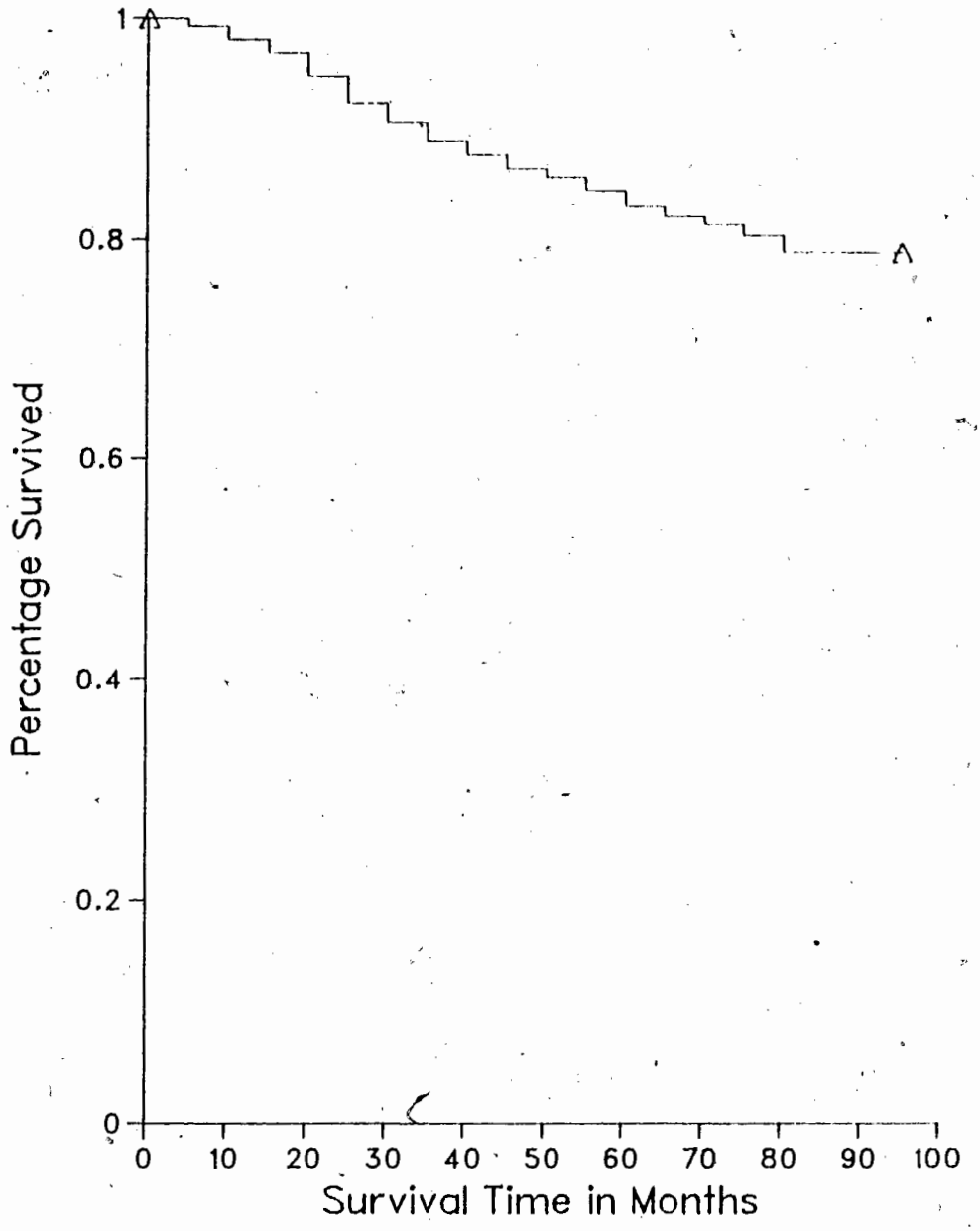


Figure 3.1.3 Estimated Survival Curve



information from all factors simultaneously, because all factors acted on the patients in that way. However, experience shows that analyzing too many unimportant factors together with the important ones will not only slow down progress but hinder one from gaining insight into the nature of the problem as well. Therefore, some selective analysis is useful for getting rid of those totally unimportant factors, so that one can start simultaneous analyses with a reasonable set of potential prognostic factors. This process is called screening for prognostic factors. ([47]).

In this section, the 13 covariates are analyzed one by one to see which of them contribute significantly to the survival of the patients. For both the continuous and the categorical covariates, the rationale of the analysis is the same: if over different ranges of a continuous covariate or over different levels of a categorical covariate, the survival of malignant melanoma patients changes only within random variation, then the covariate is of little value in predicting survival; otherwise if changes to survival are so large that they can not be explained by random fluctuation alone, the implication is that the covariate correlates to survival in a significant way.

For the continuous covariates, there is a question of how to choose cutpoints to perform grouping. The cutpoints seen in the literature for depth and age are used, and for index the first quartile, mean and the third quartile are used. The issue of cutpoint will be discussed more fully in section 3.4.3.

In order to judge when the changes of survival are significant, Mantel's test and Breslow's test are used ([6],[43]). Both tests can be viewed as analogues of nonparametric rank tests. When there are no censored observations, Mantel's test is essentially an exponential scores test. Breslow's test is a version of the Kruskal-Wallis test for censored data. The difference between the two tests is that Breslow's test puts more weight on early observations and is less sensitive to late events which occur when few patients in the study remain alive. Both tests used here are based on large sample theory, and on null hypotheses the test statistics have approximate chi-square distributions with degrees of freedom shown. For details, see section 3.3.

The results of the analyses are summarized in Table 3.2.1 and Figure 3.2.1.

Table 3.2.1 Summary of univariate analysis (The first statistic is Breslow's, the second statistic is Mantel's)

Covariates	statistics	df	P-values
Sex	21.481	1	0.0000
	23.527	1	0.0000
Site	13.725	3	0.0033
	12.553	3	0.0057
Depth	67.349	3	0.0000
	71.836	3	0.0000
Clark1	34.428	3	0.0000

	37.621	3	0.0000
Mitoses	19.239	2	0.0001
	19.115	2	0.0001
Celltype	19.581	2	0.0001
	19.118	2	0.0000
Diff	13.421	2	0.0012
	14.440	2	0.0007
Pigm	2.352	3	0.5026
	3.028	3	0.3874
Ulc	15.446	1	0.0001
	17.562	1	0.0001
Lymp	7.098	1	0.0077
	7.772	1	0.0053
Regr	2.759	1	0.0967
	2.219	1	0.1363
Age	17.798	2	0.0001
	20.056	2	0.0000
Index	8.195	2	0.0166
	8.151	2	0.0170

All the covariates are significant (α = 0.05) except pigmentation and regression. Among those significant covariates, sex, depth, Clark's levels, ulceration and age are highly significant. (α = 0.005).

Before 1971, survival analysis relied solely upon univariate techniques, and many factors were reported to be prognostic

Figure 3.2.1 Univariate analysis: by sex

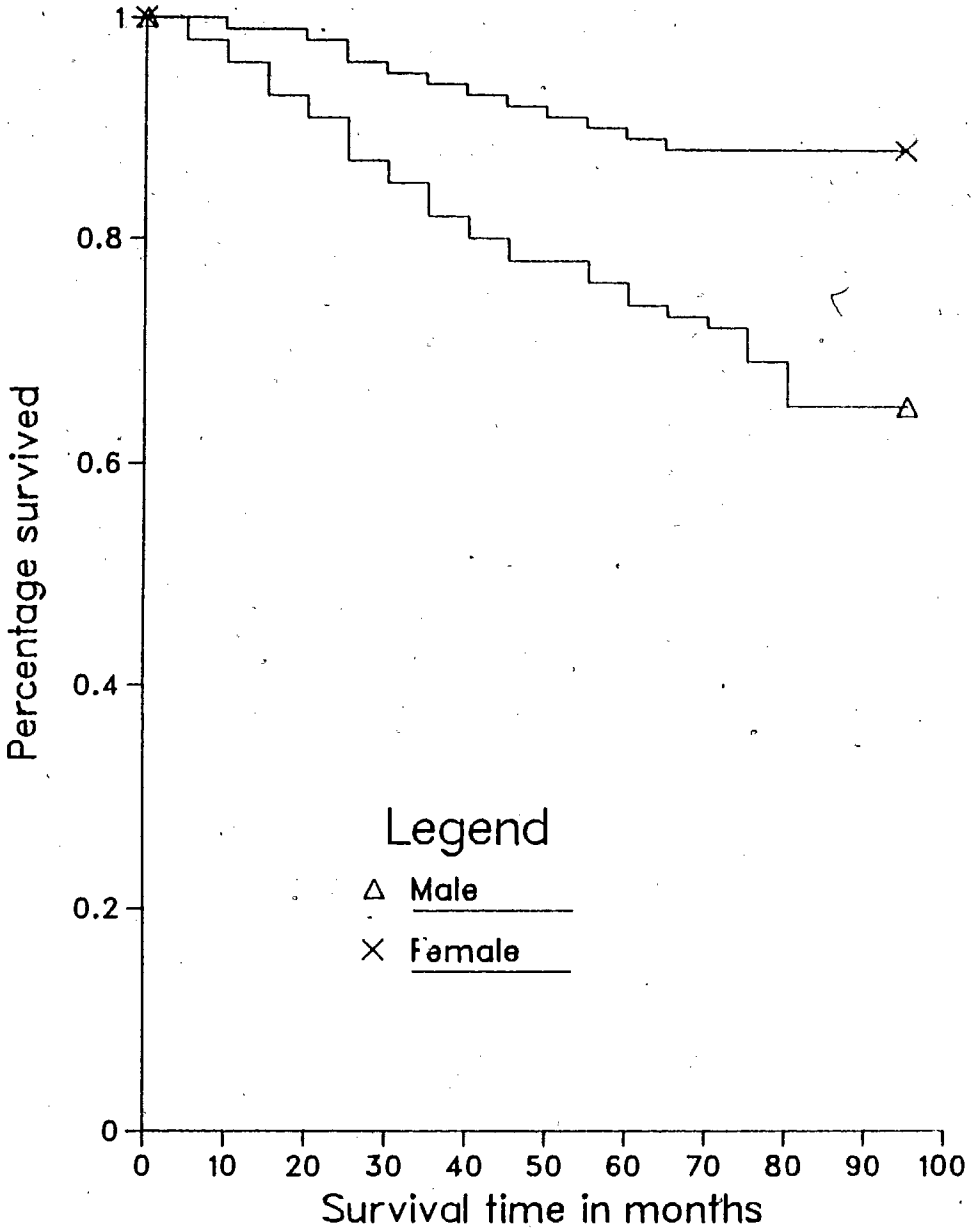


Figure 3.2.1 Univariate analysis: by age

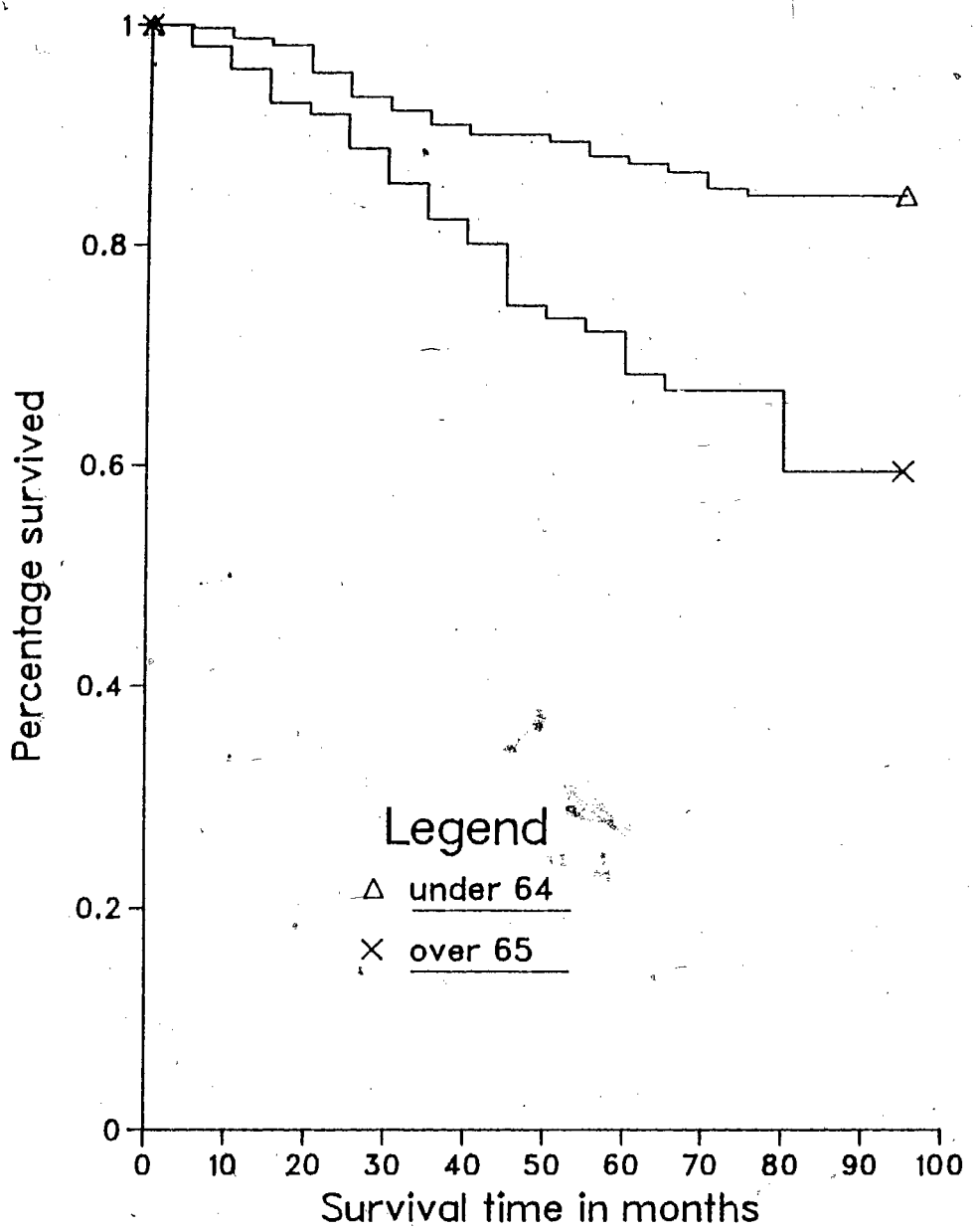
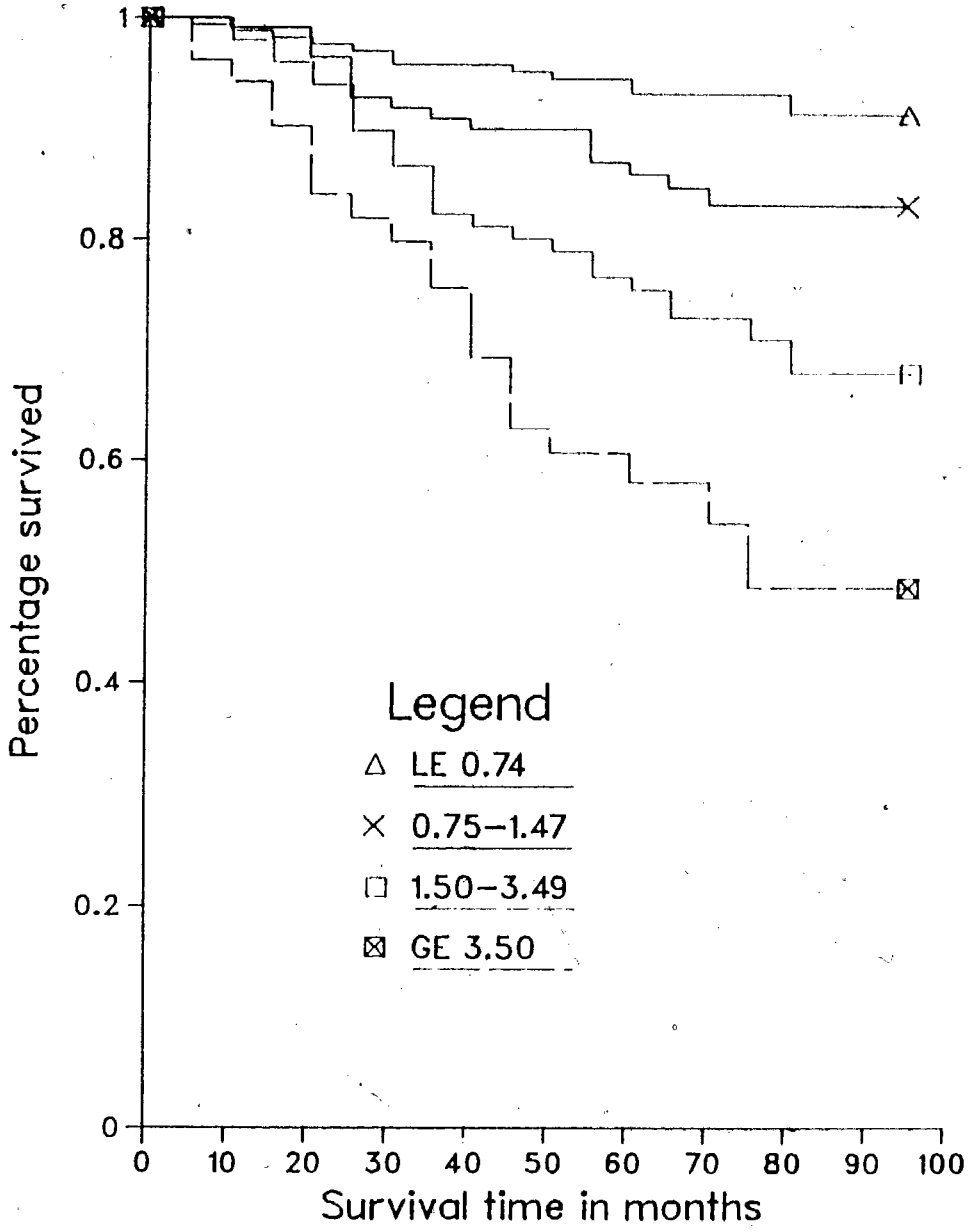


Figure 3.2.1 Univariate analysis: by depth



indicators of survival.

However, different factors are acting on a single patient at the same time. Moreover, in a large scale study like the present one, some covariates are likely to be correlated. If this is so, it is useful to know which covariates are highly correlated, so that they can be treated properly later when fitting regression models (e.g., to deal with the multicollinearity problem). Also, association between covariates might explain the prognostic value of different covariates. With this in mind, an association analysis of the 13 covariates is performed.

Seventy-eight two-way tables are examined. For each table, a test for independence of classification (Pearson's chi-square) is carried out. Moreover, Goodman-Kruskal's λ and λ_{yx} (λ_{xy}) are used to measure the association between two covariates X and Y when X and Y are of equal importance (λ) or when Y depends on X (λ_{yx}), respectively. The test procedure is described in section 3.3.

The Goodman-Kruskal's association measures are employed here because they fit the problem nicely. For example, if one can accurately predict ulceration level (yes, no) based on tumour depth, and tumour depth is found to be an important prognostic factor, then the way in which ulceration influences survival will be better understood.

In summary table 3.2.2, a "1" means both the independence test and the association tests are significant (at $\alpha = 0.05$) for

Table 3.2.2 Association Analysis

	sex	sit	dep	cla	mit	cel	dif	pig	ulc	lym	reg	age	ind
sex	1	0	0	0	0	0	0	0	0	0	0	0	
sit		0	1	0	1	0	0	0	0	0	0	0	
dep			2	1	1	1	0	1	1	0	1	2	
cla				1	1	1	1	0	1	1	1	2	
mit					1	1	0	1	0	0	0	1	
cel						0	0	1	0	0	0	1	
dif							0	0	0	0	0	1	
pig								0	0	0	0	0	
ulc									0	0	0	1	
lym										0	0	0	
reg											0	0	
age												0	
ind													

the pair of covariates involved, a "2" means the association is very strong (P-values $\simeq$ 0), and a "0" means there is no empirical evidence of association.

As expected, depth and Clark's levels are highly correlated, so are depth and index, and Clark's levels and index. Also depth correlates with all the covariates except sex, site, pigmentation and regression; Clark's levels correlates with all the covariates except sex and ulceration. Since depth and Clark's levels are frequently reported to be significant prognostic factors, the above series of associations suggest

that the other covariates influence survival through their associations with these important factors. The association between sex and site is a confirmation of the observation that women have their lesions mostly on lower limbs, while men have lesions mostly on the trunk.

The high association between depth and Clark's levels, together with their biological meanings, suggests that it may be more appropriate to treat depth and Clark's levels separately and fit models with only one of them included.

3.3 Statistical Methodology

This section provides the statistical techniques for fitting the Cox regression model. For ease of reference, the statistical methods used in the previous two sections are also described here.

1 Product Limit (or Kaplan-Meier) Estimate of Survival Curve.

Suppose that there are observations on n individuals and that deaths occur at k time points. Let m_i be the number of deaths at time t_i; let r_i be the number of individuals at risk at t_i, that is, the number of individuals alive and uncensored just prior to t_i. Then the product-limit estimate of the probability of surviving past time t, denoted S(t), is

$$\hat{S}(t) = \prod_{i:t_i<t} (r_i - m_i)/r_i .$$

If a censored time equals a death time t_i, the convention that

the censored individual is included in the set of r_i individuals at risk at t_i is adopted, since an individual censored at time t almost certainly survives past t.

As a function of t, $\hat{S}(t)$ is a right-continuous step function with $\hat{S}(0) = 1$. The heights of the k steps corresponding to the k distinct death times can be computed recursively as below:

$$\hat{S}(t_1+0) = (r_1 - m_1)/r_1,$$

$$\hat{S}(t_i+0) = \hat{S}(t_{i-1}+0)(r_i - m_i)/r_i, \qquad i = 2,3,\cdots,k.$$

When the largest observed time is a censored time, $\hat{S}(t)$ remains undefined beyond this time point. ([31],[35],[36],[41]).

Example 1. The following data are the lengths of remission times (in weeks) for a group of acute leukemia patients. The starred quantities denote censored observations and 6-MP stands for the drug 6-mercaptopurine taken by the patients.

Data for illustrations

6-MP																			
6	6	6	6*	7	9*	10	10*	11*	13	16	17*	19*	20*	22	23	25*	32*	34*	35*

The calculations leading to the product-limit estimate of the survival curve for these patients are outlined in Table 3.3.1.

Table 3.3.1 Estimated survival curve for the 6-MP data using the product-limit method

t_i	r_i	m_i	$(r_i - m_i)/r_i$	$\hat{S}(t_i+0)$
6	21	3	0.857	0.857
7	17	1	0.947	0.807
10	15	1	0.933	0.573
13	12	1	0.917	0.690
16	11	1	0.909	0.627
22	7	1	0.871	0.538
23	6	1	0.833	0.448

Note that $\hat{S}(t)$ is well defined up to $t = 35$ only.

2 Life Table (or Actuary) Estimate of Survival Curve

Suppose that the range of survival time is divided into m intervals. This method describes the whole survival experience in terms of interval survival experience.

Let

n_i = number of patients entering the i-th interval,

c_i = number of patients censored in the i-th interval,

d_i = number of patients dying in the i-th interval.

Under the assumption that in each interval censored observations occur randomly (that is, are distributed uniformly), the patients who are censored are considered to be at risk for half

of the interval time. Thus, the number of patients at risk during the i-th interval is estimated by

$$r_i = n_i - \frac{1}{2}c_i,$$

and given that an individual enters the i-th interval, the conditional probability of his / her dying in the i-th interval is estimated by

$$q_i = d_i/r_i,$$

assuming that the mortality rate is constant for each of the m intervals.

The conditional probability of surviving the i-th interval given that the individual enters the i-th interval is estimated by

$$p_i = 1 - q_i,$$

and the cumulative survival function, usually called the survival curve and meaning the cumulative proportion of patients surviving to, say, the beginning of the i-th interval, is estimated by

$$P_i = p_i P_{i-1},$$

where $P_1 = 1$ and $i = 2,3,\cdots,m$. ([26],[31],[41]).

Example 2. The 6-MP data given in example 1 are used to illustrate the life-table method (Table 3.3.2). I_i represents the i-th interval of survival time.

3 K-sample Mantel's Test

Let k be the number of groups of (or categories of a covariate for) individuals whose survival distributions are to

Table 3.3.2 Estimated survival curve for the 6-MP data using the life-table method

I_i	n_i	c_i	d_i	r_i	q_i	p_i	P_i
[0,4)	21	0	0	21	0	1	1
[4,8)	21	1	4	20.5	0.195	0.805	0.805
[8,12)	16	3	1	14.5	0.069	0.931	0.749
[12,16)	12	0	1	12	0.083	0.917	0.687
[16,20)	11	2	1	10	0.100	0.900	0.619
[20,24)	8	1	2	7.5	0.267	0.733	0.453
[24,28)	5	1	0	4.5	0	1	0.453
[28,32)	4	0	0	4	0	1	0.453
[32,36)	4	3	0	2.5	0	1	0.453

be compared. Let $t_1 < t_2 < \cdots < t_h$ be the times at which deaths occurred among the k groups and let n be the total number of individuals. The null hypothesis is that the k groups have the same survival distribution.

At time t_i, let n_{ij} be the number of individuals in group j in the study (that is, whose observation time t is greater than or equal to t_i). Let x_{ij} be the number of individuals who died at exactly time t_i in group j. (If there are no tied times, x_{ij} is zero for all but one group; $x_{ij} = 1$ for the group where death occurs.)

Conditioned on the n_{ij} and the sum $x_{i+} = \sum_{j=1}^{k} x_{ij}$, the vector $\underset{\sim}{x}_i = (x_{i1},\cdots,x_{i(k-1)})^T$ has a $k-1$ dimensional hypergeometric distribution with mean vector

$$E(\underset{\sim}{x}_i) = (E(x_{i1}),\cdots,E(x_{i(k-1)}))^T,$$

where $E(x_{ij}) = (x_{i+}n_{ij})/n_{i+}$, and $i = 1,2,\cdots,h$, $j = 1,2,\cdots,k-1$.

The covariance matrix $\underset{\sim}{V}_i$ of $\underset{\sim}{x}_i$ has elements

$$\text{cov}(x_{ij},x_{il}) = \frac{n_{ij}(\delta_{jl} - n_{il}/n_{i+})x_{i+}(n_{i+} - x_{i+})}{n_{i+}(n_{i+} - 1)},$$

where $\delta_{jl} = 1$ if $j = l$ and $\delta_{jl} = 0$ if $j \neq l$, $n_{i+} = \sum_{j=1}^{k} n_{ij}$, $j, l = 1,2,\cdots,k-1$.

Let

$$\underset{\sim}{E} = \sum_{i=1}^{h} E(\underset{\sim}{x}_i)$$

$$\underset{\sim}{Q} = \sum_{i=1}^{h} \underset{\sim}{x}_i,$$

$$\underset{\sim}{V} = \sum_{i=1}^{h} \underset{\sim}{V}_i.$$

Then the j-th element of $\underset{\sim}{Q} - \underset{\sim}{E}$ is like

$$\sum_{i=1}^{h} (x_{ij} - x_{i+}n_{ij}/n_{i+}).$$

Mantel's test statistic is

$$X_M^2 = (\underset{\sim}{Q} - \underset{\sim}{E})^T \underset{\sim}{V}^{-1} (\underset{\sim}{Q} - \underset{\sim}{E}),$$

which is asymptotically distributed as chi-square with $k-1$ degrees of freedom and large values of X_M^2 indicate that the null hypothesis is false. ([26],[41]).

Example 3. Consider the following three groups of hypothetical data. Starred numbers are censored observations.

group 1: 1 3* 4 5* 10*

group 2: 2* 3 7* 9* 13*

group 3: 2 5 8* 8* 11*

The Mantel test statistic can be calculated as below.

t_i	n_{i1}	n_{i2}	n_{i3}	n_{i+}	x_{i1}	x_{i2}	x_{i3}	x_{i+}
1	5	5	5	15	1	0	0	1
2	4	5	4	13	0	0	1	1
3	4	4	4	12	0	1	0	1
4	3	4	4	11	1	0	0	1
5	2	4	4	10	0	0	1	1

$h = 5$ $k = 3$ $n = 15$

$\underline{x}_1 = (1, 0)^T$ $E(\underline{x}_1) = (5/15, 5/15)^T$

$\underline{x}_2 = (0, 0)^T$ $E(\underline{x}_2) = (4/13, 5/13)^T$

$\underline{x}_3 = (0, 1)^T$ $E(\underline{x}_3) = (4/12, 4/12)^T$

$\underline{x}_4 = (1, 0)^T$ $E(\underline{x}_4) = (3/11, 4/11)^T$

$\underline{x}_5 = (0, 0)^T$ $E(\underline{x}_5) = (2/10, 4/10)^T$

$\underline{Q} = (2, 1)^T$ $\underline{E} = (1.447, 1.815)^T$

The two by two matrices $\underline{V}_i$, *et al.*, are as below.

$\underline{V}_1$		$\underline{V}_2$		$\underline{V}_3$		$\underline{V}$	
2/9	-1/9	36/169	-20/169	2/9	-1/9	24/121	-12/121
-1/9	2/9	-20/169	40/169	-1/97	2/9	-12/121	28/121

$\underline{V}_5$		$\underline{V}$		$\underline{V}^{-1}$	
4/25	-2/25	1.016	-0.520	1.280	-0.557
-2/25	6/25	-0.520	1.153	-0.557	1.128

Therefore, $X_M^2 = 0.6201$.

4 K-sample Breslow's Test

With the same notation used above, let

$$w_j = \sum_{i=1}^{h} \frac{n_{i+}}{n+1}(x_{ij} - x_{i+}n_{ij}/n_{i+}),$$

where $j = 1,2,\cdots,k-1$.

Then the vector $\underline{W} = (w_1,\cdots,w_{k-1})^T$ has zero mean and covariance matrix

$$\underline{V}_w = \sum_{i=1}^{h} (\frac{n_{i+}}{n+1})^2 \underline{V}_i .$$

Breslow's test statistic is

$$X_B^2 = \underline{W}^T \underline{V}_w^{-1} \underline{W},$$

which is also asymptotically distributed as chi-square with k-1 degrees of freedom and large values of X_B^2 indicate that the null hypothesis is false.

Note that the Mantel test puts equal weight 1 to the terms $x_{ij}-x_{i+}n_{ij}/n_{i+}$, while the Breslow test gives decreasing weights $n_{i+}/(n+1)$ to the same terms. ([26],[41]).

Example 4. Continue example 3, the Breslow test statistic is calculated as below.

$w_1 = (15/16)[1-5/15]+(13/16)[-4/13]+(12/16)[-4/12]$
$+(11/16)[1-3/11]+(10/16)[-2/10] = 0.5$.

$w_2 = (15/16)[-5/15]+(13/16)[-5/13]+(12/16)[1-4/12]$
$+(11/16)[-4/11]+(10/16)[-4/10] = -0.625$

$\underline{W} = (0.5, -0.625)^T$

$$\underline{V}_w = (\frac{15}{16})^2 \underline{V}_1 + (\frac{13}{16})^2 \underline{V}_2 + (\frac{12}{16})^2 \underline{V}_3 + (\frac{11}{16})^2 \underline{V}_4 + (\frac{10}{16})^2 \underline{V}_5$$

Finally, $X_B^2 = 0.6679$.

5 The Goodman-Kruskal Test

Let x stand for the row and y stand for the column of an r by c contingency table. Let a_{ij} be the entries of the table (i = 1,2,···,r and j = 1,2,···,c). Let $r_i = \sum_{j=1}^{c} a_{ij}$ be the row totals, $c_j = \sum_{i=1}^{r} a_{ij}$ be the column totals and N be the table total.

The statistic λ_{yx} involves a comparison of the following two situations: an individual is chosen at random from the population and one is asked to guess to which y category the individual belongs, either (a) given no further information or (b) given the individual's x category. If x and y are totally uncorrelated then one can do no better in the second situation than one can do in the first, but otherwise there will be an improvement. The measure λ_{yx} quantifies this improvement as the relative decrease in the probability of error in guessing the y category as between the two situations under the assumption that the guess consists of selecting the most likely of the y categories on each occasion. The formula for λ_{yx} is

$$\lambda_{yx} = (\sum_{i=1}^{r} \max_j a_{ij} - \max_j c_j)/(N - \max_j c_j).$$

An approximate formula for the variance of λ_{yx} is

$$(N-\sum_{i=1}^{r} \max_j a_{ij})(\sum_{i=1}^{r} \max_j a_{ij}+\max_j c_j-2\Sigma\max_j a_{ij})/(N-\max_j c_j)^3,$$

where Σ denotes the summation over those values of i such that $\max_j a_{ij}$ occurs in the column which has the the largest column total. For large N, the ratio of λ_{yx} to the approximate standard error has asymptotically a standard normal distribution.

Exchanging the roles of x and y will give λ_{xy}, that is, the statistic which measures the association between x and y by means of predicting x based on y.

When x and y are of equal importance, the appropriate statistic is λ, which is a combination of λ_{xy} and λ_{yx} (combined in a symmetrical manner) and is defined as

$$\lambda = \frac{\sum_{i=1}^{r} \max_j a_{ij} + \sum_{j=1}^{c} \max_i a_{ij} - \max_j c_j - \max_i r_i}{2N - \max_j c_j - \max_i r_i}.$$

The approximate variance for λ is very complicated and is not presented here. ([26],[63]).

6 The Cox Regression Model

The model to be used in the remainder of this project is called *The Cox Proportional Hazards Regression Model* for survival or failure data analysis (Cox 1972). Some basic terms are defined below.

Suppose that the random variable T denotes the time elapsed between some specified event (in this project, this is the diagnosis date) and the time at death (failure), or at drop-out, or at the time the study terminates, of a living organism or an inanimate device. In the present context, T is called the survival time. The cumulative distribution function of T is defined as

$$F(t) = \text{Prob}(T \leq t) \qquad 0 \leq t < +\infty$$

$$= \text{Prob(death by time } t);$$

the survival function of T as

$$S(t) = \text{Prob}(T > t) \qquad 0 \leq t < +\infty$$

$$= \text{Prob(surviving past time } t);$$

the density function of T as

$$f(t) = \frac{dF(t)}{dt};$$

the hazard rate function of T as

$$h(t) = \frac{f(t)}{S(t)};$$

and the cumulative hazard function of T as

$$H(t) = \int_0^t h(u)du = -\ln S(t).$$

For a more advanced treatment, see Kalbfleisch and Prentice (1980).

In survival analysis, h(t) is the function which plays a fundamental role [20]. For this reason other functions are usually expressed in terms of h(t), e.g.,

$$\hat{S}(t) = \exp(-\int_0^t h(u)du).$$

The Cox proportional hazards regression model is also formulated in terms of h(t). A new feature here is the introduction of a number of covariates denoted by vector $\underline{z} = (z_1, \cdots, z_s)^T$, which is introduced in the hope that it can explain h(t). There are various ways of doing this, but only the most commonly used form of the Cox model, namely, the exponential form is described here and is used later.

Let $h(t;\underline{z})$ be the hazard rate for an individual with covariate vector $\underline{z}$. The exponential form expresses $h(t;\underline{z})$ as

$$h(t;\underline{z}) = h_0(t)\exp(\beta^T\underline{z}) \qquad (1)$$

where β is a column vector of unknown regression coefficients and $h_0(t)$ is an unknown underlying hazard rate function for an individual with covariate vector $\underline{z} = 0$. Notice that no parametric model is assumed for the underlying hazard function; it is completely arbitrary.

The above model contains two implicit assumptions.

Assumption I: The relationship betwen the underlying hazard function and the covariates is *multiplicative* as shown in (1). Thus the ratio of the hazard functions for two individuals with different covariate vectors does not depend on time. For this reason, this assumption is usually called the *proportionality assumption.*

Assumption II: The effect of the covariates on the hazard rate function is of the exponential form also shown in (1).

Of the two assumptions, the first one is more basic, since it describes the general manner in which the covariates act upon the hazard rate function. The second one is in fact chosen for convenience, as the exponential form is just one of the many possible forms. However, the exponential form is quite flexible, in particular, it is good at describing increasing / decreasing phenomena such as survival.

Estimates of the regression coefficients are obtained as below.

Suppose that $t_1 < t_2 < \cdots < t_k$ are the k distinct times to

death of the individuals $1_1, 1_2, \cdots, 1_k$ among n individuals in the study. Let R_i be the group of individuals at risk just prior to time t_i (called risk group at t_i). Then given that only a single death can occur at t_i and that the risk group is R_i, the conditional probability that individual 1_i with covariate vector z_i dies at t_i is

$$\exp(\beta^T z_i) / \sum_{j \in R_i} \exp(\beta^T z_j).$$

Multiplying these probabilities together for each of the k death times gives the partial likelihood function (Cox 1975)

$$L(\beta) = \prod_{i=1}^{k} (\exp(\beta^T z_i)) / (\sum_{j \in R_i} \exp(\beta^T z_j)),$$

and maximizing the partial likelihood function with respect to β yields estimators of β with similar asymptotic properties to those of the usual maximum likelihood estimators ([19],[20],[35]). If there are ties among the death times, say, there are m_i deaths at time t_i, let s_i be the vector sum of the covariates of the m_i individuals. A modified likelihood function ([7])

$$L(\beta) = \prod_{i=1}^{k} \exp(\beta^T s_i) / [\sum_{j \in R_i} \exp(\beta^T z_j)]^{m_i}$$

is then maximized.

7 A Graphical Method for Checking the Proportionality Assumption

When fitting the Cox model to the data, it is necessary to know whether the two assumptions, especially the proportionality assumption, hold or not. There is a graphical method for this

purpose, in which $\ln(-\ln\hat{S}(t;\bar{z}))$ is plotted against survival time t, where $\hat{S}(t;\bar{z})$ is the Kaplan-Meier estimate of $S(t;\bar{z})$ and $\bar{z}$ is the mean vector of the covariates of a certain stratum, when an (independent) covariate is stratified. (For a categorical covariate, its categories form natural strata; for a continuous covariate, one needs to choose appropriate cut points.) If the proportionality assumption holds, the plot should exhibit constant differences between strata because

$$\ln(-\ln S(t;\bar{z})) = \beta^T\bar{z} + \ln(-\ln S_0(t)).$$

In this project, the plotting is done for categorical covariates only ([35],[37],[40]).

8 The Global Chi-square Test of Fit

Whenever a model is fitted to the data, it is important to know whether the fit is good so that valid statements can be made about the fitted model. This is, for the time being, only a partially solved problem. A ready-to-use method is called the global chi-square test, which tests the null hypothesis that all regression coefficients are identically zero.

Let U(0) represent the vector of first derivatives (with respect to β) of the partial likelihood function evaluated at $\beta = 0$; let I(0) represent the observed information matrix evaluated at $\beta = 0$. Then under the null hypothesis

$$X^2 = U^T(0)I^{-1}(0)U(0)$$

has asymptotically a chi-square distribution with degrees of freedom equal to the number of covariates in the model. Large

values of X^2 imply that the null hypothesis is not true ([26],[35]).

9 Significance Tests

Three large sample significance tests are available for deciding which covariates are significant.

The first is the maximum partial likelihood ratio test. The log partial likelihood is maximized first under the full model (all covariates) and then under the restricted model (maximization is restricted to covariates not being tested). The difference between the above two log likelihoods times 2 is the test statistic, that is,

$$\text{likelihood ratio} = 2[\ln L(\hat{\beta}_{Full}) - \ln L(\hat{\beta}_{Restricted})].$$

The second test is the Wald test based on the asymptotic normality property of maximum likelihood estimates.

Let $\hat{\beta}^*$ represent the subset of MLE's (obtained under the full model) corresponding to the coefficients to be tested. The Wald test statistic is

$$\text{Wald} = \hat{\beta}^{*T} I(\hat{\beta}^*) \hat{\beta}^*.$$

The third test is called the score function test, which is built on the derivatives of the partial likelihood function.

Let $U(\beta_0)$ and $I(\beta_0)$ represent the vector of first derivatives and the observed information matrix. These functions are then evaluated using parameter estimates calculated under

the restricted model (the coefficients to be tested are constrained to zero). The scores test statistic is

$$\text{score} = U^T(\beta_0)I^{-1}(\beta_0)U(\beta_0).$$

All three statistics are compared with the chi-square distribution with degrees of freedom equal to the number of parameters being tested. The asymptotic distributions of the likelihood ratio statistic and the score function statistic have not been proved to be chi-square. However, the three statistics generally give close results for large samples. For small samples, the likelihood ratio and score tests usually produce similar values, but the Wald test is often different, because it depends critically on the normality of β estimates, whereas the other tests do not ([26],[53]).

10 Testing Time-dependent Covariates to Check the Proportionality Assumption

Suppose that a categorical covariate has k+1 levels. To include this covariate into the Cox model, k dummy variables are generated. For each of the k dummy variables, a time-dependent covariate can be introduced. Then one can fit a model containing the k dummy variables and the k time-dependent covariates and test the hypothesis that the coefficients corresponding to the k time-dependent covariates are all zero. The proportionality assumption is acceptable if the above test accepts the null hypothesis.

For a continuous covariate, time-dependent covariates of a suitable form, say, some piecewise defined covariates can be introduced and significance tests performed. However, this is not a mature method and experience is needed to use it properly ([20]).

Among the suggested forms of time-dependent covariates, the following one seems to be useful:

(a dummy variable) times (ln(survival time))

denoted by z_i = wln(t). The idea is to use the power function family, which is quite flexible, to describe $h(t;z)$, because $h(t;z)$ is proportional to $\exp(\beta_i z_i) = t^{(\beta_i w)}$.

In this project, the above method is applied to the categorical covariates only. ([20],[35]).

11 Covariate Selection Procedure

Among the procedures available, the backward selection procedure is to be used for most of the analyses to be done in this project. The advantage of backward selection over forward selection is that the former does better than the latter when some of the covariates act together---a likely situation in the data (see section 3.2).

The backward selection procedure goes like this:

1. Include all the covariates into the Cox model and decide the significance level α to be used. $\alpha = 0.05$ is used in this project.

2. Test each of the (13) covariates in turn by the maximum partial likelihood ratio test and record the P-values.
3. Leave out the covariate with the largest P-value, which must be greater than α, and repeat step 2 with the (12) covariates left.
4. When there is no covariate to leave out, the procedure terminates. The covariates not left out from the model are considered the significant prognostic factors.

([31],[44]).

12 The PHH Method of Variable Selection

As an alternative to the (partial) likelihood ratio statistic, Peduzzi, Hardy and Holford (1980) proposed the PHH statistic to calculate significance probabilities to enter or remove a candidate variable in a variable selection process.

For the candidate variable z_j, the vector of first derivatives of the (partial) likelihood function $U(\hat{\beta})$ and the observed information matrix $I(\hat{\beta})$ are calculated using current parameter estimates for those variables already in the model and zero for the candidate variable. Let $U_j(\hat{\beta})$ be the j-th component of $U(\hat{\beta})$ and $I_{jj}(\hat{\beta})$ be the j-th diagonal element of $I(\hat{\beta})$. The tail probability of

$$U_j^2(\hat{\beta})\ I_{jj}^{-1}(\hat{\beta})$$

is calculated from the chi-square distribution with one degree of freedom.

The distribution of the PHH statistic is unknown. But its close resemblance to the score test statistic and many numerical calculations suggest that the chi-square aproximation is reasonable. Also, since it is much cheaper to use than the (partial) likelihood ratio statistic, it can be used to eliminate variables with little or no relationship to survival. ([26],[51]).

There are some practical problems when using the statistical techniques described above. Discussions will be offered when these problems arise.

3.4 Fitting Cox's Regression Model and Testing Planned Hypotheses

3.4.1 Planned Hypothesis 1

In this section, multivariate analyses will be performed to study the first planned hypothesis, that is, to identify the significant prognostic factors covered by the data.

Pigmentation and regression will not be included in the multivariate analyses because they were "screened out" on the univariate analysis, and Cox's regression models will be fitted to the rest of the covariates except index, which is to be discussed in section 3.5.2.

As was discussed in section 3.2, the strong association between tumour depth and Clark's levels of invasion suggests that it be appropriate to include only one of them when fitting

the Cox model. Therefore, two parallel analyses, one including depth and another including Clark's levels, are carried out below.[1]

First, the graphical method is used to check the proportionality assumption of the Cox model. The plots are shown in Figure 3.4.1 and Figure 3.4.2. The upper plots are obtained using the original coding explained in Table 2.2 on pages 8-9; the lower plots are obtained using the same covariates but revised coding. The revised coding is in Table 3.4.1.

Table 3.4.1 Recoding information

Covariate	Levels	Codes	Names	Base-level
sex	male	1	sex1	sex1
	female	2	sex2	
site	non-limbs	2	site1	site1
	limbs	4	site4	
depth	continuous	/	depth	/
Clark's	level II and III	2	clark1	clark1
	level IV and V	4	clark4	
mito	level I and II	1	mito1	mito1
	level III	3	mito3	
cell	lentigo	1	cell1	
	superficial spreading	2	cell2	cell2
	nodular	3	cell3	

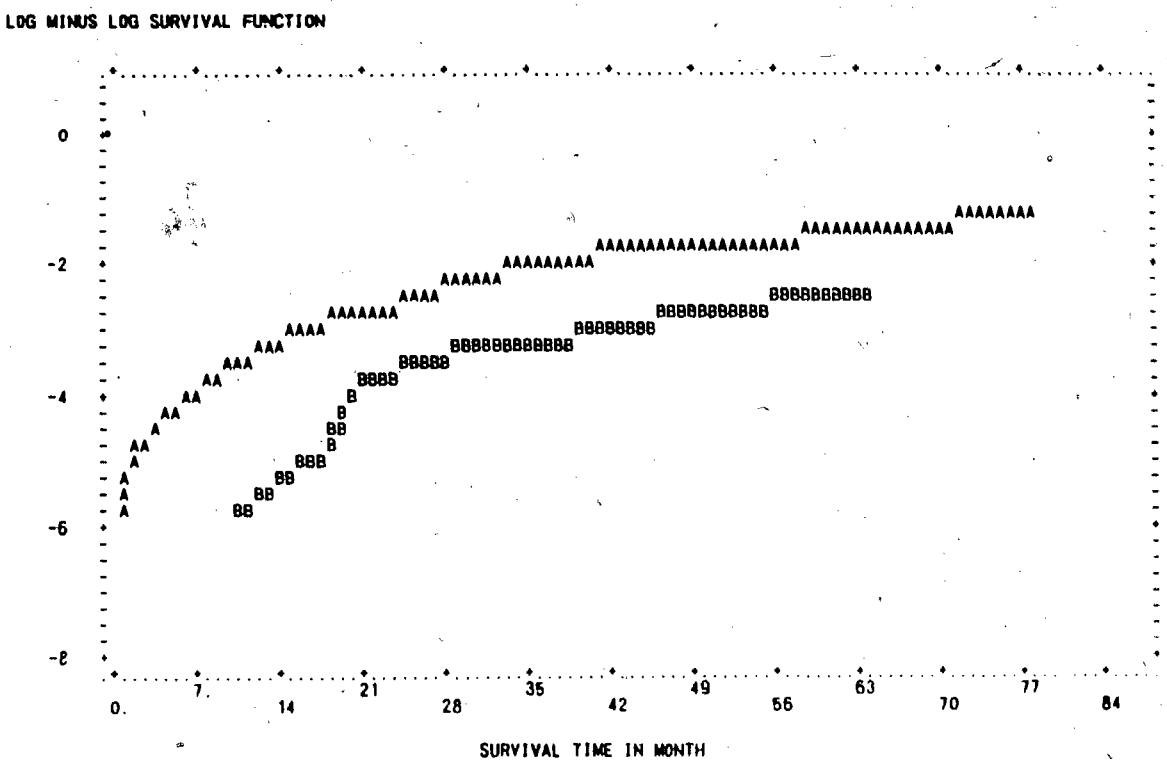
[1] To achieve computational stability, individual tumour depth minus the mean depth is used in all the model fitting calculations. Age and index are treated similarly.

diff	level I	1	diff1	
	level II and III	3	diff3	diff3
ulce	absent	0	ulce0	ulce0
	present	1	ulce1	
lymp	absent	0	lymp0	lymp0
	present	1	lymp1	
age	continuous	/	age	/

Improvements are seen by comparing the upper plots with the lower plots. For celltype, the distance is not constant between the level representing lentigo melanoma and the other two levels representing superficial spreading and nodular melanomas; for differentiation, the two curves cross. Results like these are expected, because strictly speaking, everything is changing as time changes. However, since most covariates do seem to have a proportional effect on survival, the above results are encouraging.

Secondly, the proportionality assumption is checked by testing time-dependent covariates as explained in the previous section. Some practical problems are: the method is very costly; there is no clear guide as to what kind of time-dependent covariates should be introduced and to which covariate(s) in the data one should introduce time-dependent covariate(s). More importantly, a dilemma exists: ideally one wants to check the assumption before fitting a model to the data by testing time-dependent covariates, but in order for the tests to give

Figure 3.4.1 Checking proportionality when depth is used (sex: upper plot for original codes, lower plot for new codes).



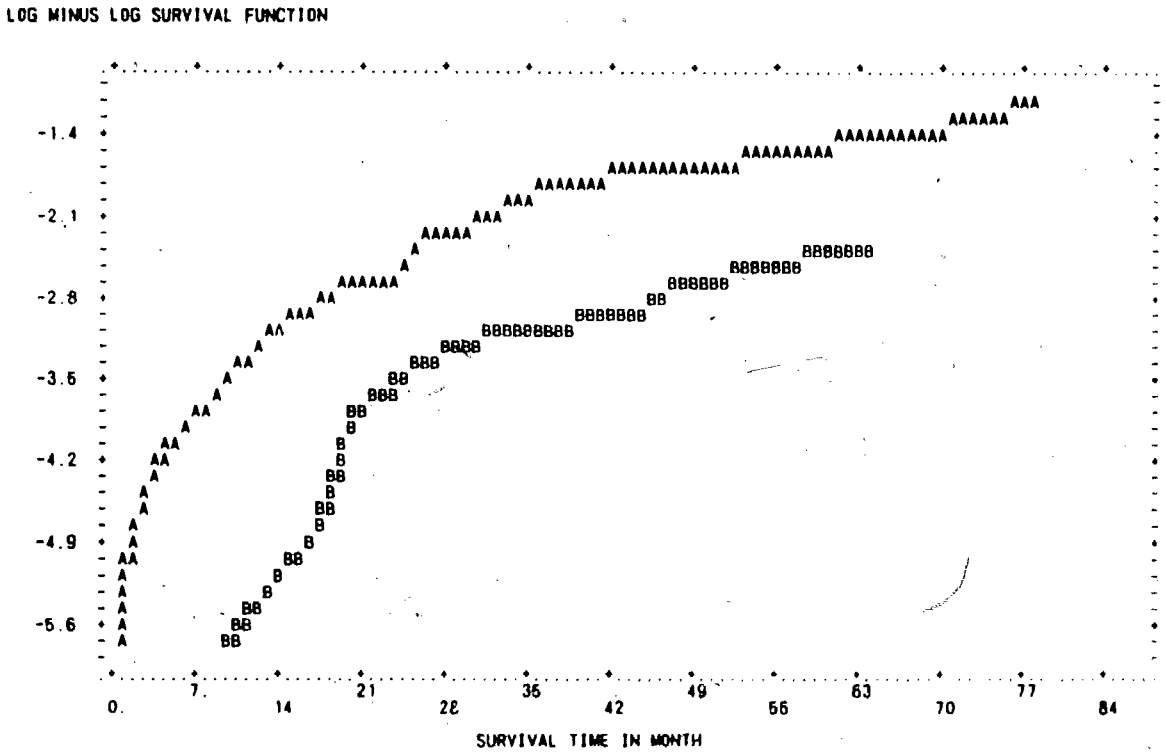
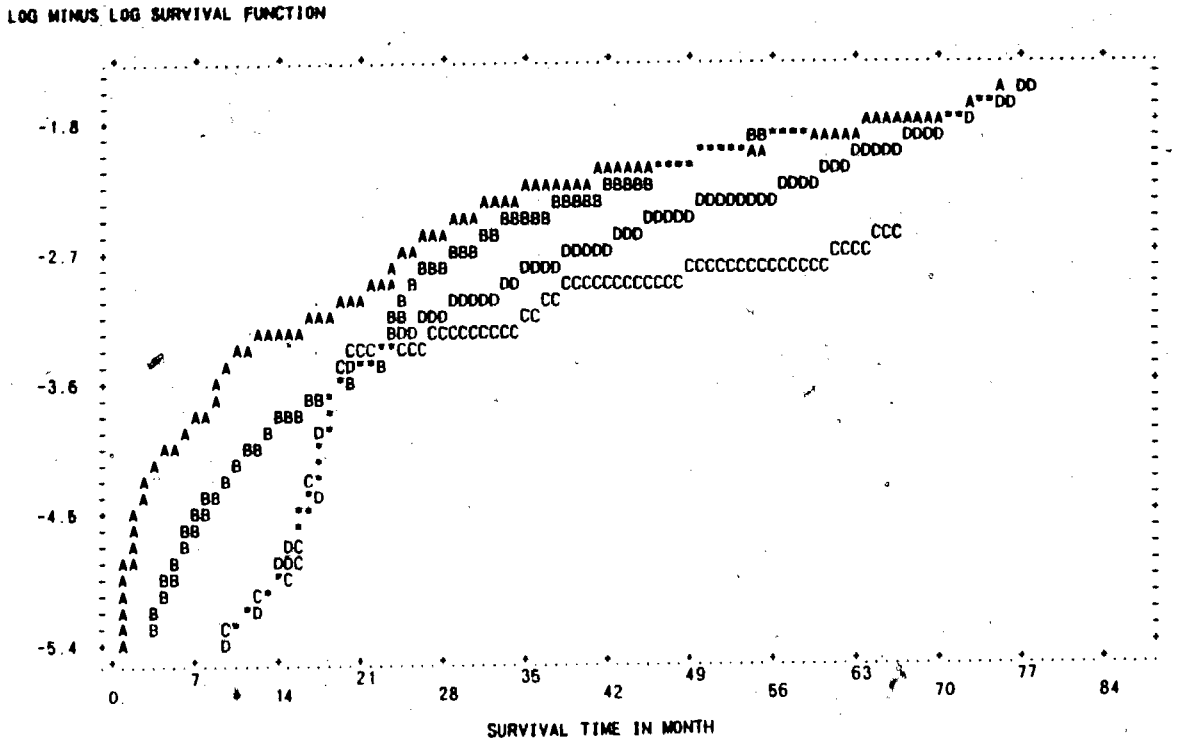


Figure 3.4.1 continued (site: upper plot for original codes, lower plot for new codes).



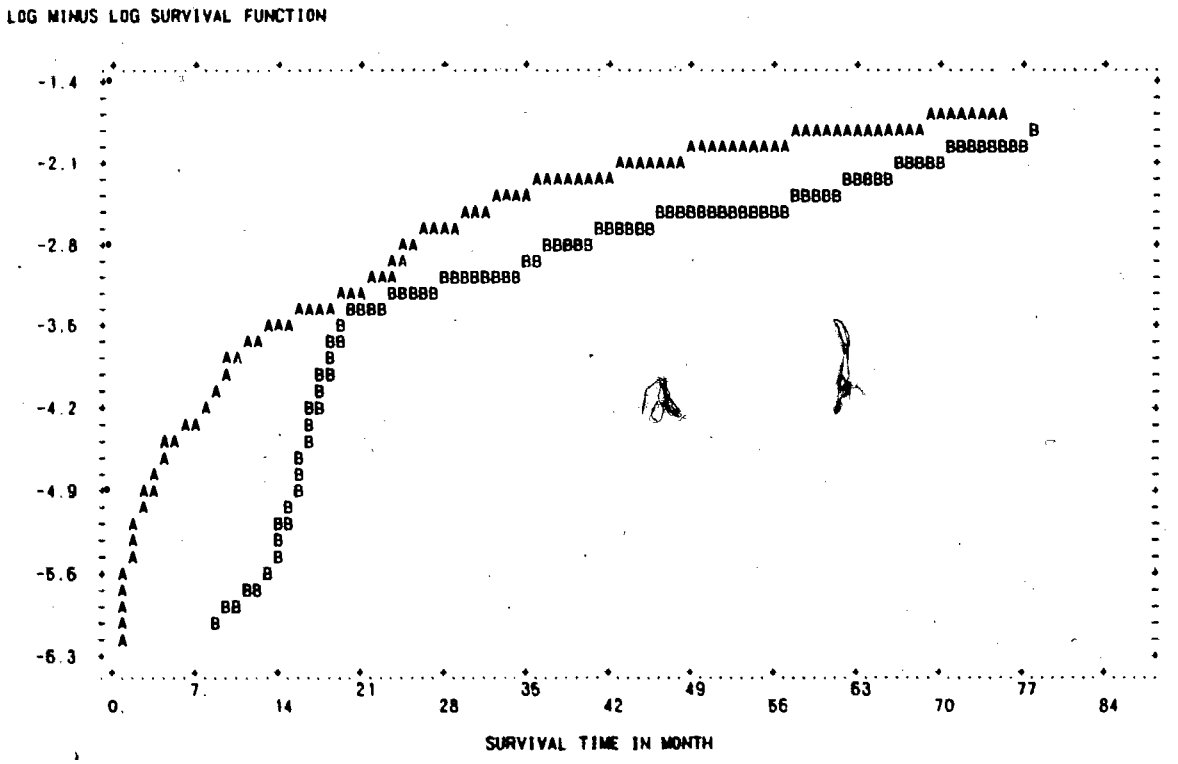
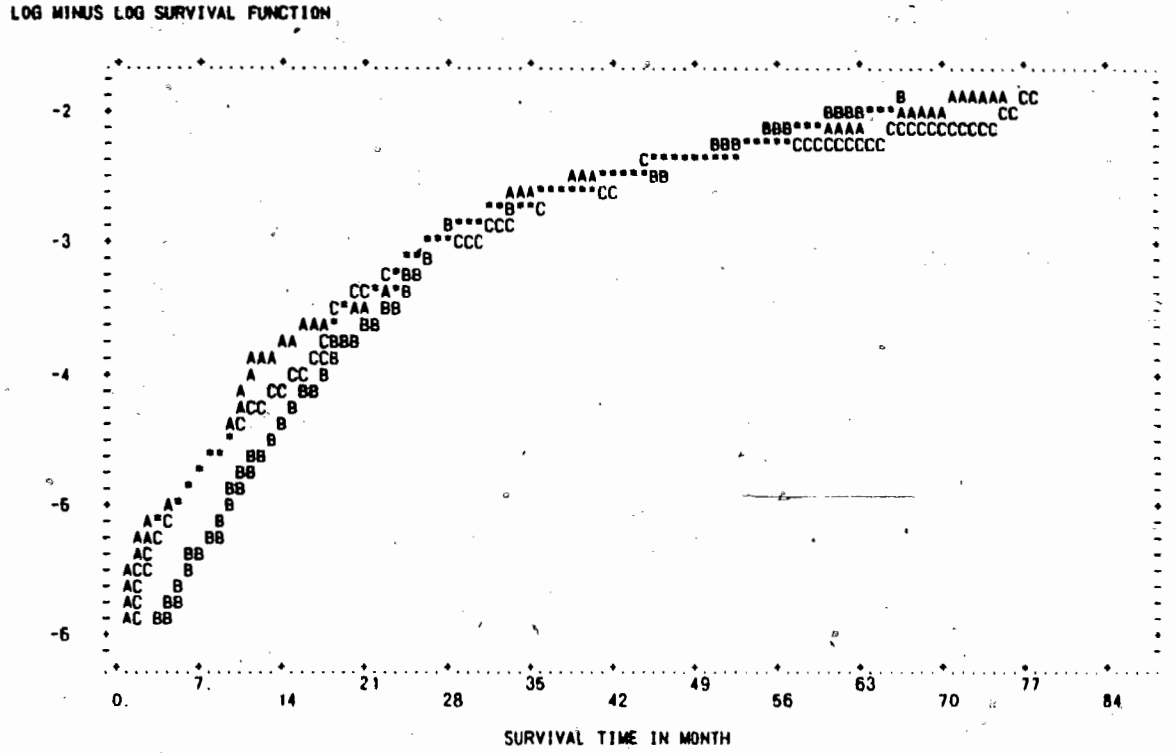


Figure 3.4.1 continued (mitoses: upper plot for original codes, lower plot for new codes).



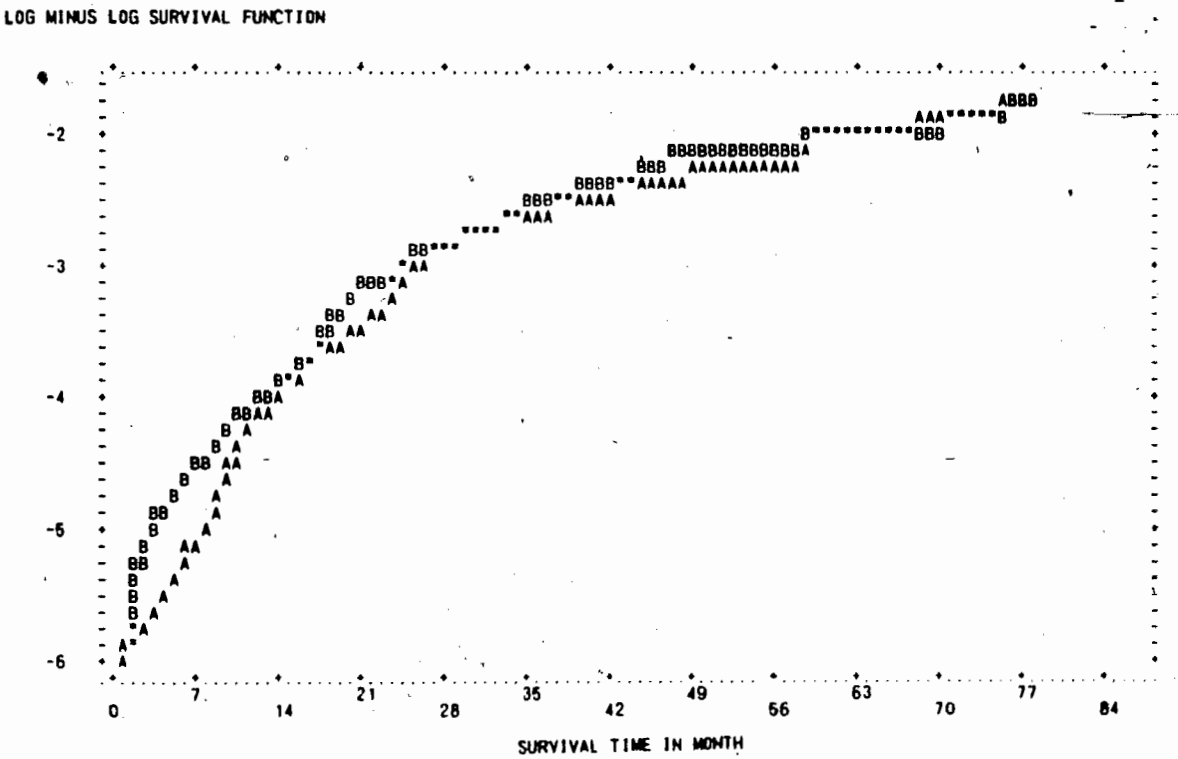
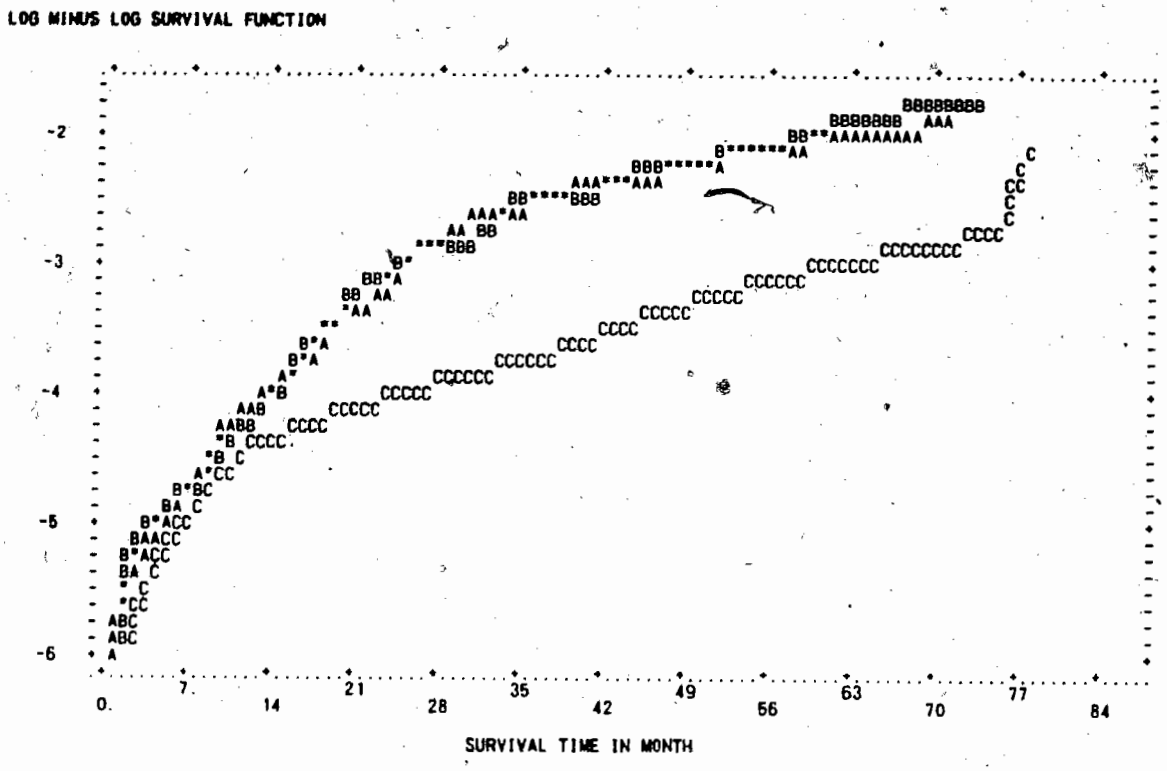


Figure 3.4.1 continued (celltype: upper plot for original codes, lower plot for new codes).



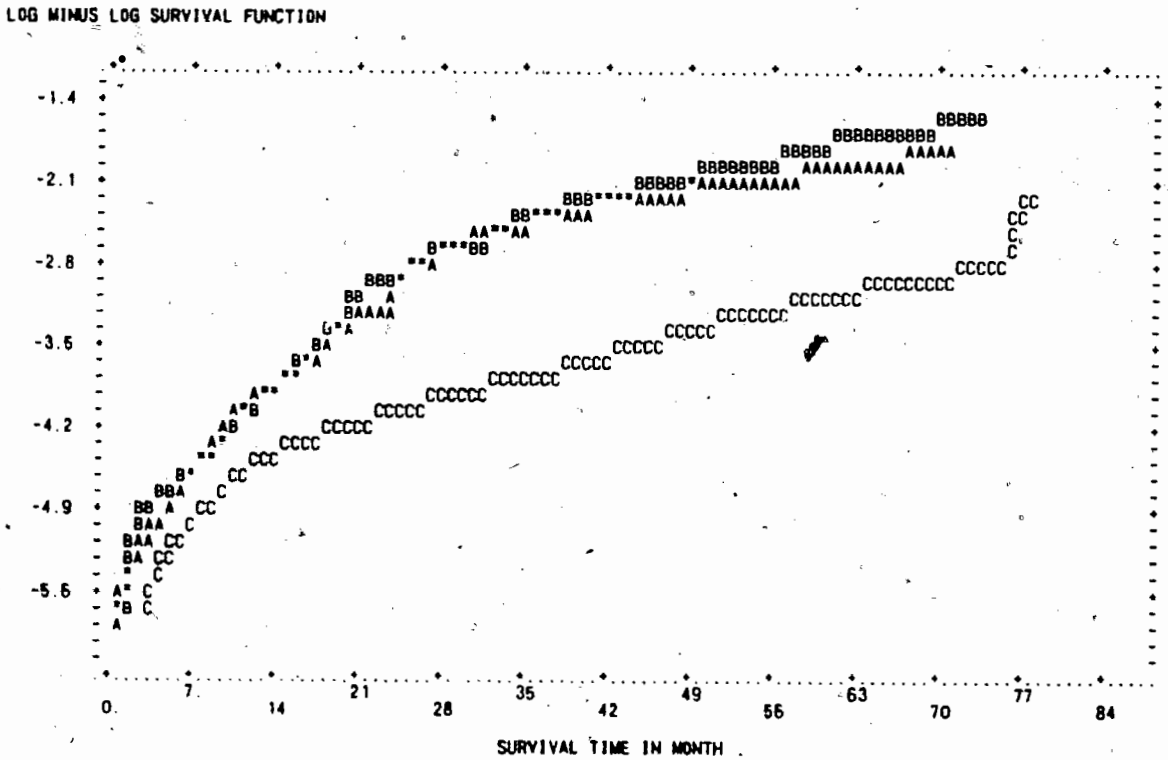
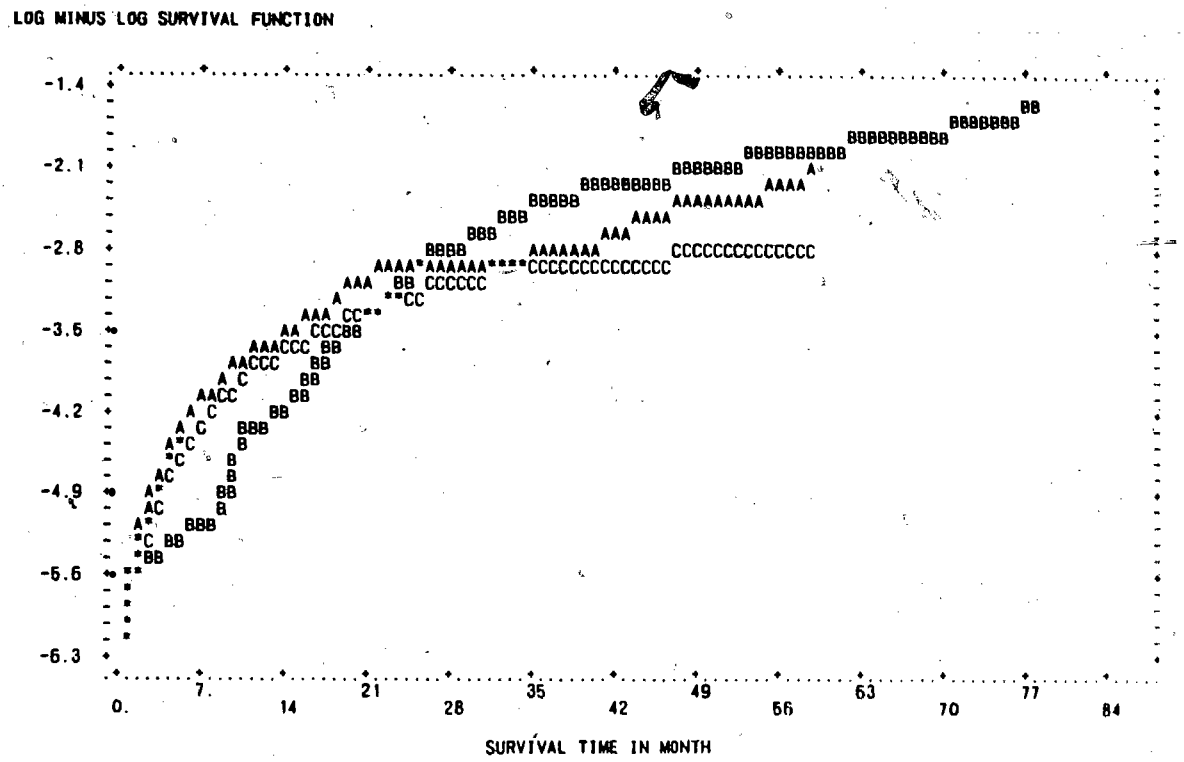


Figure 3.4.1 continued (differentiation: upper plot for original codes, lower plot for new codes).



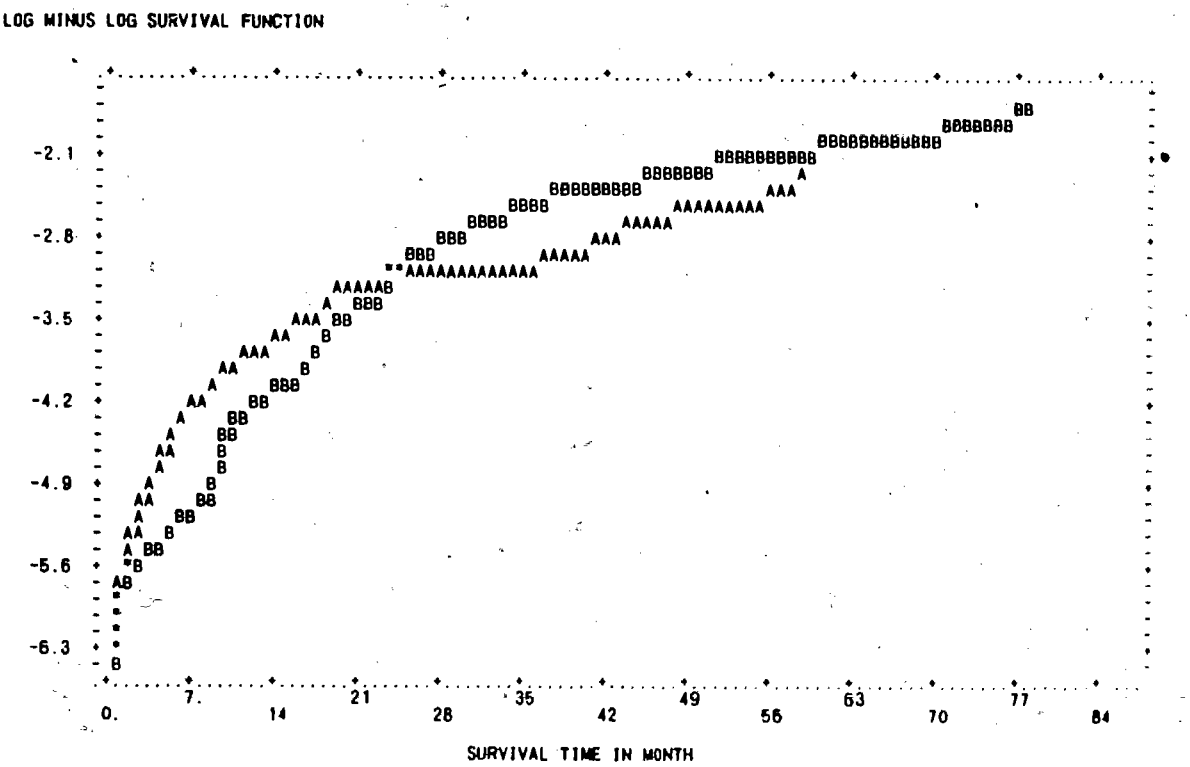
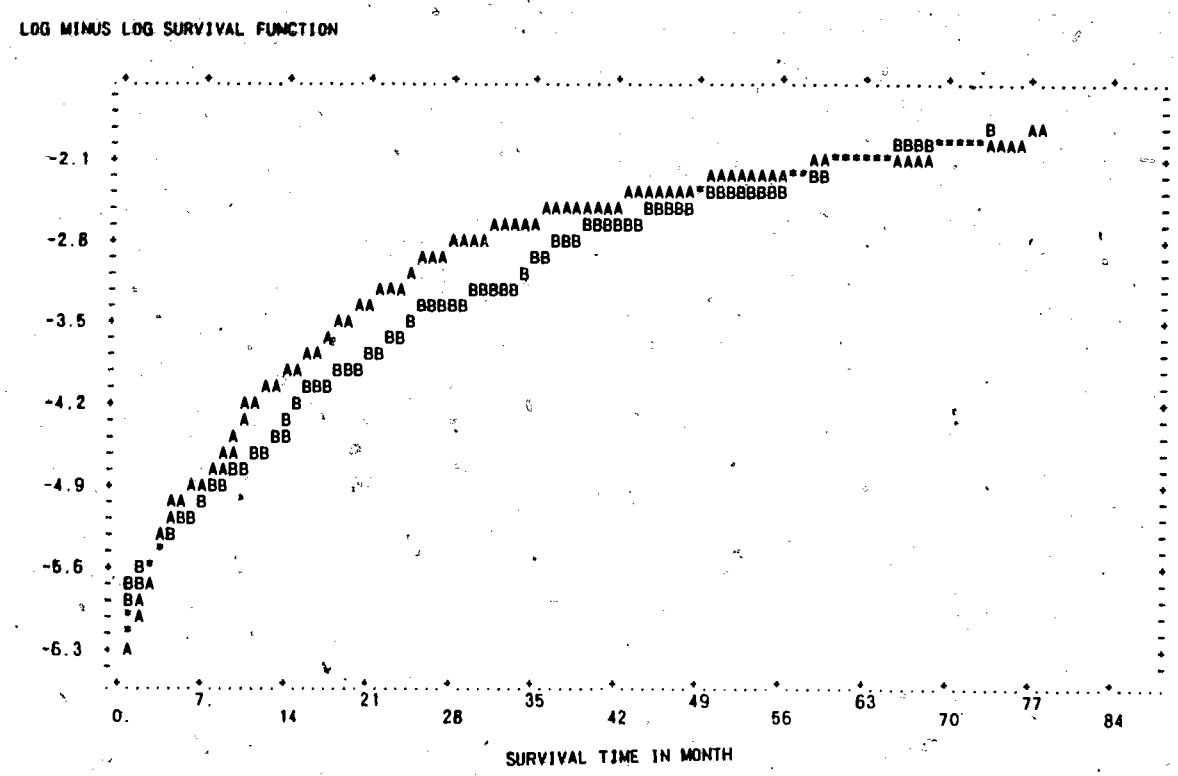


Figure 3.4.1 continued (ulceration: upper plot for original codes, lower plot for new codes).



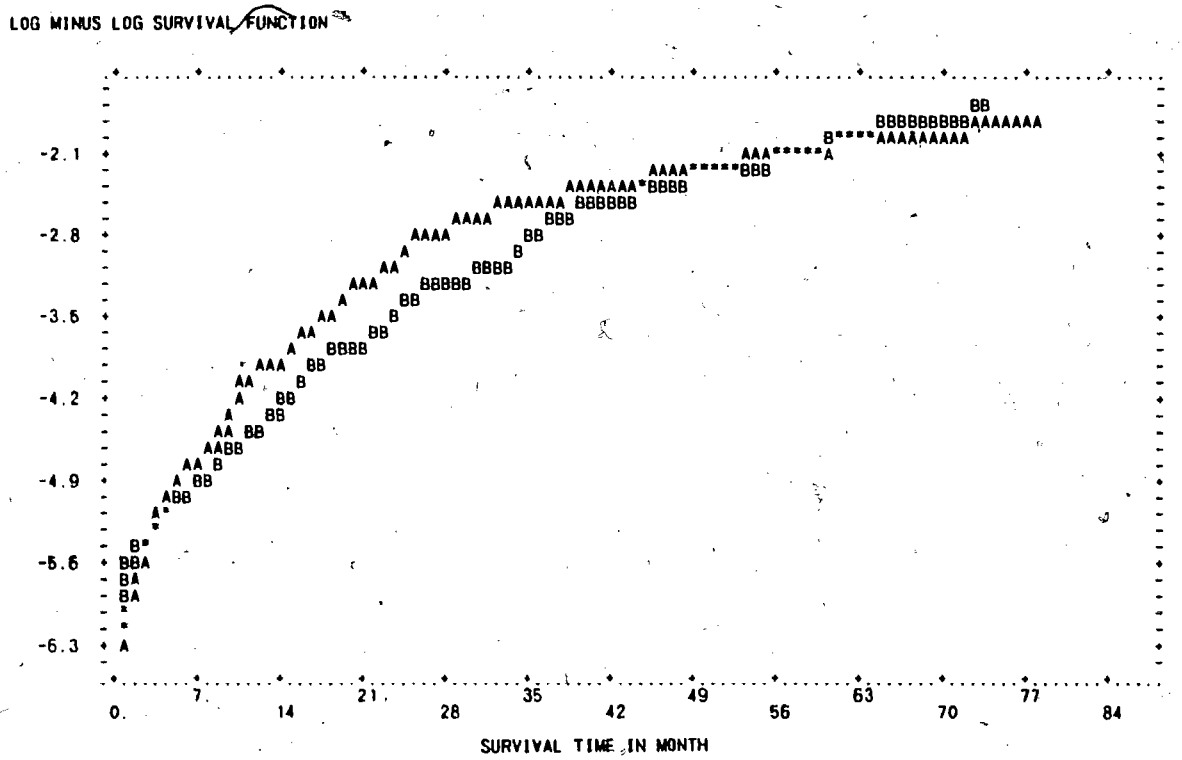
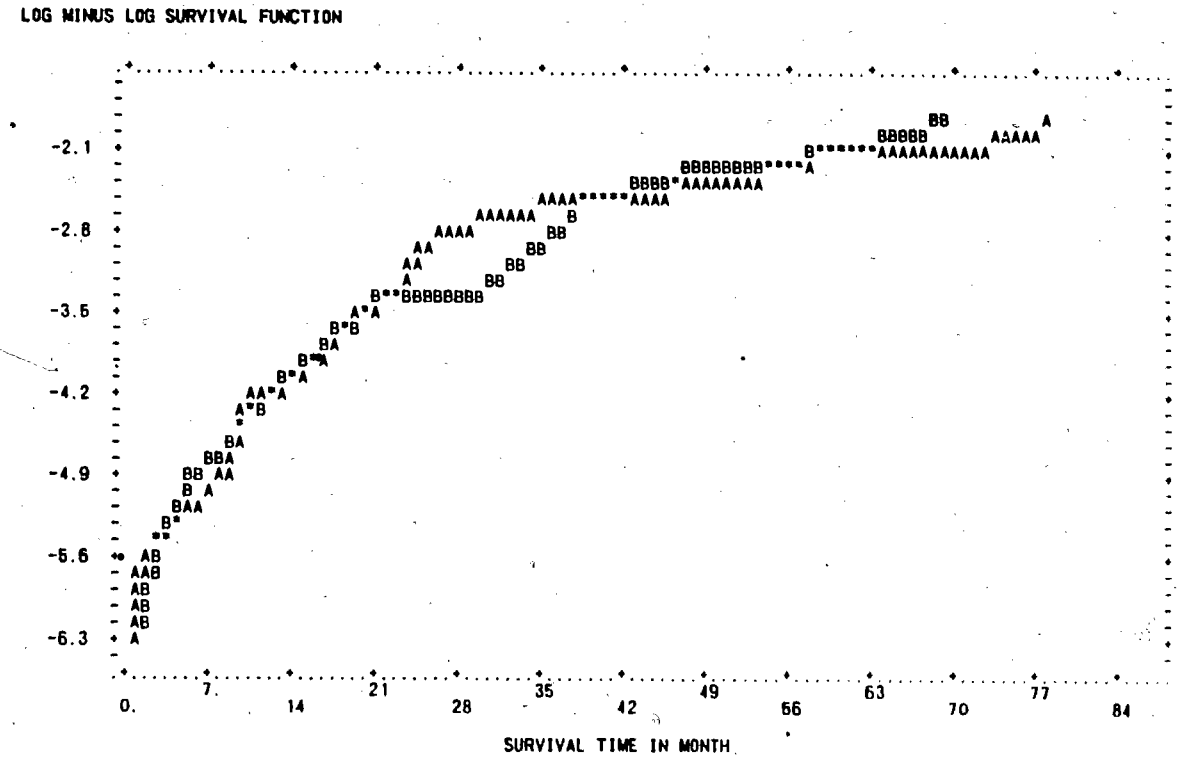


Figure 3.4.1 continued (lymphatic inflammation: upper plot for original codes, lower plot for new codes).



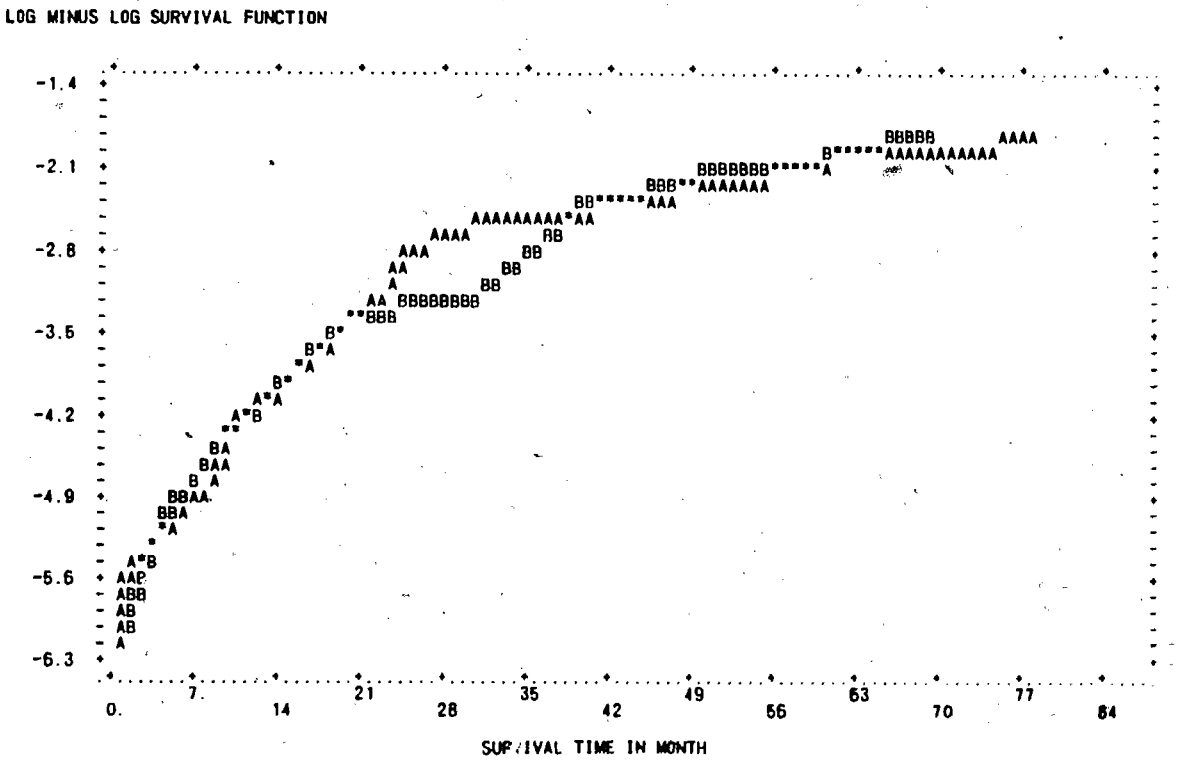
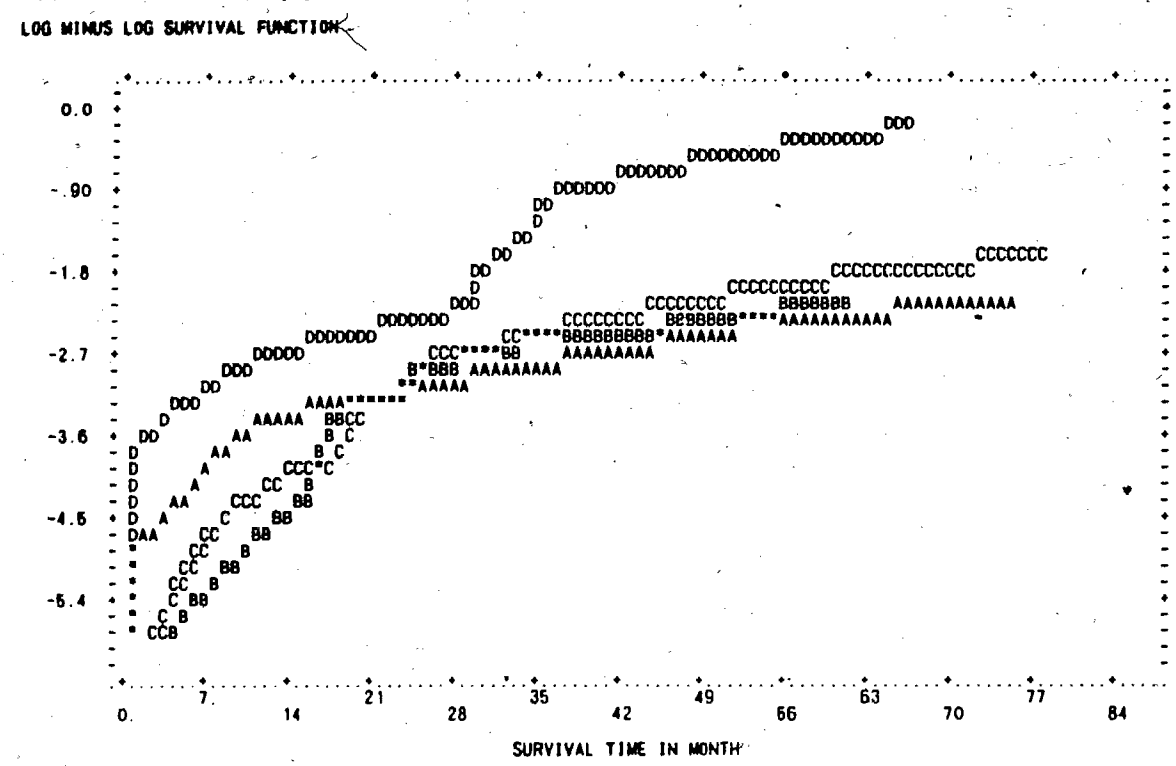


Figure 3.4.2 Checking proportionality when Clark's levels are used (Clark's levels: upper plot for original codes, lower plot for new codes).



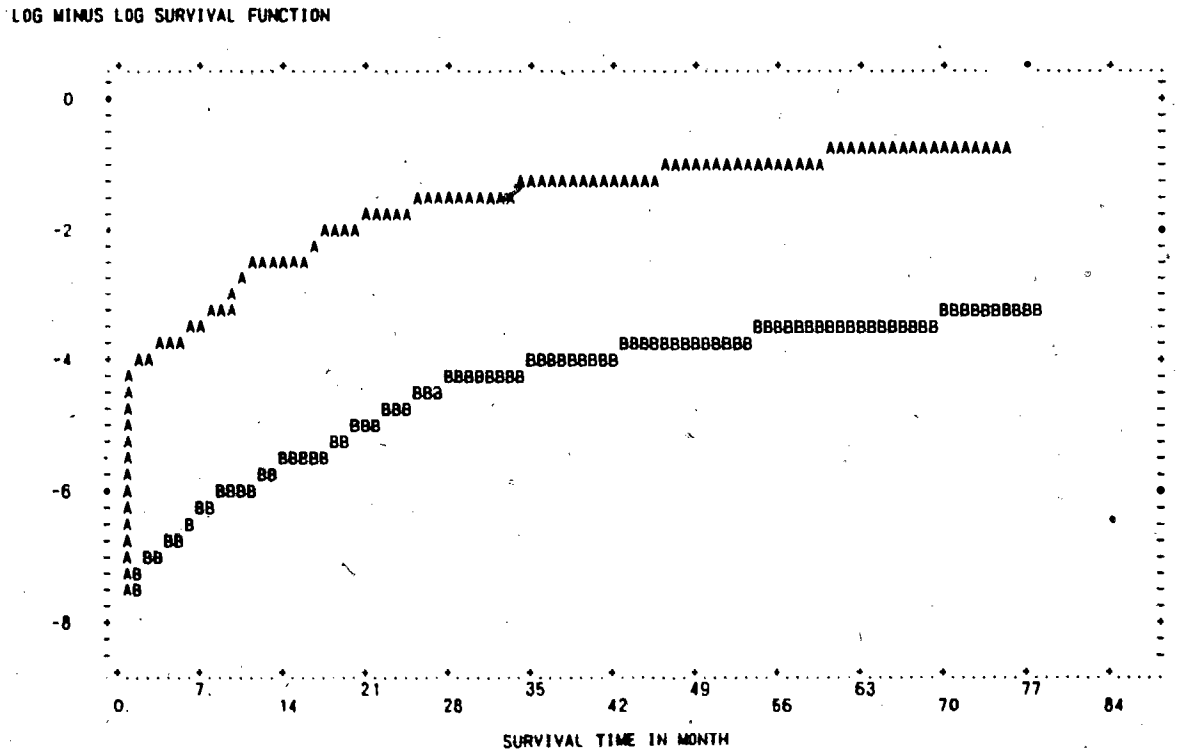
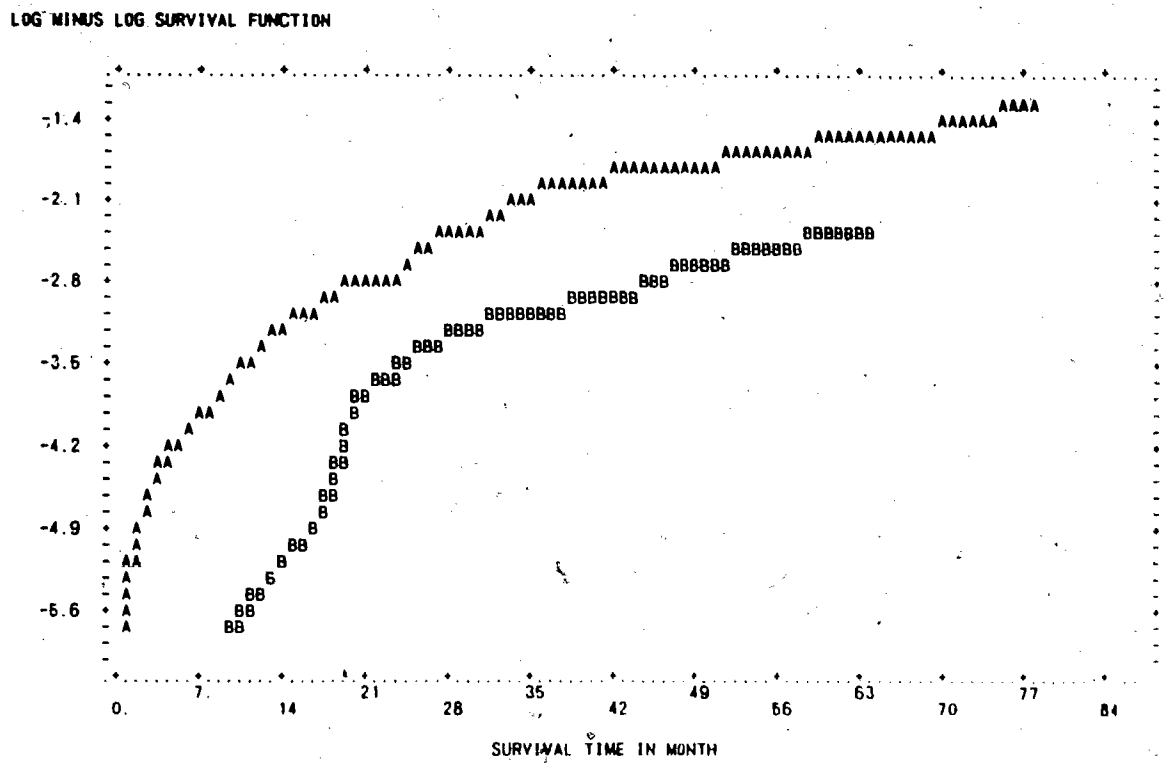


Figure 3.4.2 continued (sex: upper plot for original codes, lower plot for new codes).



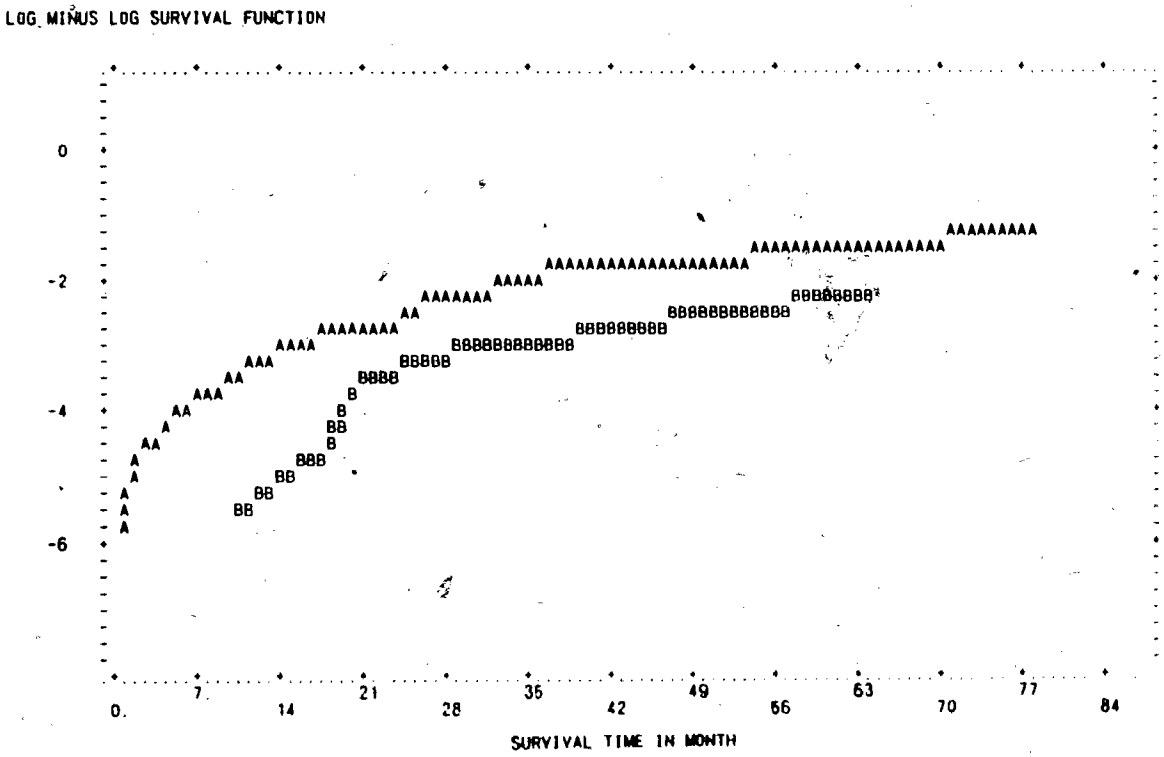
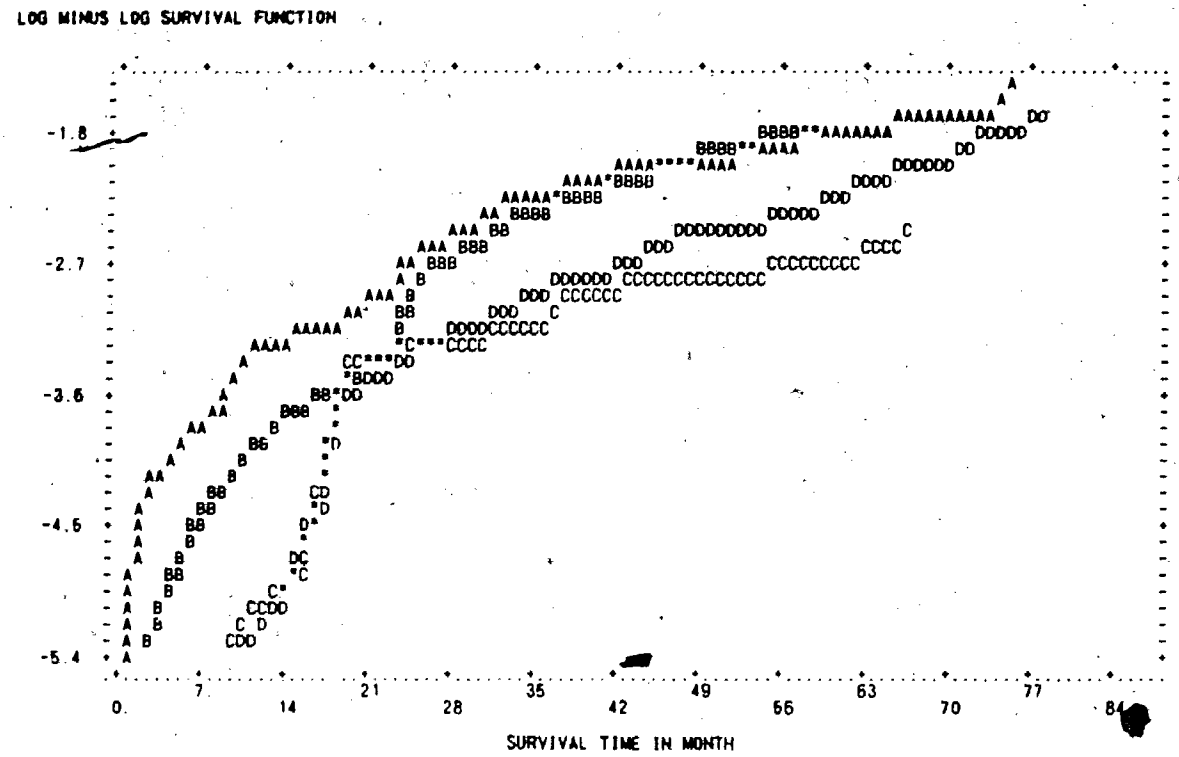


Figure 3.4.2 continued (site: upper plot for original codes, lower plot for new codes).



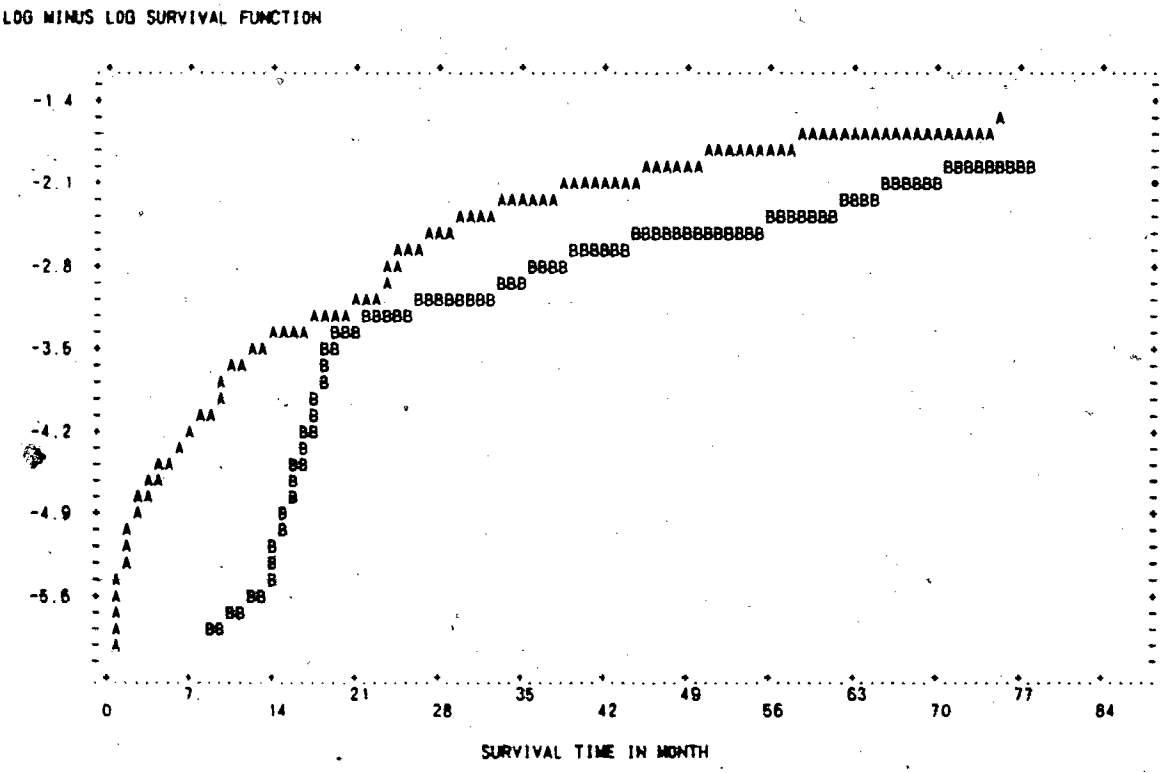
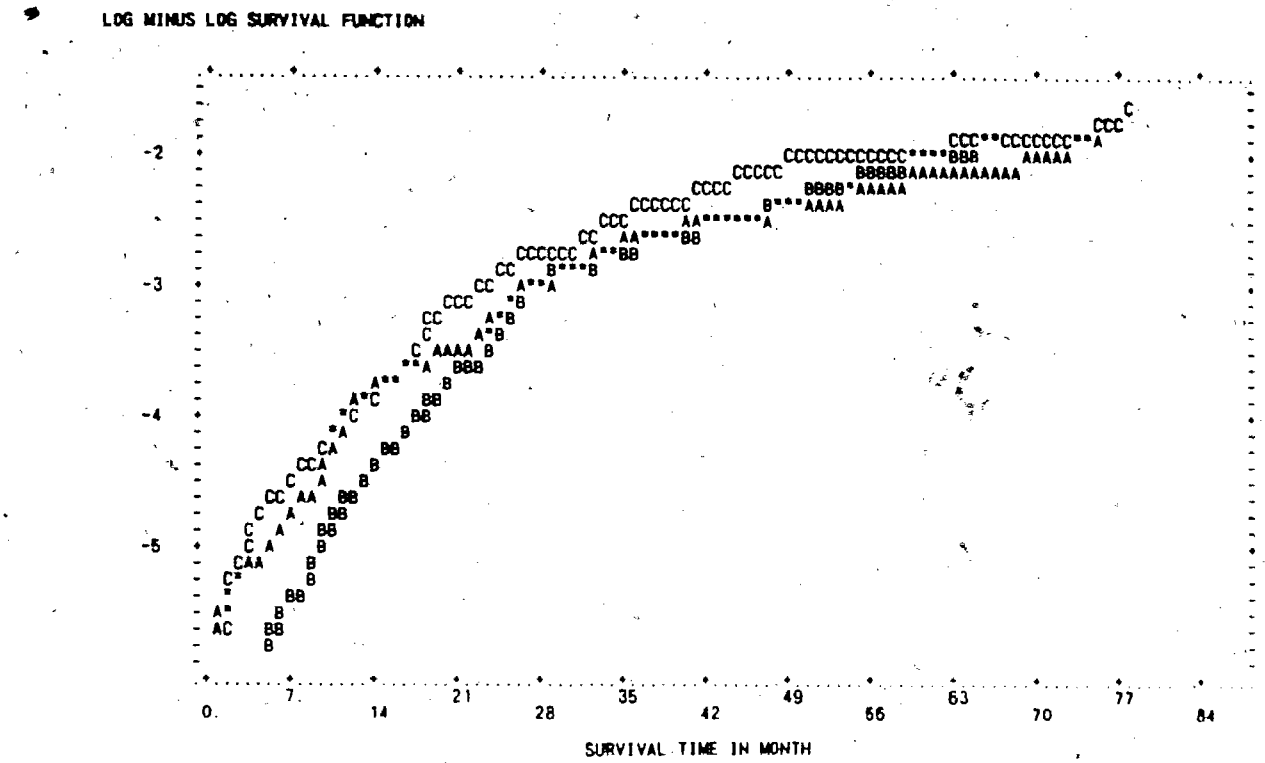


Figure 3.4.2 continued (mitoses: upper plot for original codes, lower plot for new codes).



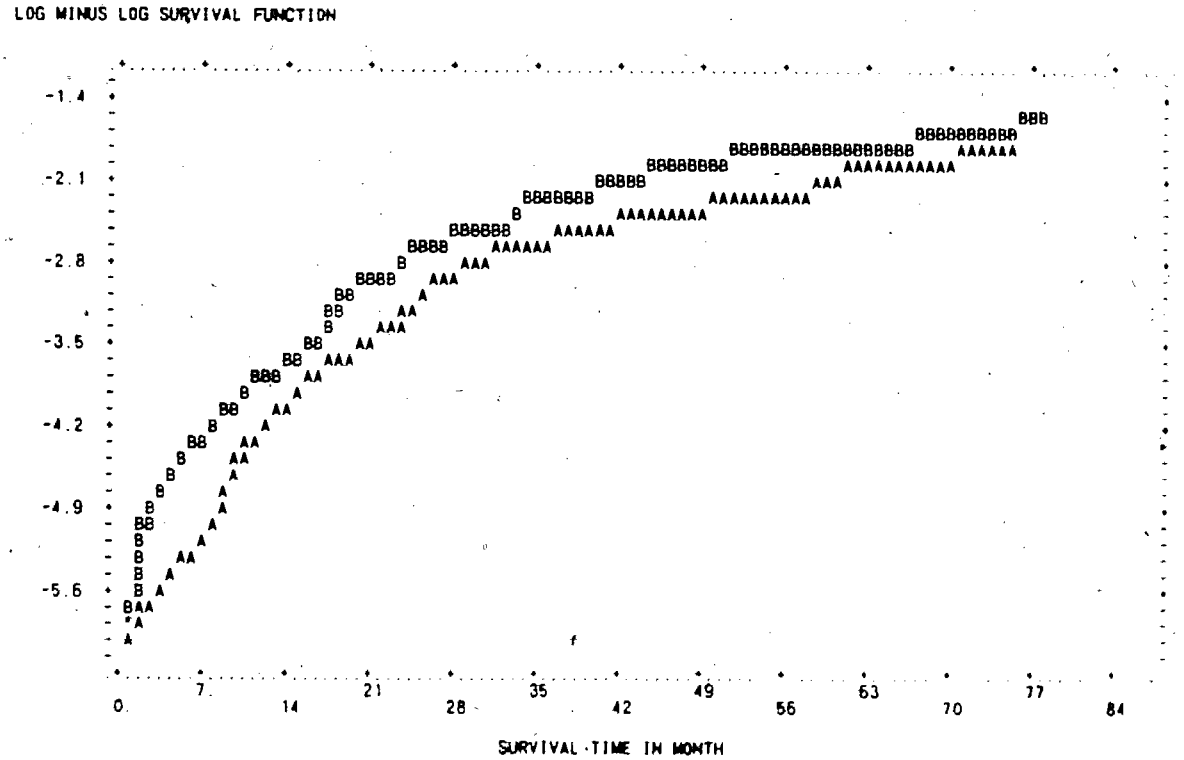
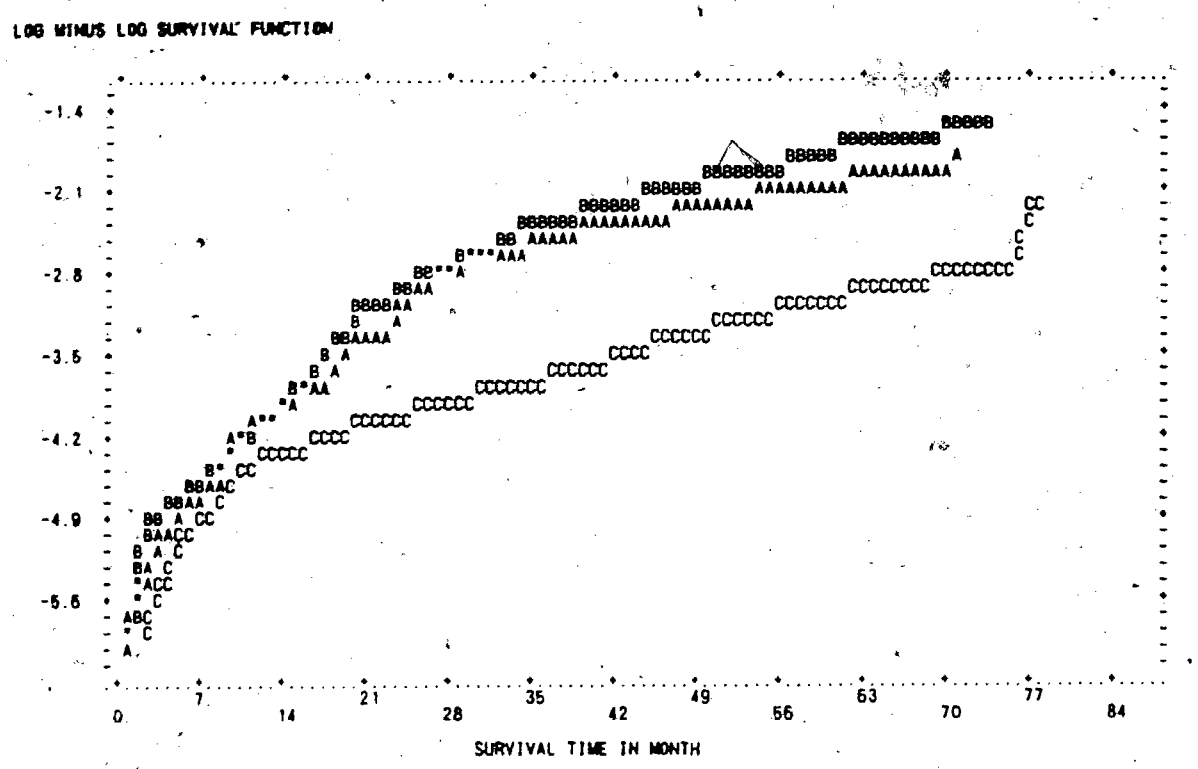


Figure 3.4.2 continued (celltype: upper plot for original codes, lower plot for new codes).



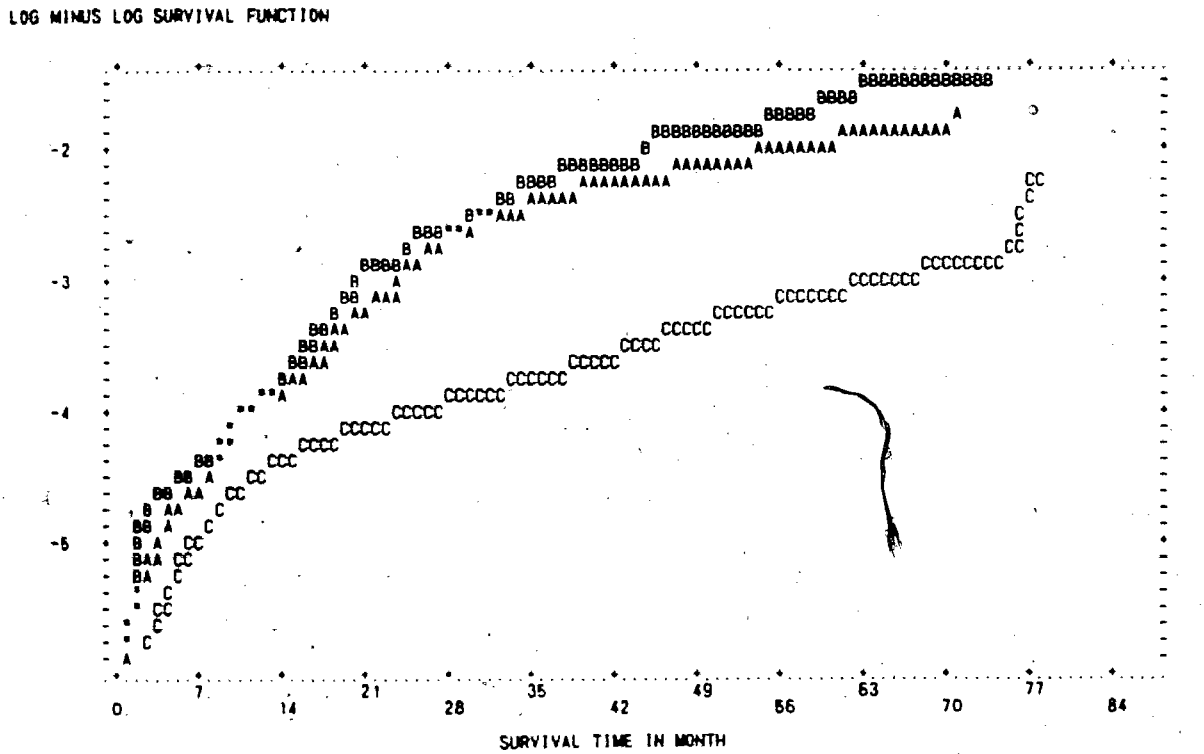
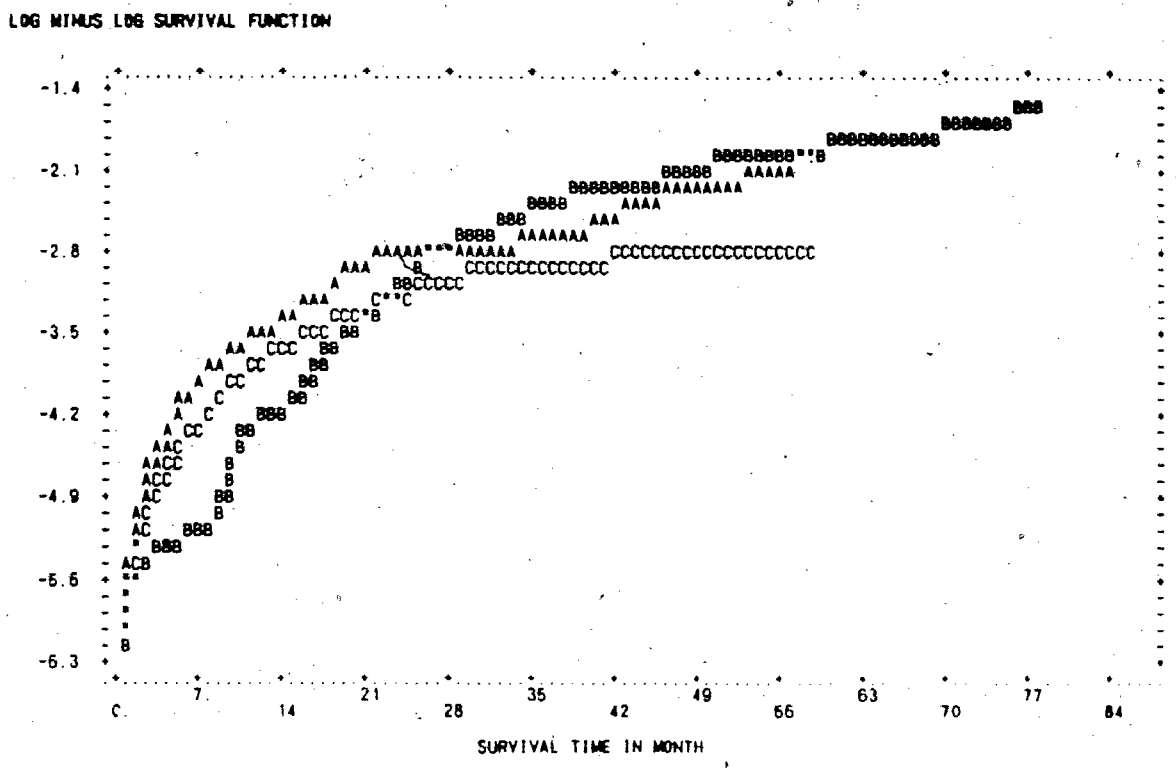


Figure 3.4.2 continued (differentiation: upper plot for original codes, lower plot for new codes).



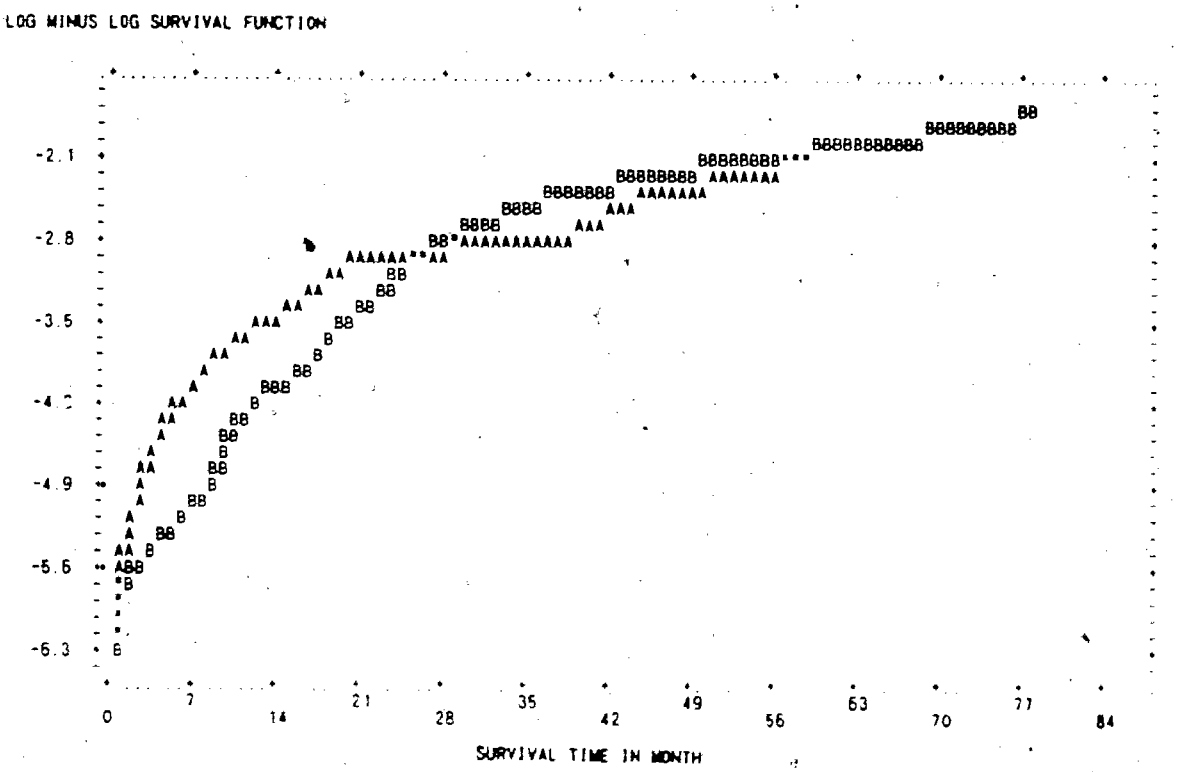
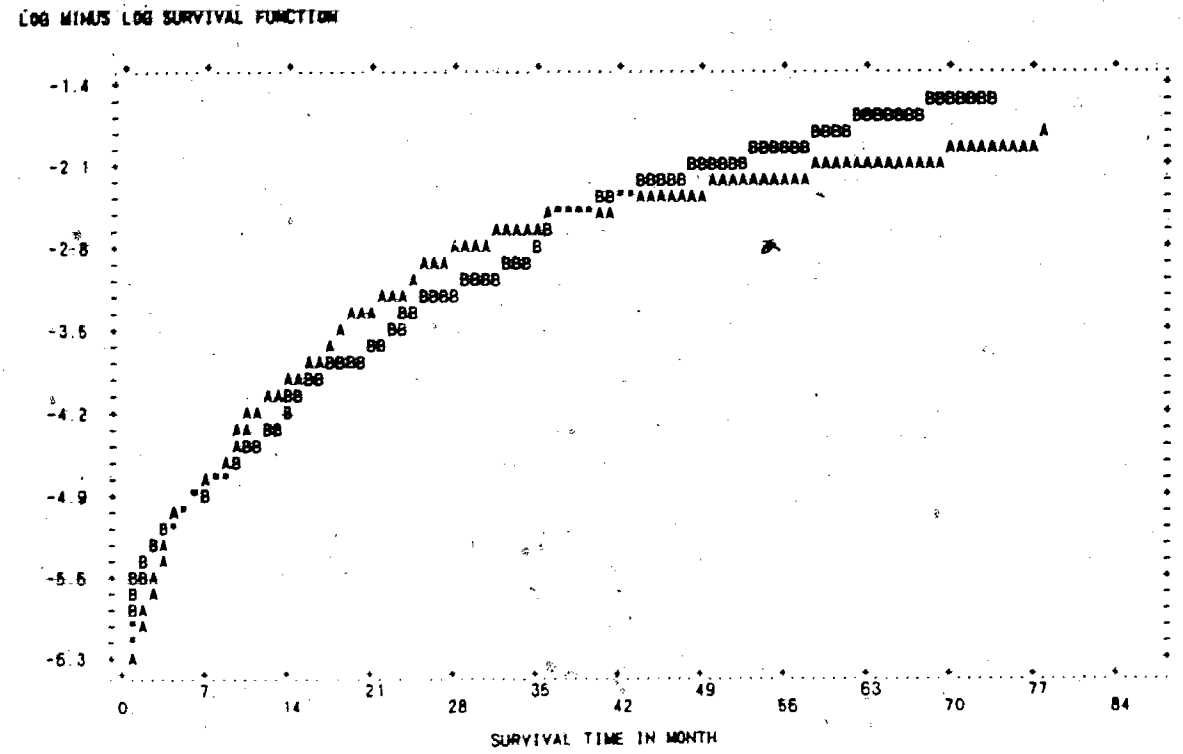


Figure 3.4.2 continued (ulceration: upper plot for original codes, lower plot for new codes).



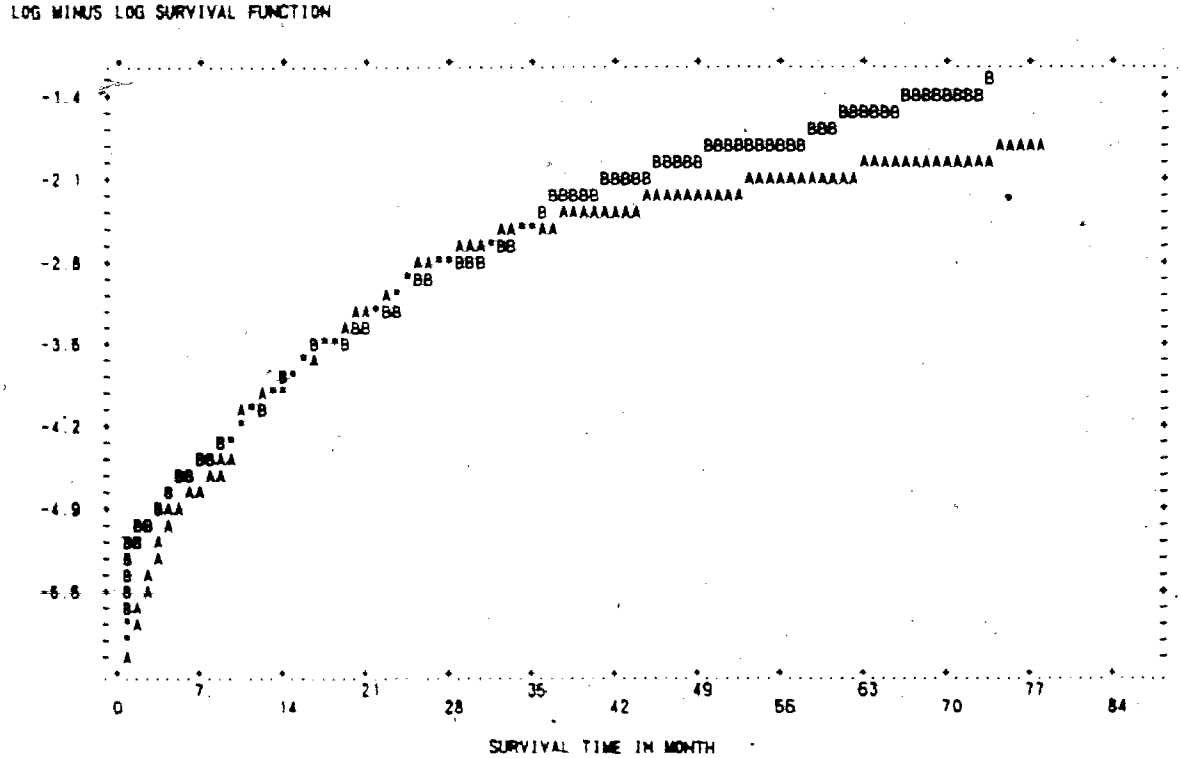
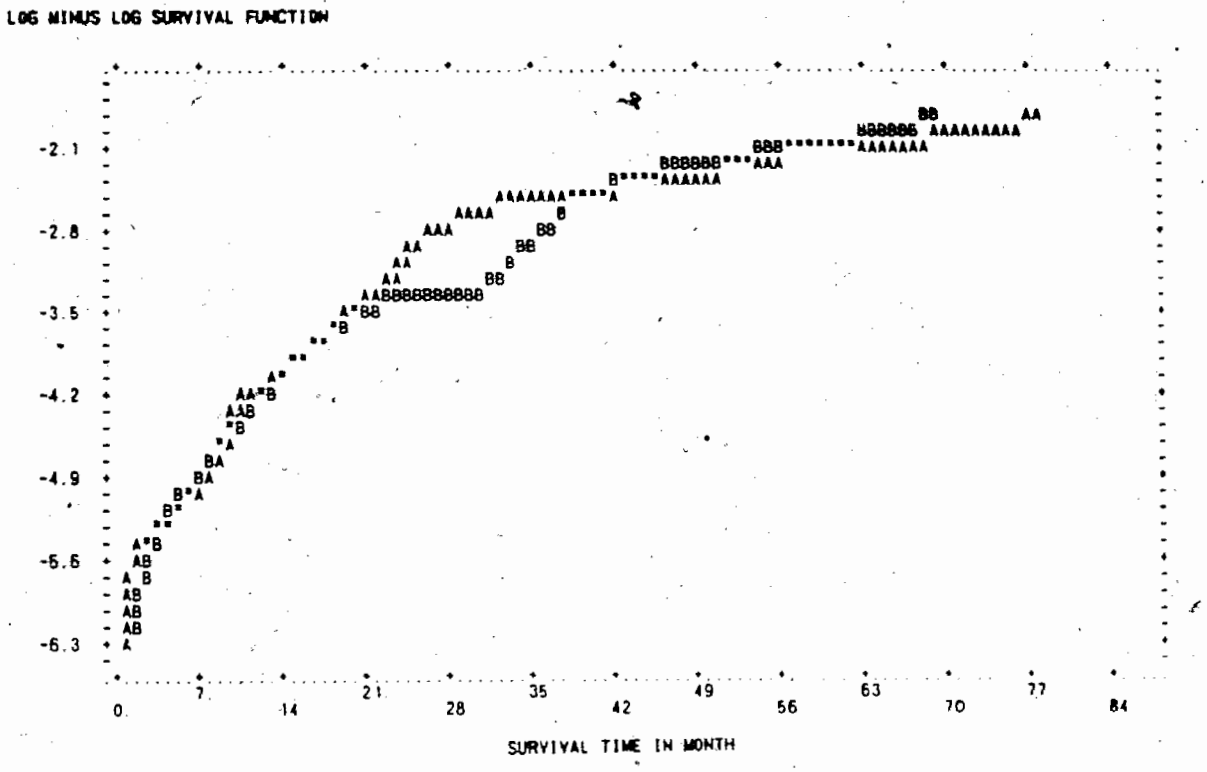
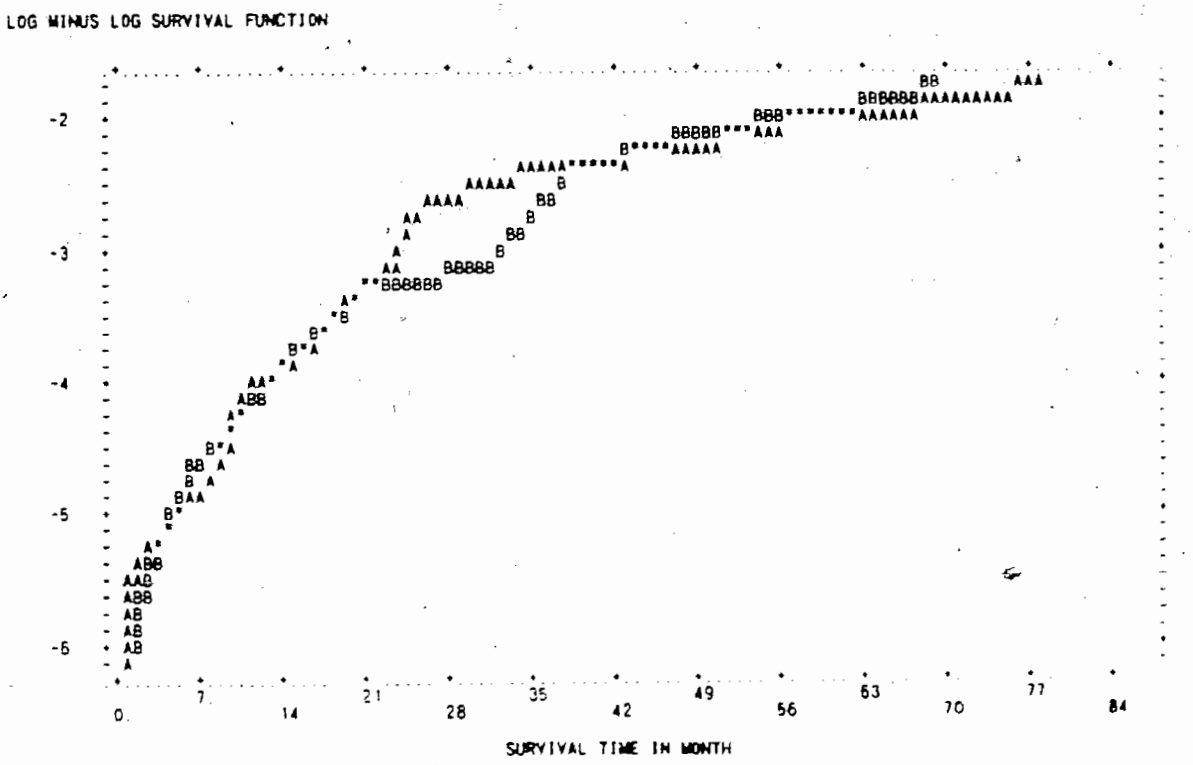


Figure 3.4.2 continued (lymphatic inflammation: upper plot for original codes, lower plot for new codes).





valid results, one needs to ensure that the model containing the time-dependent covariates is well-fitted, which gives rise to a new and at least equally difficult assumption to be checked.

In this project, the following approach is adopted. Regarding a continuous covariate as having one level, the levels of the covariates in the data are treated as individual covariates and a backward selection is carried out using the PHH statistic. Then time-dependent covariates are introduced for the significant levels selected in the above backward selection. This approach is based on the intuition that those levels which appear to influence survival more than other levels are more likely to violate the assumption and thus should be checked. The results are shown below and the new codes are used.

Table 3.4.2 Backward selection for significant levels.

	When DEPTH is used	
Covariates	Chi-square to keep in	P-values
depth	43.40	0.0
sex2	11.13	0.0008
age	23.20	0.0000
	When CLARK'S LEVELS are used	
Clark4	7.57	0.0059
sex2	10.21	0.0014
age	20.88	0.0000
cell1	4.13	0.0421
cell3	7.33	0.0068

Table 3.4.3 Testing time-dependent covariates. (z = ln(survival time) and all the time-dependent covariates are tested simultaneously)

Covariates in the model	Time-dependent covariates tested	Test statistics	df	P-values
When DEPTH is used				
depth		lratio	1	0.2342
sex2	z1 = sex2 * z	score	1	0.2522
age		Wald	1	0.2592
z1				
When CLARK'S LEVELS are used				
clark4	z1 = clark4 * z			
sex2	z2 = sex2 * z	lratio	4	0.4078
age		score	4	0.4397
cell2	z3 = cell2 * z	Wald	4	0.4576
cell3	z4 = cell3 * z			
z1,z2,z3,z4				

The above checks show that the proportionality assumption is acceptable for the data after recoding some of the covariates. Two backward selections are then performed (using the partial likelihood ratio test and the recoded covariates) to identify significant prognostic factors. The results are summarized in Table 3.4.4 and Table 3.4.5.

Table 3.4.4 Backward selection when depth is used.

Step no.	Covariates left out	Chi-square to leave out	df	P-values	Global test P-values
1	mitoses	0.0032	1	0.9581	0.0000
2	ulceration	0.0236	1	0.8780	0.0000
3	lymp	0.0240	1	0.8769	0.0000
4	diff	0.7918	1	0.3735	0.0000
5	site	1.9400	1	0.1637	0.0
6	celltype	4.9938	2	0.0823	0.0
7	no covariate to be left out.				

The kept-in covariates are:

Name	P-values
depth	0.0
age	0.0007
sex	0.0000

The estimation of the kept-in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.
depth	0.3309	0.0502	6.5881
age	0.0231	0.0069	3.3360
sex(female)	-1.1655	0.2420	-4.8162

Table 3.4.5 Backward selection when Clark's levels are used.

Step no.	Covariates left out	Chi-square to leave out	df	P-values	Global test P-values

1	lymp	0.0046	1	0.9452	0.0000
2	diff	0.1546	1	0.6942	0.0000
3	mito	0.6738	1	0.4117	0.0000
4	ulceration	1.6122	1	0.2042	0.0000
5	site	2.3718	1	0.1236	0.0
6	no covariate to be left out.				

The kept-in covariates are:

Name	P-values
Clark's	0.0066
age	0.0003
sex	0.0000
celltype	0.0037

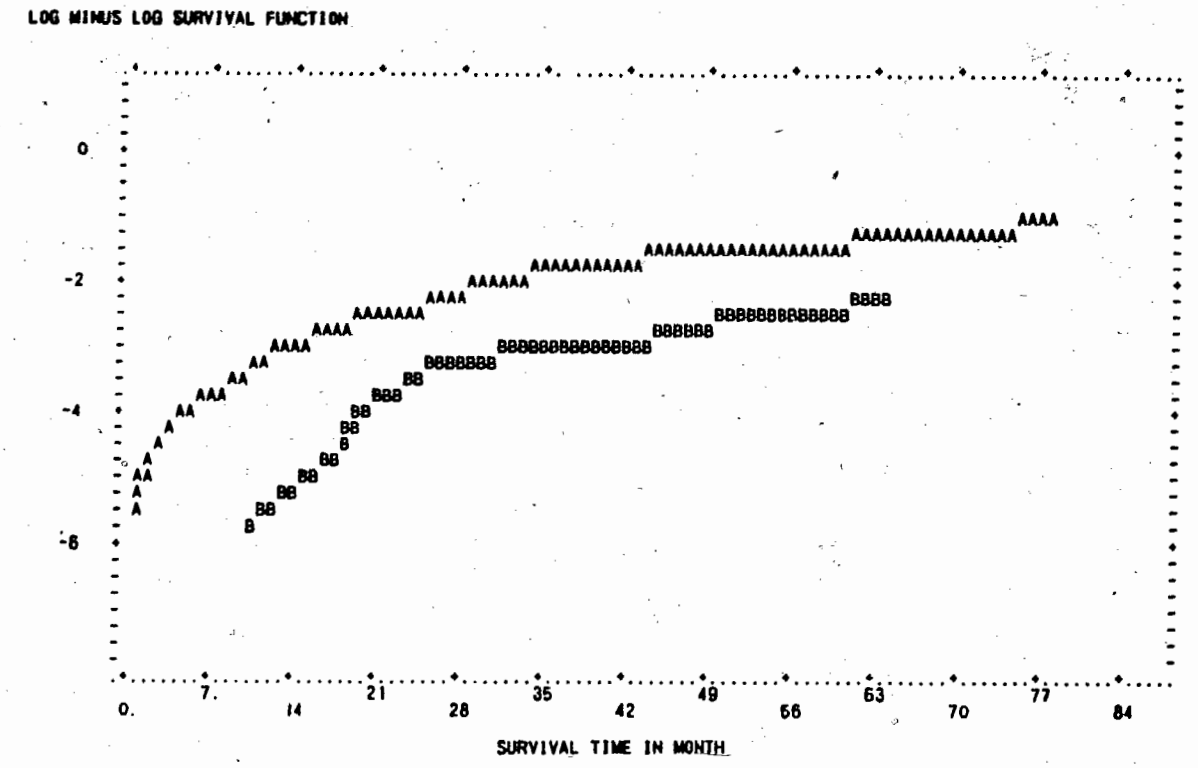
The estimation of the kept-in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.
Clark(IV,V)	0.7005	0.2607	2.6874
age	0.0271	0.0075	3.5987
sex(female)	-1.1615	0.2402	-4.7941
cell1(lentigo)	-0.9698	0.5367	-1.8067
cell3(nodular)	0.5439	0.2548	2.1347

With depth in the model, depth, age and sex are identified as significant prognostic factors; with Clark's levels in the model, Clark's levels, age, sex and celltype are identified as significant prognostic factors.

It is understood that model fitting is an evolutionary process. Thus, the graphical method is used once again to check

Figure 3.4.3 Checking proportionality when only the significant factors are used (upper plot for sex under depth, lower plot for Clark's levels under Clark's levels).



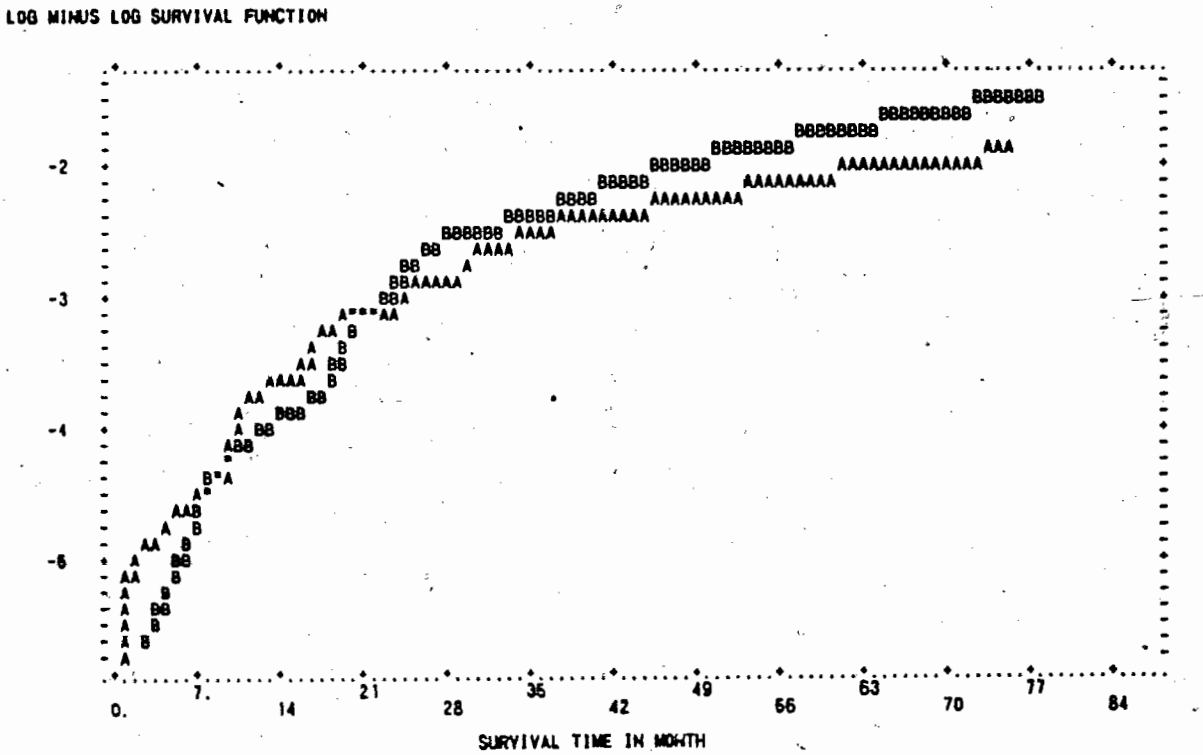
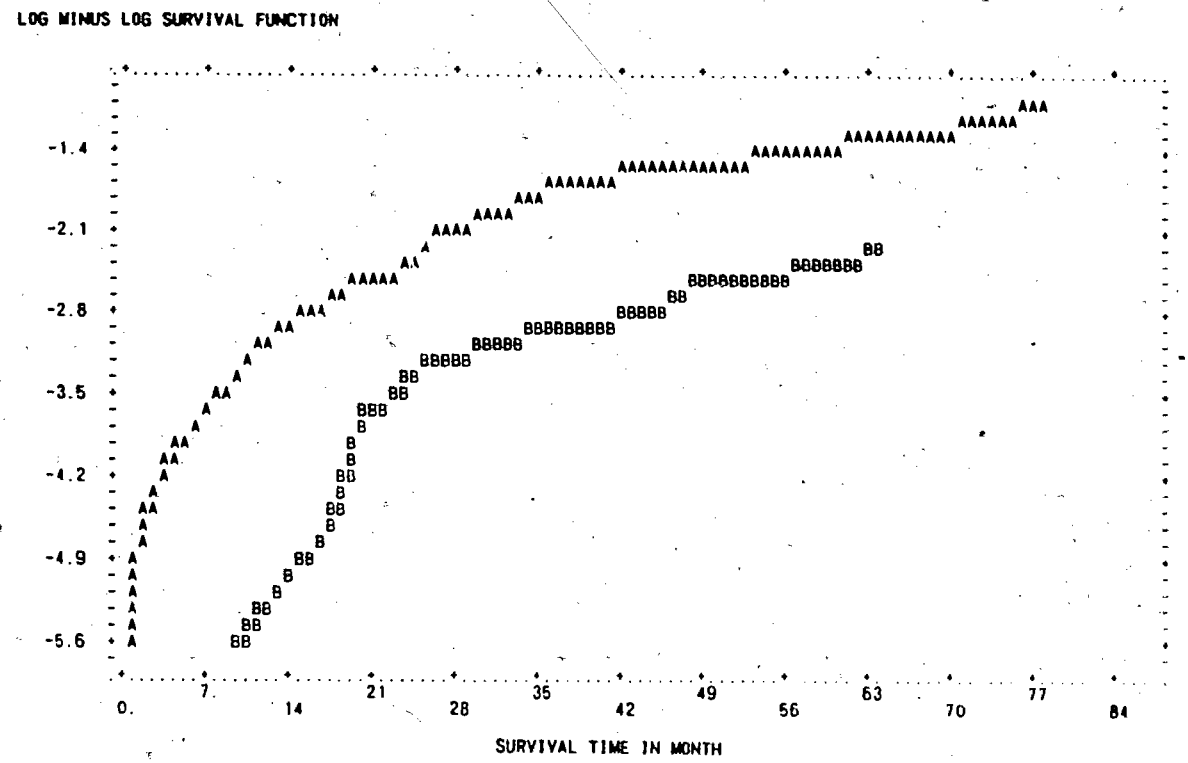


Figure 3.4.3 continued (upper plot for sex under Clark's, lower plot for celltype under Clark's).



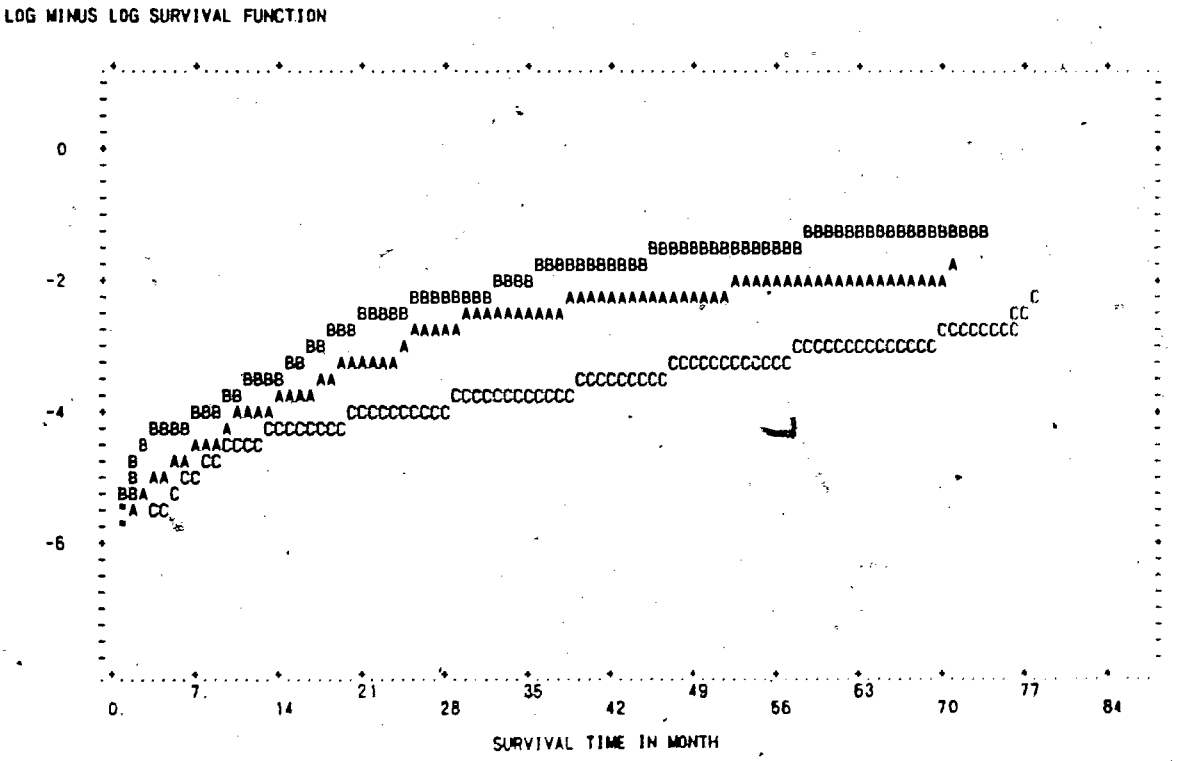


Figure 3.4.4 Model fit when depth is used

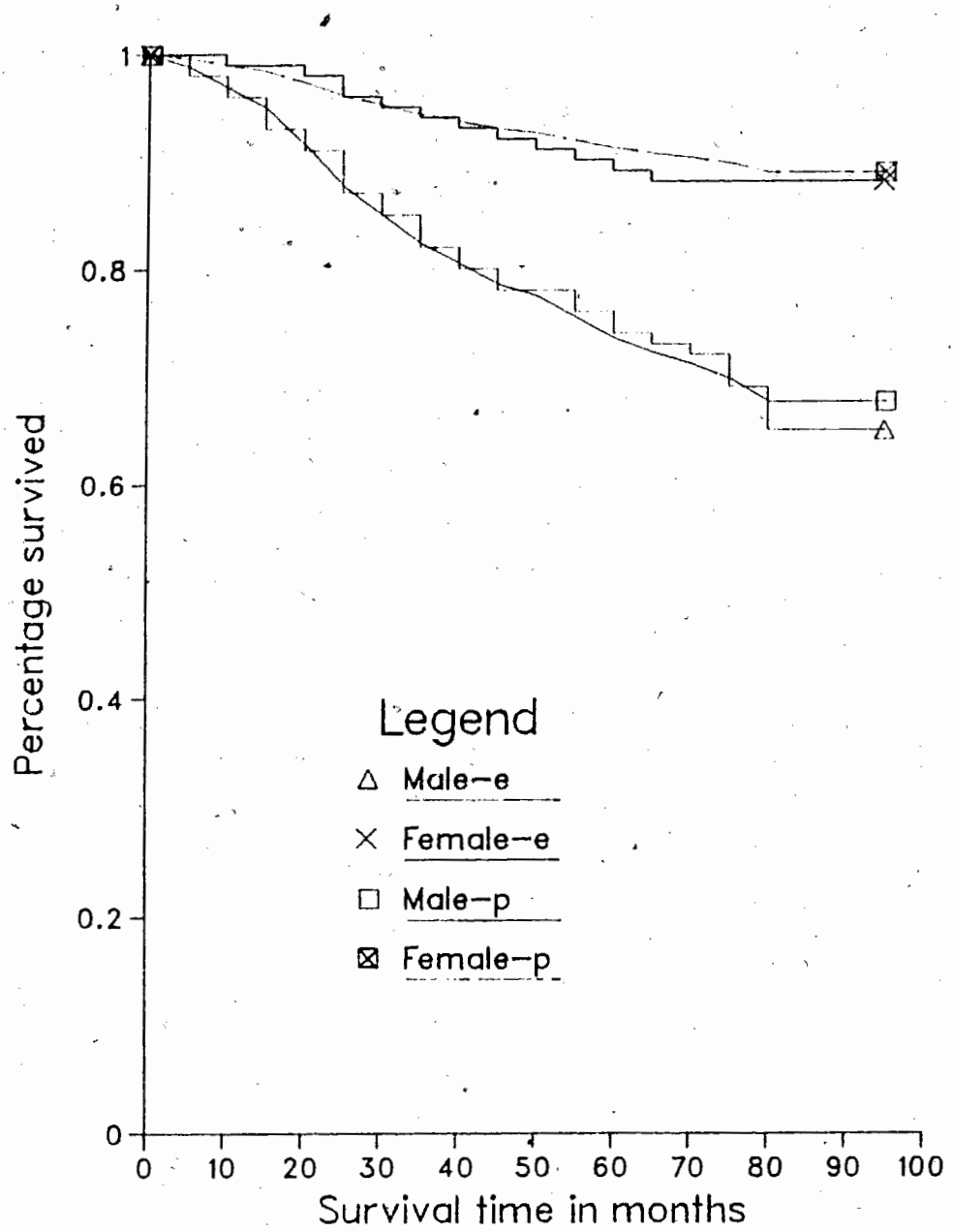


Figure 3.4.5 Model fit when Clark's levels II and III are used

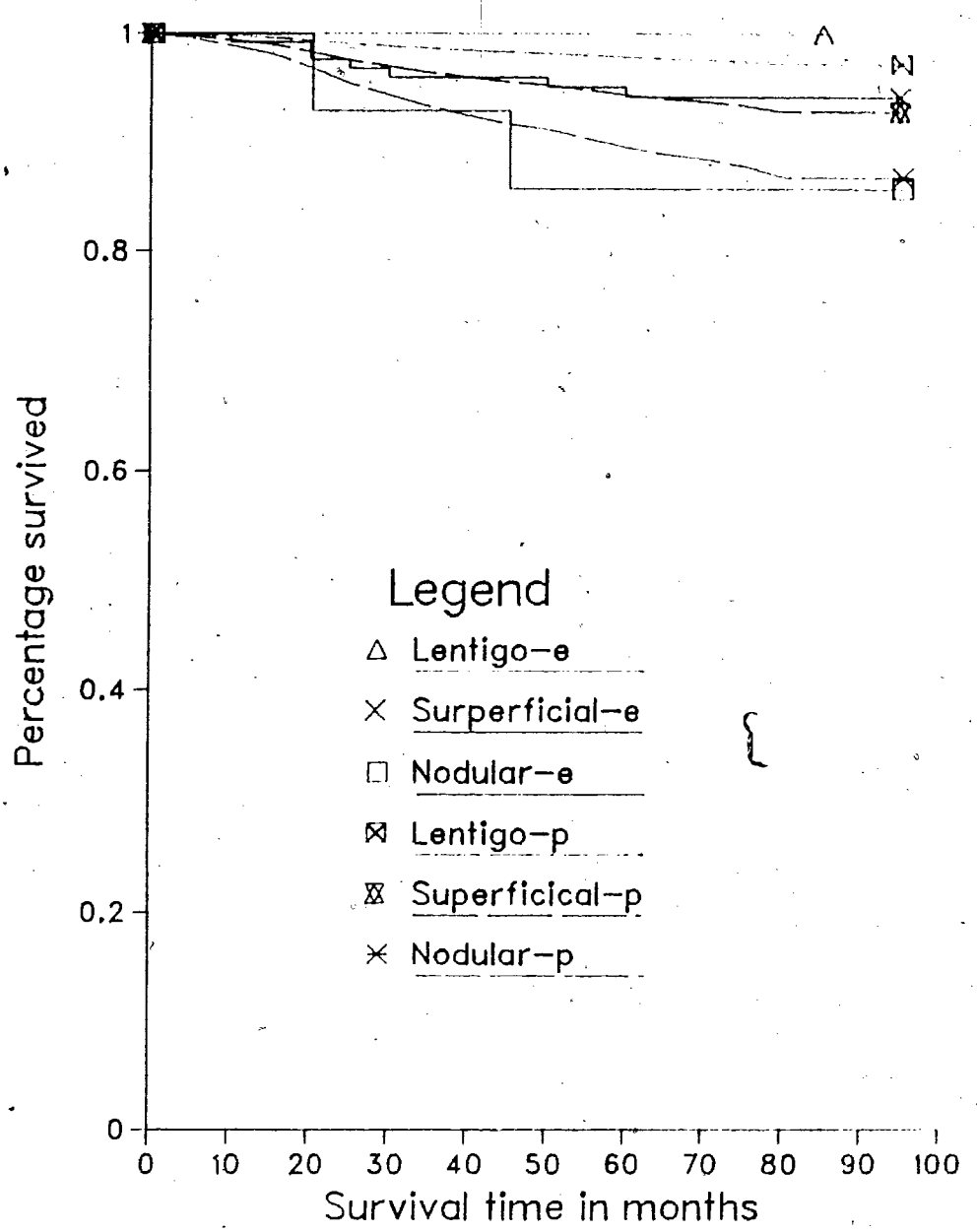
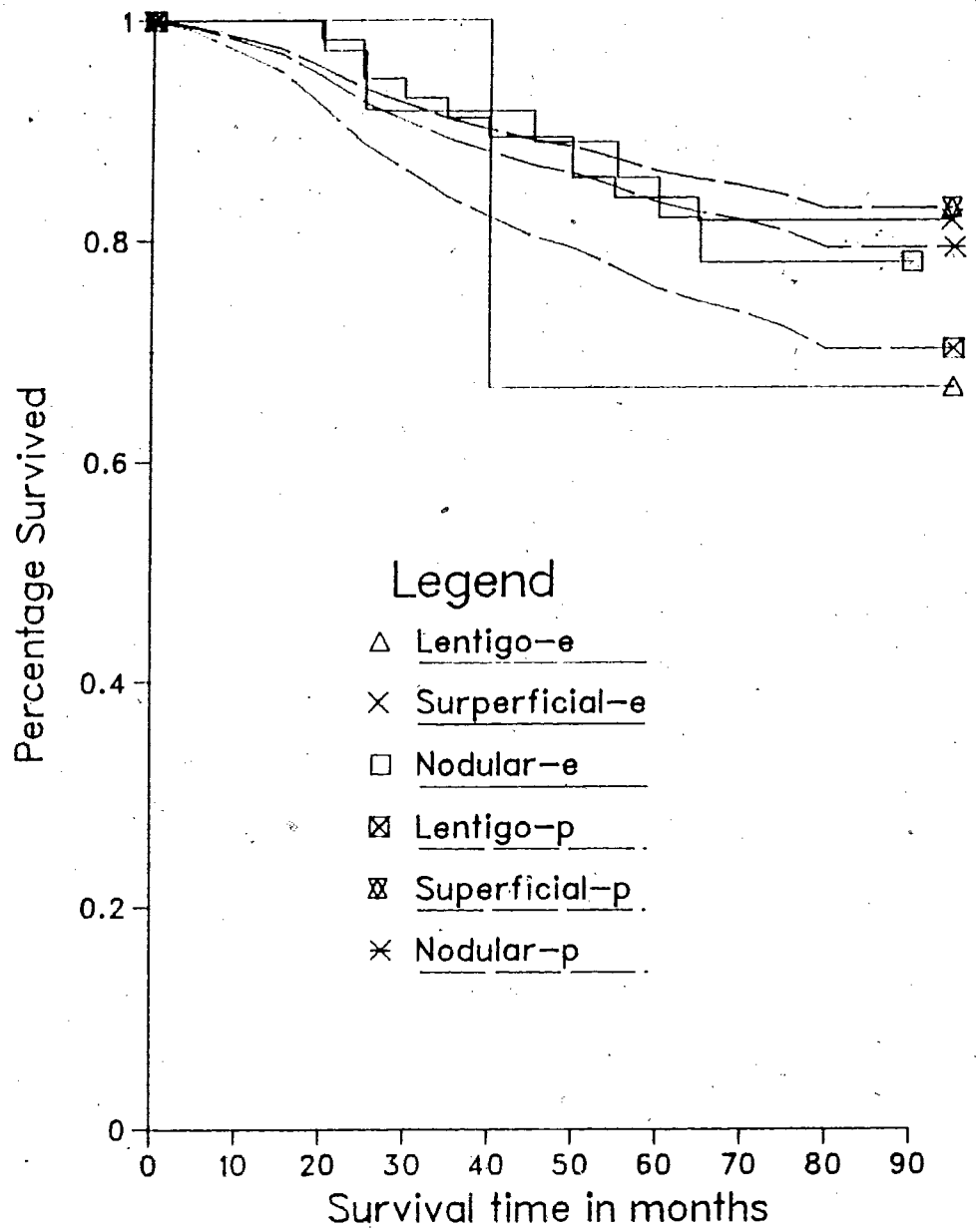


Figure 3.4.5 Model fit when Clark's levels IV and V are used



the proportionality assumption when only the significant factors are in the model. Figure 3.4.3 contains the plots.

Also, the results of testing time-dependent covariates shown in Table 3.4.3 happen to be the results needed to support the models containing only the significant factors. For these two models, the proportionality assumption is reasonable.

As another check of model fit, the estimated survival curves (by the life table method) and the predicted survival curves are compared on pages 74-75. With depth in the model, the comparison is focused on the men's and women's survival; with Clark's levels in the model, several combinations are considered. The comparisons are displayed in Figure 3.4.4 and Figure 3.4.5. Satisfactory agreement is observed.

In summary, based on the data and the multivariate analyses using Cox's regression model, the significant prognostic factors are tumour depth, sex and age if tumour depth is to be considered, or are Clark's levels of invasion, sex, age and cell type if Clark's levels of invasion are to be considered. (cf section 3.5.1).

3.4.2 Planned Hypothesis 2

In a paper by Koh *et al.* (1984) the authors claimed that they determined the following finding: after careful depth and location considerations were accounted for, there is no difference in prognosis between stage I LMM (lentigo maligna melanomas) and stage I non-LMM (superficial spreading and

nodular melanomas). The authors used 44 LMM patients out of 1,130 melanoma patients in their study as the basis. 44 people were chosen from the non-LMM patients to form 44 pairs with the 44 LMM patients matched on body site (face vs posterior scalp vs posterior neck vs arm vs leg) and depth (< 0.85, 0.86 to 1.69, 1.70 to 3.64, ≥ 3.65). Site was chosen to be matched because the authors highly appreciate its importance ([24]). An attempt was also made to match patients with similar age, length of follow-up and sex (no detail was provided). Patients who died from causes other than melanoma were recorded as being alive, and their survival time being censored. The log-rank test (Mantel's test) was used to compare survival between the LMM and the chosen non-LMM patients (P-value = 0.68); McNemar's test was used to analyze the 44 pairs for death (P-value = 0.37); Cox's proportional hazards model was used to analyze the 88 patients from the 44 pairs, focusing on histology (LMM vs non-LMM), age (≤ 60 vs > 60), location, depth, associated nevus and sex, with depth alone found to be significant (p-value = 0.0007).

Although the authors of the above paper did a fine job (matching and multivariate analysis), the following aspects seem to have not been covered well.

First, the results of the analysis of all 1,130 patients were not mentioned, especially, the significant factors identified for the whole study were not mentioned. Therefore the important general background is missing.

Secondly, lesion site, especially the concept of posterior scalp and posterior neck, was an important criterion used to form the 44 pairs, but it has not been widely accepted as an important prognostic factor. ([55],[62]).

Thirdly, of the 44 LMM patients, 40 had their tumours on head and neck (90.9%), 2 on arms and 2 on legs. So to match arms and legs does not make much sense in this case, and consequently the results thus obtained are unlikely to apply to arms and legs in general. For instance, it may not be true that LMM and non-LMM have the same prognosis if they have the same depth and are on the same site of arms and legs.

Fourthly, the cutpoints of depth determined from large data sets may not catch the characteristics of the LMM population, and to treat age as a categorical covariate in such a special way (≤ 60 vs > 60) can have similar problems. If this is so, the matching could have been misguided.

Finally, the sample size might not be large enough as to make the definite conclusion reached by the authors.

It should be emphasized that matching is a useful technique (controlling confounding effects or increasing precision of an analysis), but it needs careful consideration to use it ([14]). Also one should realize the extent within which results obtained through the matching technique will apply.

In this section, the following second planned hypothesis is tested based on the data, that is, Is there any difference in prognosis between lentigo maligna melanomas and other types of melanomas ?

It is shown first that it is not easy to construct good matches. For this, depth is listed in ascending order and sex, age, cell and status are listed accordingly in Table 3.4.6. One may try to form a few pairs, then one will see the point. The matching technique will not be employed in this project.

Table 3.4.6 Data for constructing matches.

SEX	SITE	DEPTH	CELL	AGE	STATUS
1	2	.0500	1	66	1
2	4	.0500	1	51	2
2	1	.100	1	63	2
1	1	.150	3	70	2
2	1	.150	3	52	2
2	3	.200	1	31	2
1	1	.200	1	84	1
1	2	.200	1	53	2
2	4	.200	1	17	0
2	2	.200	1	67	2
1	4	.200	1	57	2
1	1	.200	3	35	2
1	1	.200	1	42	2
2	2	.250	1	50	2
1	4	.250	1	51	2
2	1	.250	3	52	2
2	3	.270	3	56	2
1	3	.290	1	61	2
2	3	.300	1	46	0
1	1	.300	3	73	2
1	2	.300	1	71	2
2	3	.300	1	26	2
1	4	.300	1	30	0
2	2	.300	1	57	2
2	4	.300	1	57	2
1	1	.300	3	75	2
2	2	.300	1	43	2
1	2	.300	1	54	2

2	4	.300	1	69	2
1	3	.300	1	72	2
2	2	.310	1	27	2
1	3	.320	1	58	2
2	3	.340	1	71	2
1	2	.350	1	75	2
2	2	.350	1	55	2
1	4	.350	1	41	0
2	4	.350	1	67	2
2	3	.350	1	62	2
2	4	.360	1	40	2
1	3	.360	1	31	2
1	1	.370	1	50	1
2	3	.370	1	70	2
2	4	.370	1	45	2
1	2	.370	3	64	2
2	3	.370	1	56	2
1	3	.370	1	50	2
1	1	.380	3	77	2
1	1	.390	1	36	2
2	1	.400	3	51	2
2	1	.400	1	36	2
2	3	.400	1	24	2
1	3	.400	1	38	2
1	3	.400	1	32	2
1	3	.400	1	67	2
2	2	.400	1	59	2
2	3	.400	1	30	2
2	4	.400	1	36	2
1	2	.400	1	42	2
2	3	.400	1	64	2
2	1	.400	1	40	2
2	1	.400	1	47	2
1	1	.420	3	73	2
1	3	.430	1	45	2
2	4	.430	1	67	2
2	2	.430	1	35	2
2	3	.440	3	67	2
2	3	.450	1	62	2
2	3	.450	1	30	2
2	4	.450	1	57	2
1	1	.450	3	56	2
2	4	.450	1	55	2
2	1	.460	3	73	2
1	1	.470	3	68	1
2	3	.470	1	73	2
2	3	.470	1	74	2
1	2	.470	1	34	2
1	1	.480	1	31	2
2	3	.490	1	41	2
1	2	.500	1	51	2
2	4	.500	1	49	2
1	4	.500	1	80	1

2	4	.500	1	76	2
1	4	.500	1	48	2
1	1	.500	1	43	2
2	2	.500	1	34	2
2	4	.500	1	68	1
1	2	.500	1	42	0
1	2	.500	1	70	1
2	3	.500	1	49	2
1	2	.500	1	23	2
2	4	.500	1	33	2
2	4	.520	1	44	2
2	3	.520	1	53	2
2	4	.520	1	34	2
1	2	.520	1	41	2
2	1	.520	1	41	2
2	4	.530	1	67	2
2	3	.530	1	61	2
2	3	.530	1	38	2
1	2	.540	1	50	1
1	3	.540	1	40	2
2	2	.540	1	43	2
1	4	.550	1	50	2
2	2	.550	1	41	2
2	2	.550	1	32	2
2	2	.550	1	29	2

(portion of the whole data set)

The general features of LMM and non-LMM patients are described below.

Table 3.4.7 Description of LMM and non-LMM patients.

	LMM		non-LMM	
Depth / Age				
median	0.65	69.00	1.07	51.00
mean	1.24	67.17	1.69	51.24
s.e.	1.53	12.87	1.60	16.88
mimimum	0.15	35	0.05	12
maximum	6.00	100	9.00	93

skewness	2.10	-0.35	1.85	0.14
kurtosis	3.47	0.78	3.62	-0.68
Sex				
male	20		160	
female	15		235	
Site				
hdnk	26		61	
others	9		334	
Diff				
levels I and II	20		315	
level III	15		80	
Status				
dead	4		74	
censored	31		321	
Type specific death rate	11.4%		18.7%	

LMM patients tend to have thinner invasion depth and older age than non-LMM. The male to female ratio is 1:0.75 for LMM and 1:1.47 for non-LMM. Most LMM are on head and neck and the type specific death rate is high with non-LMM.

Univariate analysis then shows that the five-year survival for LMM and non-LMM are 94.20% and 82.03%, respectively. The results are in Table 3.4.8.

Univariate analysis then shows that the five-year survival for LMM and non-LMM are 94.20% and 82.03%, respectively. The results are in Table 3.4.8.

Table 3.4.8 Univariate analysis of LMM vs non-LMM.

Summary Table

	Total	Dead	Lost	Censored	Percent Censored
non-LMM	395	74	16	305	0.8127
LMM	35	4	2	29	0.8857
	430	78	18	334	TOTAL

Test Statistics

	Statistic	df	P-value
Breslow	2.551	1	0.1102
Mantel	1.362	1	0.2431

CUMULATIVE PROPORTION SURVIVING

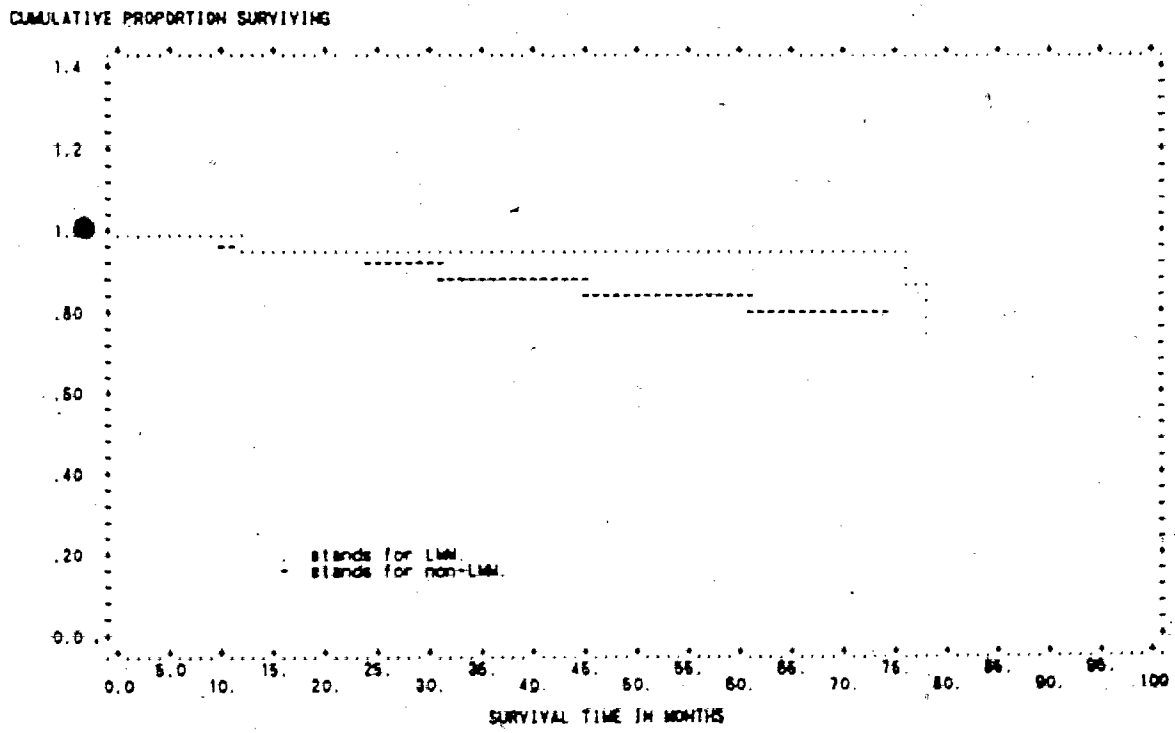
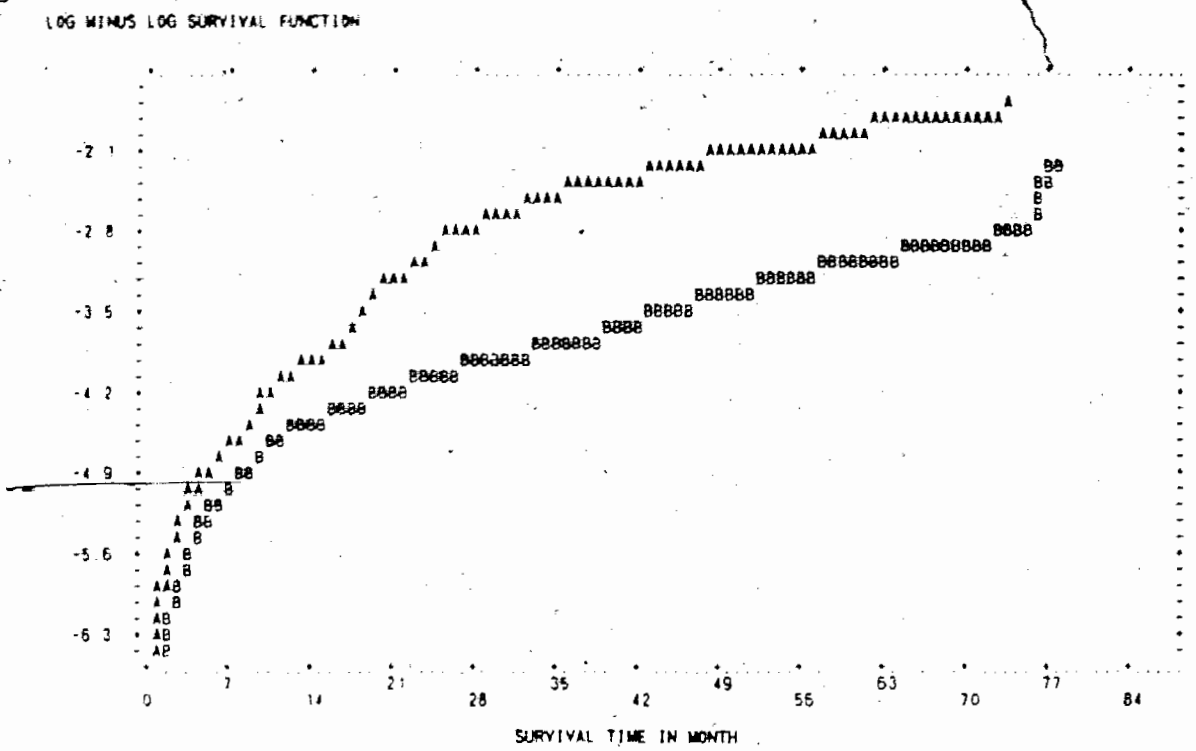


Figure 3.4.6 Checking proportionality for LMM vs non-LMM (under depth: upper plot for all factors, lower plot for significant factors).



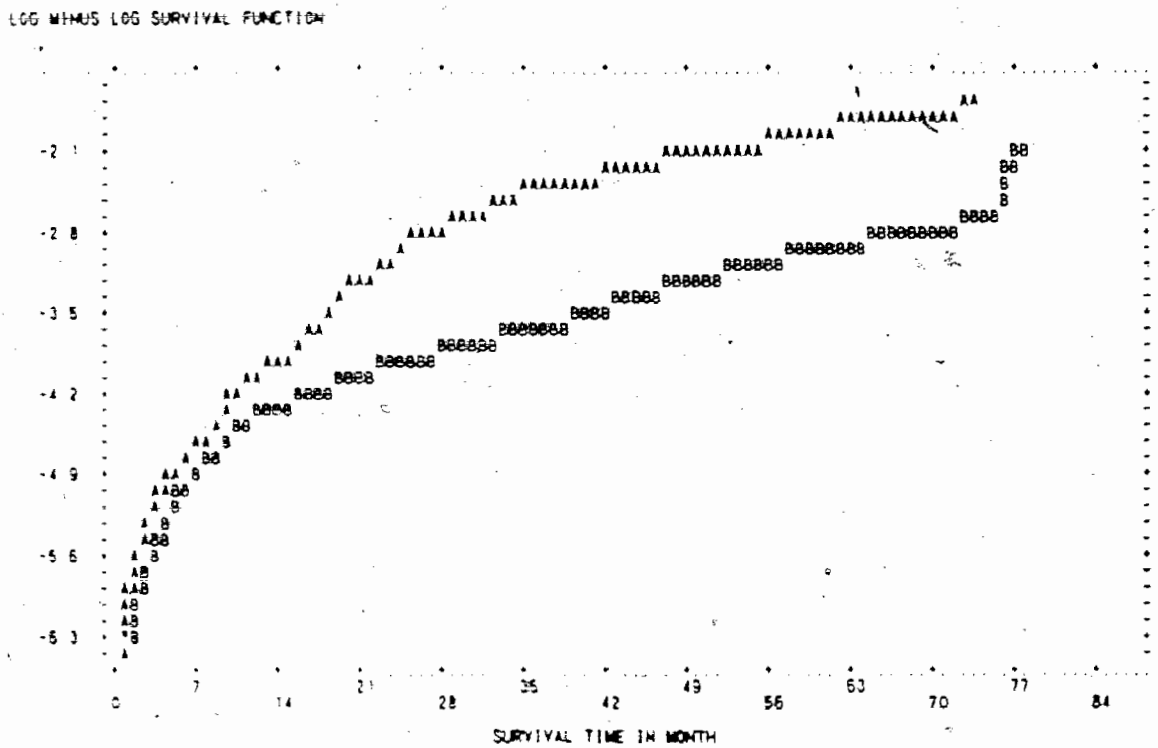
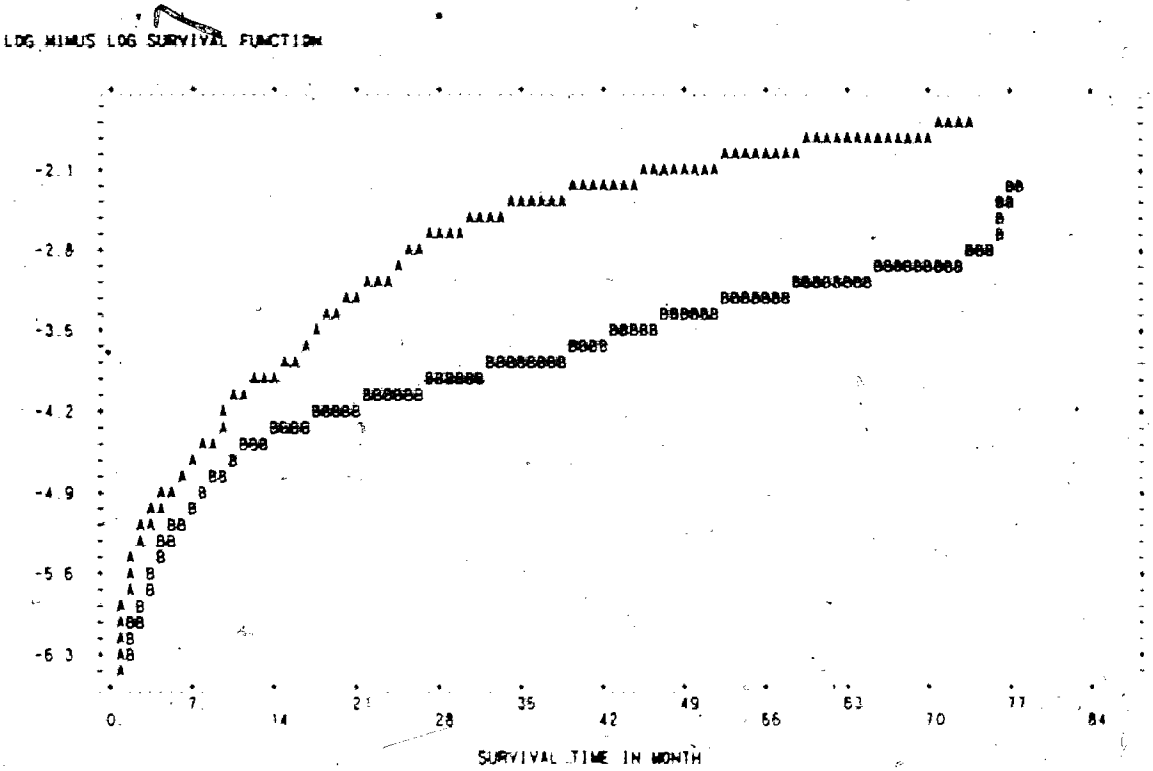
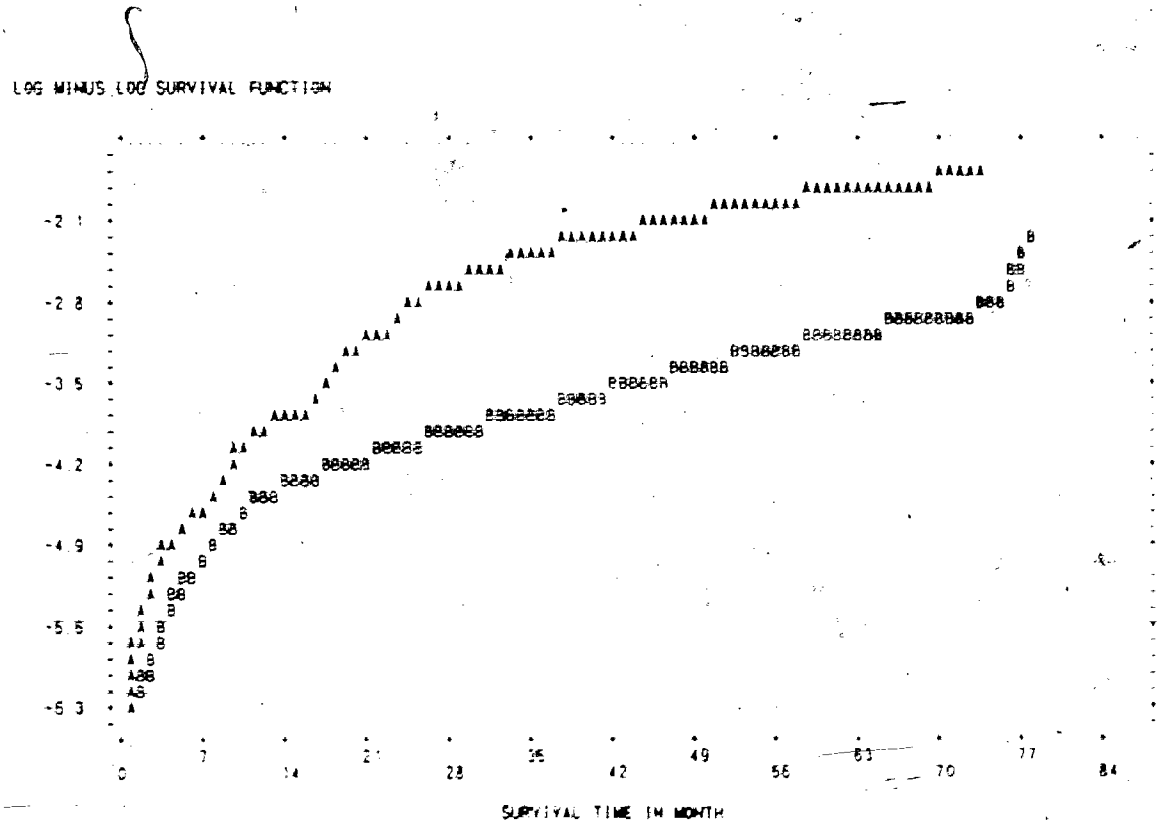


Figure 3.4.6 continued (under Clark's: upper plot for all factors, lower plot for significant factors).





The upper plots are produced when all the covariates used in the previous section are included into the model. The effect of this covariate (LMM vs non-LMM) on survival is not proportional for the first two years and the last few months in the range of survival time. Since one can not change the coding (LMM vs non-LMM) and since for a large portion of the range of survival, the proportionality assumption seems to be reasonable, multivariate analyses will be performed using the above coding.

Two backward selections (led by tumour depth and Clark's levels, respectively) are carried out with the results shown in the following two tables.

Table 3.4.9 Backward selection led by depth.

Step no.	Covariates left out	Chi-square to leave out	df	P-values	Global test P-values
1	ulceration	0.014	1	0.9070	0.0000
2	lymp	0.021	1	0.8835	0.0000
3	mito	0.037	1	0.8480	0.0000
4	diff	0.694	1	0.4050	0.0000
5	site	2.326	1	0.1272	0.0000
6	no covariate to be left out.				

The kept-in covariates are:

Name	P-values
depth	0.0
age	0.0001
sex	0.00

LMM	0.0360	

The estimation of the kept-in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.
depth	0.3133	0.0514	6.1005
age	0.0271	0.0072	3.7529
sex(female)	-1.2152	0.2431	-5.0000
LMM	-0.9721	0.5280	-1.8412

Table 3.4.10 Backward selection led by Clark's levels.

Step no.	Covariates left out	Chi-square to leave out	df	P-values	Global test P-values
1	lymp	0.000	1	0.9918	0.0000
2	diff	0.072	1	0.7880	0.0000
3	mito	1.552	1	0.2129	0.0000
4	ulceration	3.115	1	0.0776	0.0000
5	site	3.611	1	0.0572	0.0000
6	no covariate to be left out.				

The kept-in covariates are:

Name	P-values
Clark's	0.0002
age	0.0001
sex	0.0000
LMM	0.0094

The estimation of the kept-in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.

Clark(IV,V)	0.8739	0.2430	3.5966
age	0.0285	0.0076	3.7628
sex(female)	-1.1377	0.2397	-4.7466
LMM	-1.1673	0.5264	-2.2175

In the first selection, depth, age, sex and celltype are selected; in the second selection, Clark's levels, age, sex and celltype are selected.

The lower plots shown in Figure 3.4.6 correspond to the covariate (LMM vs non-LMM) and are produced when only the above selected covariates are in the model. Some, but not dramatic, improvement can be seen. Table 3.4.11 contains the results of testing time-dependent covariates to check the proportionality assumption. The large P-values support the use of the Cox model.

Table 3.4.11 Testing time-dependent covariates for LMM vs non-LMM (z = ln(survival time) and the time-dependent covariates are tested simultaneously).

When DEPTH is used				
Covariates in the model	Time-dependent covariates tested	Test statistics	df	P-values
depth				
sex2	z1 = sex2*z	lratio	2	0.2887
age		score	2	0.3400
LMM	z2 = LMM*z	Wald	2	0.3487

z1

When CLARK'S LEVELS are used

clark4	z1 = clark4*z			
sex2	z2 = sex2*z	lratio	3	0.2608
age		score	3	0.2916
LMM	z3 = LMM*z	Wald	3	0.3061
z1,z2,z3				

To summarize, the LMM patients have different characteristics from those of the non-LMM patients, especially, the LMM patients have better prognosis than the non-LMM patients.

3.4.3 Planned Hypothesis 3

In another paper by C. L. Day *et al.* (1982), the authors studied 203 patients with clinical stage I melanoma and with primary tumours 0.76 to 1.69 mm thick. 12 deaths were observed and 11 of them occured in patients with primary tumours located on the upper Back, posterior Arm, posterior Neck and posterior Scalp, thus the **BANS** concept. A poorer survival of the BANS patients was reported (84% BANS vs 99% non-BANS). The authors concluded their paper by claiming that their finding shows little sex difference in survival after accounting for depth, 'subsites,' and clinical stage.

Since the publication of the above paper, a series of studies have been conducted and conflicting results reported

([13],[55],[61]). This section studies the third planned hypothesis: Do the BANS areas have worse prognosis than other areas ?

The BANS patients and the non-BANS patients are compared in three depth ranges. The first range is the one used by Day *et al.*; the second range covers, according to the literature, most serious melanomas except the extreme cases; the third range is used to provide an overall picture.

Table 3.4.12 Depth from 0.76 to 1.69 mm (range I).

Depth and Age	BANS		non-BANS	
frequency	53	53	67	67
mean	1.15	46.2	1.13	53.6
standard dev.	0.24	14.3	0.26	16.4
minimum	0.80	22	0.77	19
maximum	1.60	83	1.65	100
range	0.80	61	0.88	81
Sex				
male	29		24	
female	24		43	
Celltype				
lentigo	0		8	
super	43		50	
nodular	10		9	
Status				
dead	11		7	

censored	42	60
Cell specific death rate due to nodular	23.3%	13.4%
Site specific death rate	20.7%	10.4%

Table 3.4.13 Depth from 0.76 to 3.65 mm (range II).

Depth and Age	BANS		non-BANS	
frequency	90	90	123	123
mean	1.70	48.6	1.74	55.2
standard dev.	0.75	16.2	0.78	17.3
minimum	0.80	21	0.77	12
maximum	3.60	86	3.65	100
range	2.80	65	2.88	88
Sex				
male	51		47	
female	39		76	
Celltype				
lentigo	0		11	
super	61		74	
nodular	29		38	
Status				
dead	23		22	
censored	67		111	

Cell specific death rate due to nodular	32.2%	30.9%
Site specific death rate	25.6%	16.5%

Table 3.4.14 The whole depth range (range III).

Depth and Age	BANS		non-BANS	
frequency	171	171	259	259
mean	1.55	48.9	1.72	54.9
standard dev.	1.42	16.0	1.71	17.5
minimum	0.20	21	0.05	12
maximum	8.00	86	9.00	100
range	7.80	65	8.95	88
Sex				
male	87		93	
female	84		166	
Celltype				
lentigo	2		33	
super	128		170	
nodular	41		56	
Status				
dead	32		46	
censored	139		213	
Cell specific				

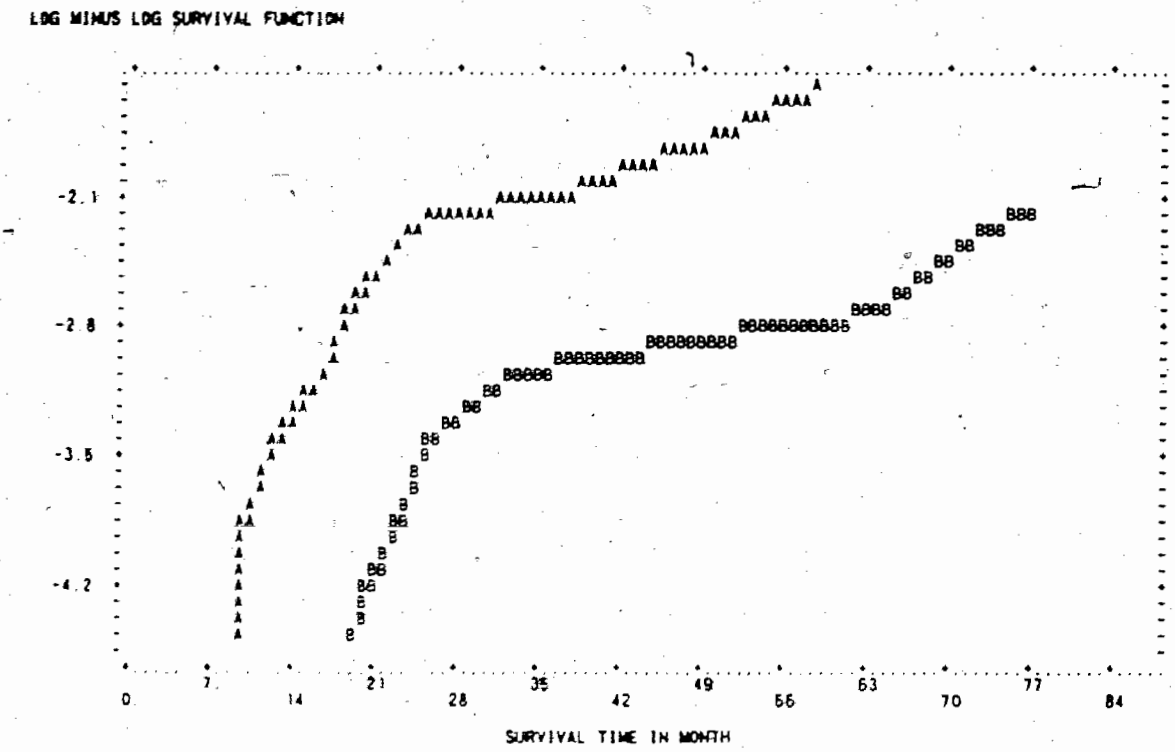
death rate due to nodular	24%	21.6%
Site specific death rate	18.7%	17.8%

In the BANS area, more male cases were observed but in the whole data set there are 250 females and only 180 males. In the three depth ranges, the average depth of BANS patients changes from being greater than that of non-BANS patients to being smaller than that of non-BANS patients. The cell specific death rate due to nodular type of melanoma (nodular type has the worst prognosis) is much higher with BANS in the first range but becomes slightly higher in the second and third ranges. Also, the site specific death rate is noticablely higher with BANS in the first and the second ranges than with non-BANS.

These observations seem to be supporting the BANS concept, but they are univariate descriptions only. Six multivariate analyses are performed, using the approach used in the previous two sections. In these analyses, site is recoded as BANS vs non-BANS. Also, celltype is recoded as lentigo plus superficial spreading vs nodular to allow each level to have a necessary number of cases to fit the Cox model.

As before, pigmentation and regression are not included in the following multivariate analyses. The proportionality assumption is checked for site (BANS vs non-BANS) and celltype only in Figure 3.4.7 (pages 95-100).

Figure 3.4.7 Checking proportionality for the BANS concept (under depth and range I: upper plot for site, lower plot for celltype).



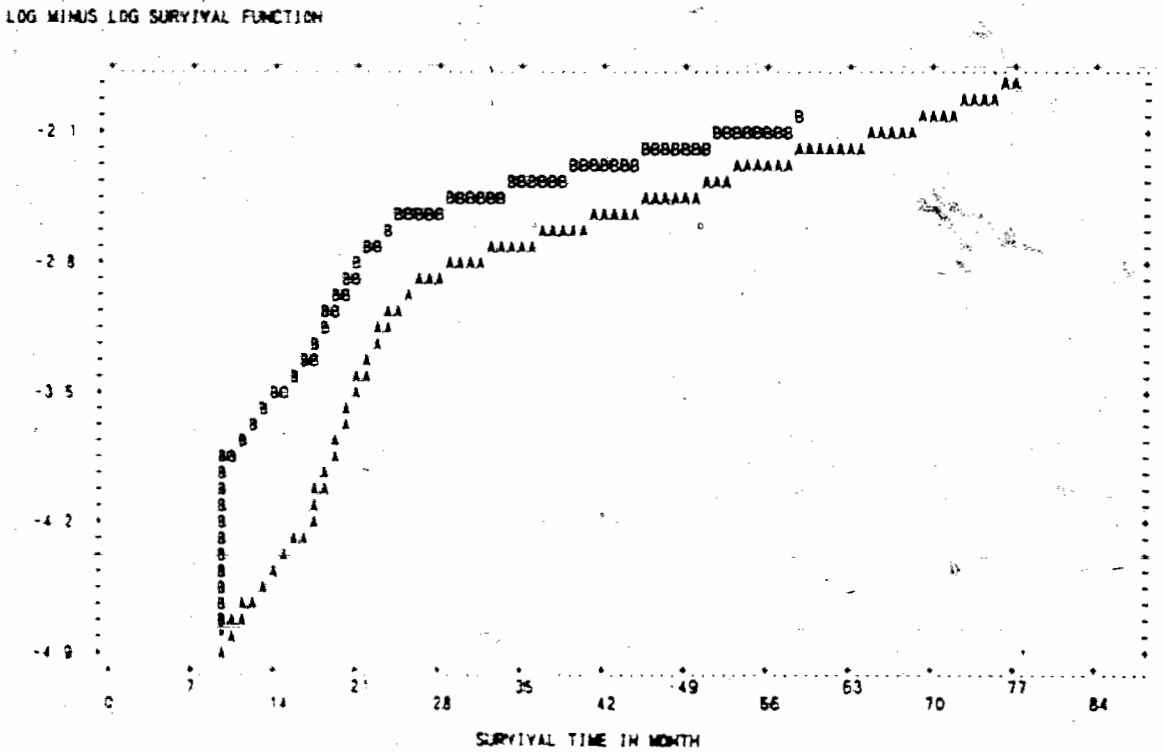
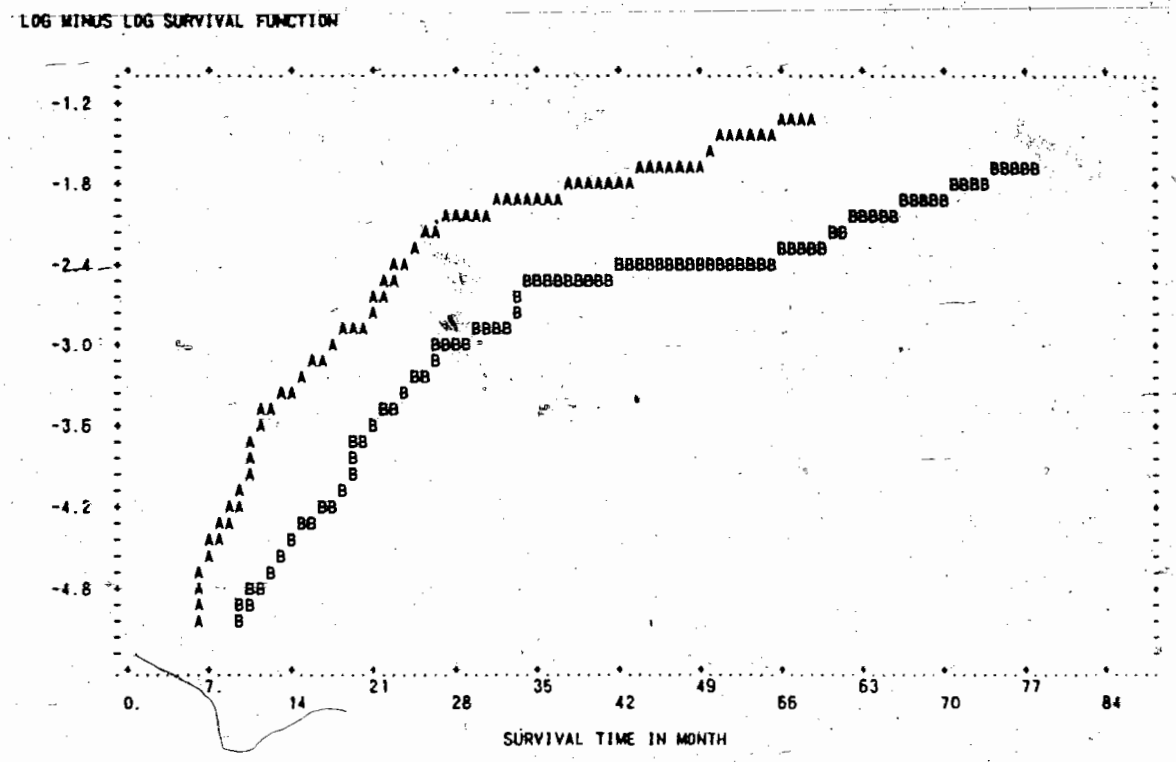


Figure 3.4.7 continued (under depth and range II: upper plot for site, lower plot for celltype).



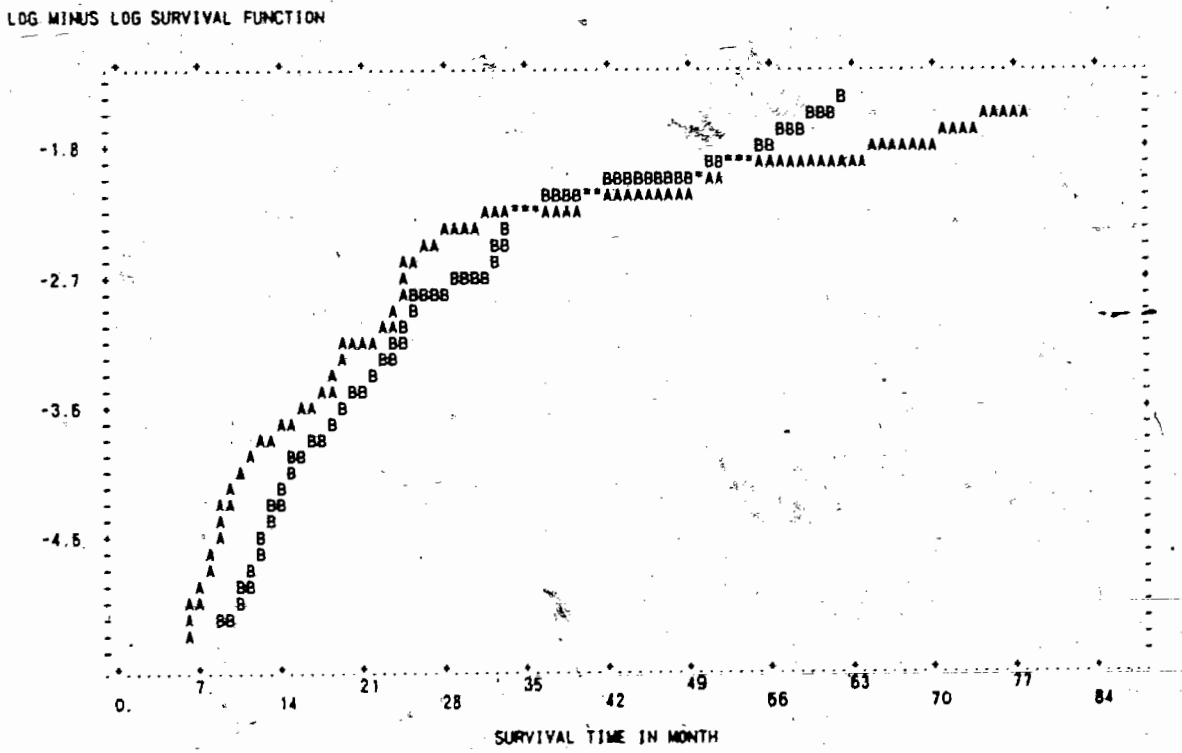
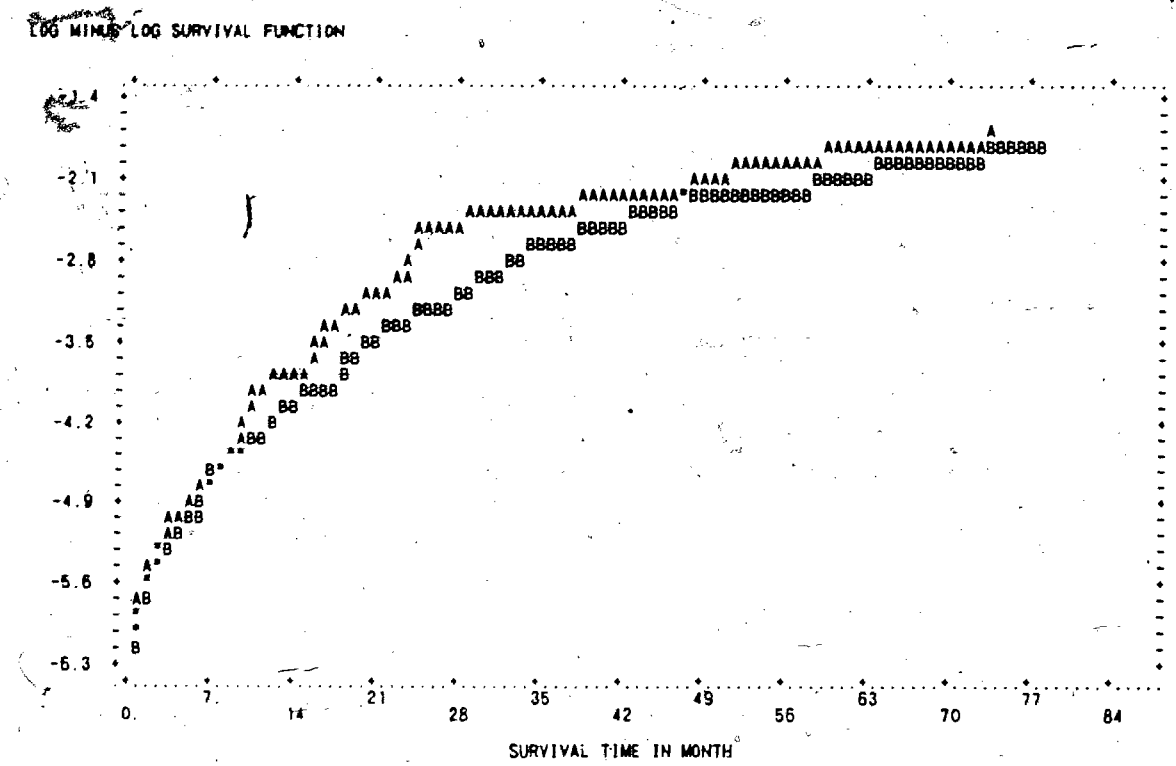


Figure 3.4.7 continued (under depth and range III: upper plot for site, lower plot for celltype).



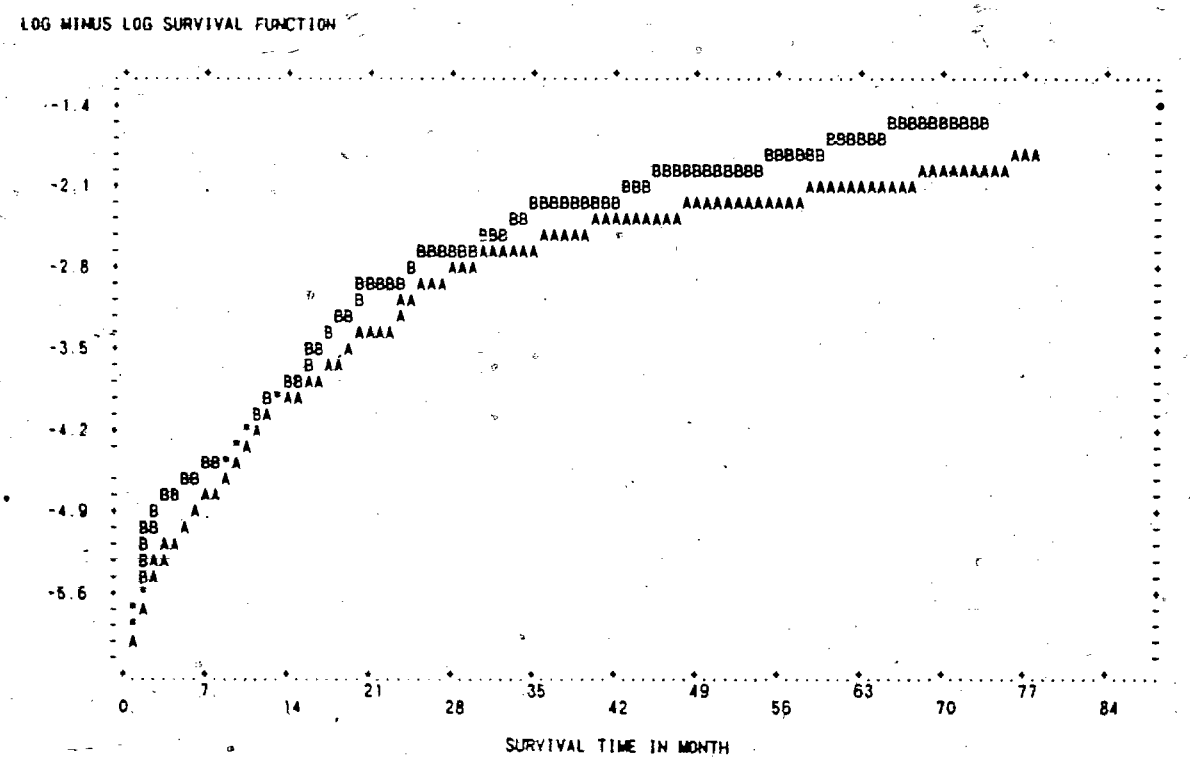
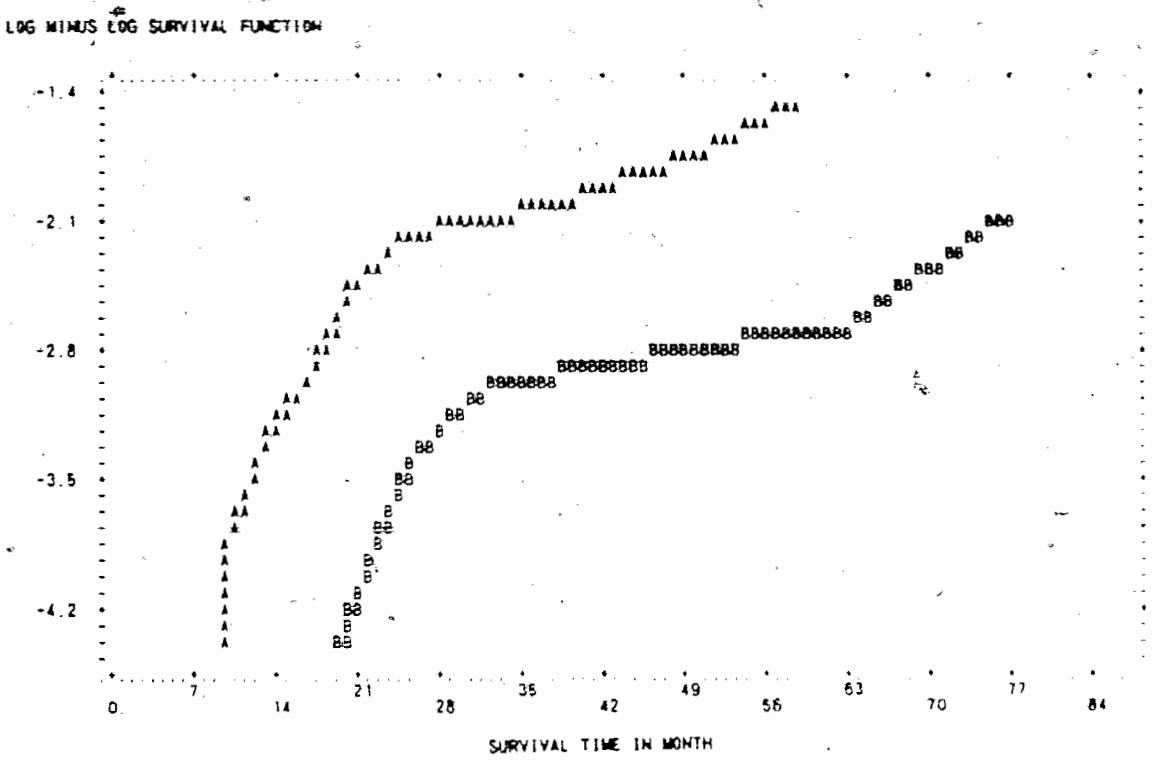


Figure 3.4.7 continued (under Clark's and range I: upper plot for site, lower plot for celltype).



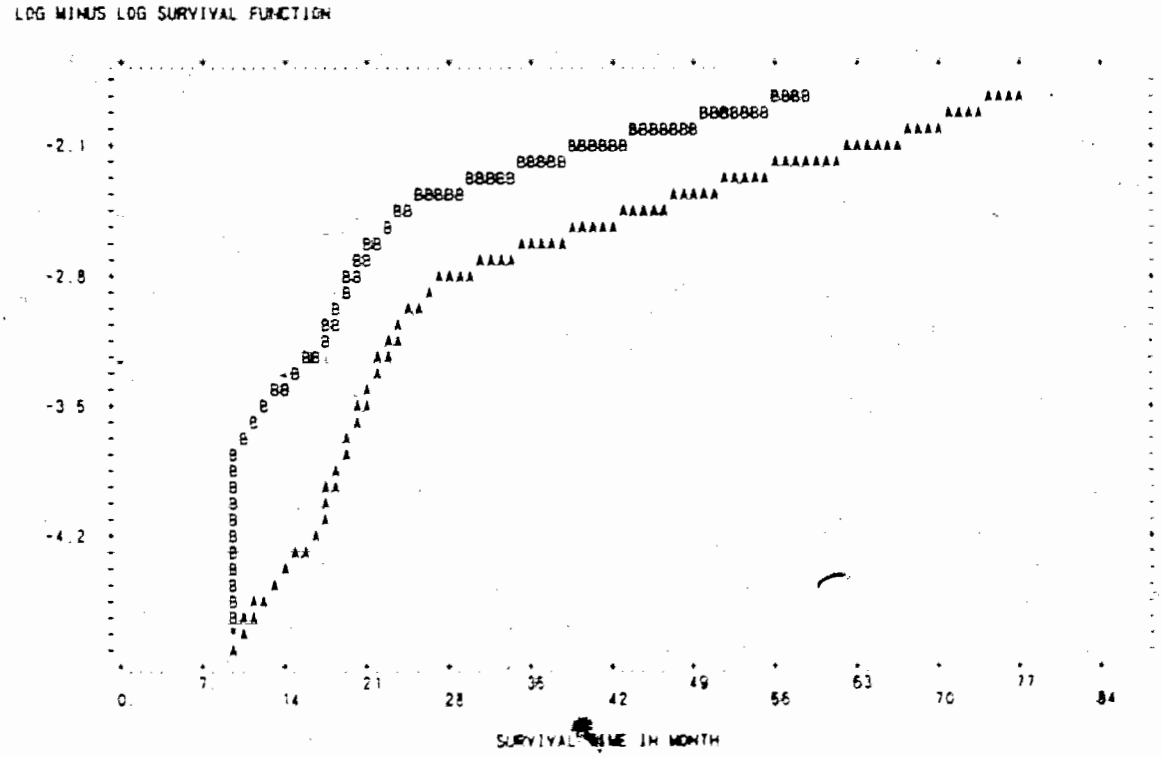
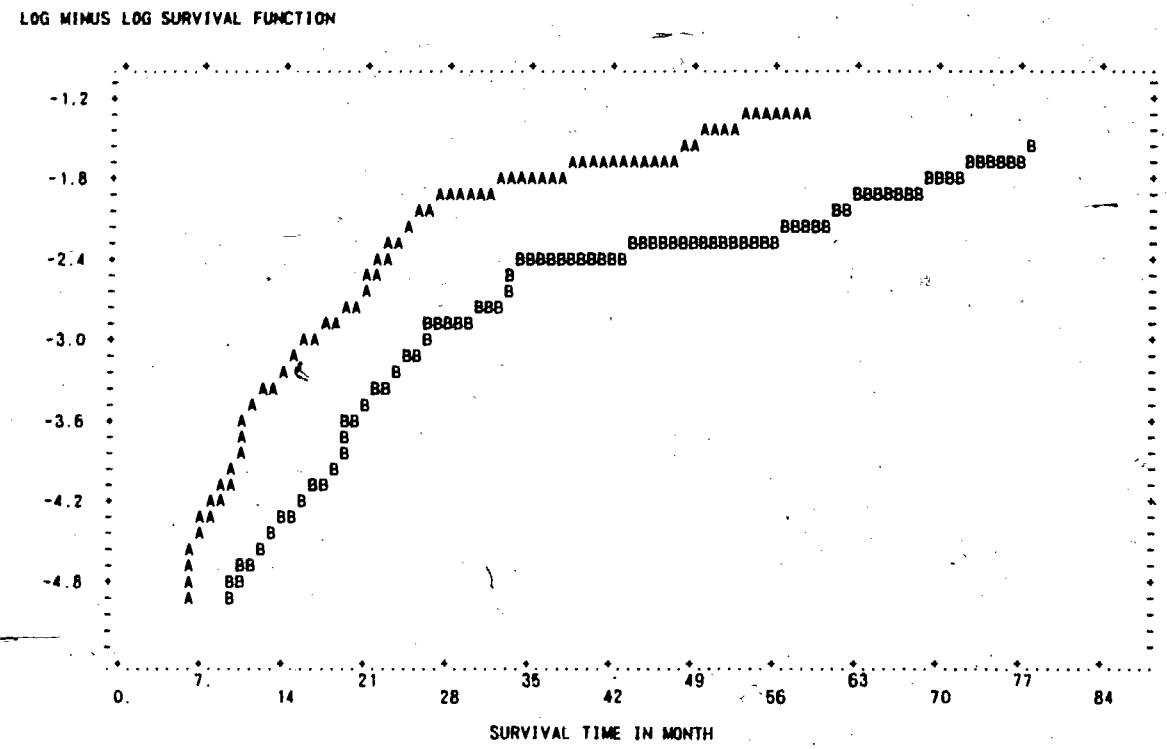


Figure 3.4.7 continued (under Clark's and range II: upper plot for site, lower plot for celltype).



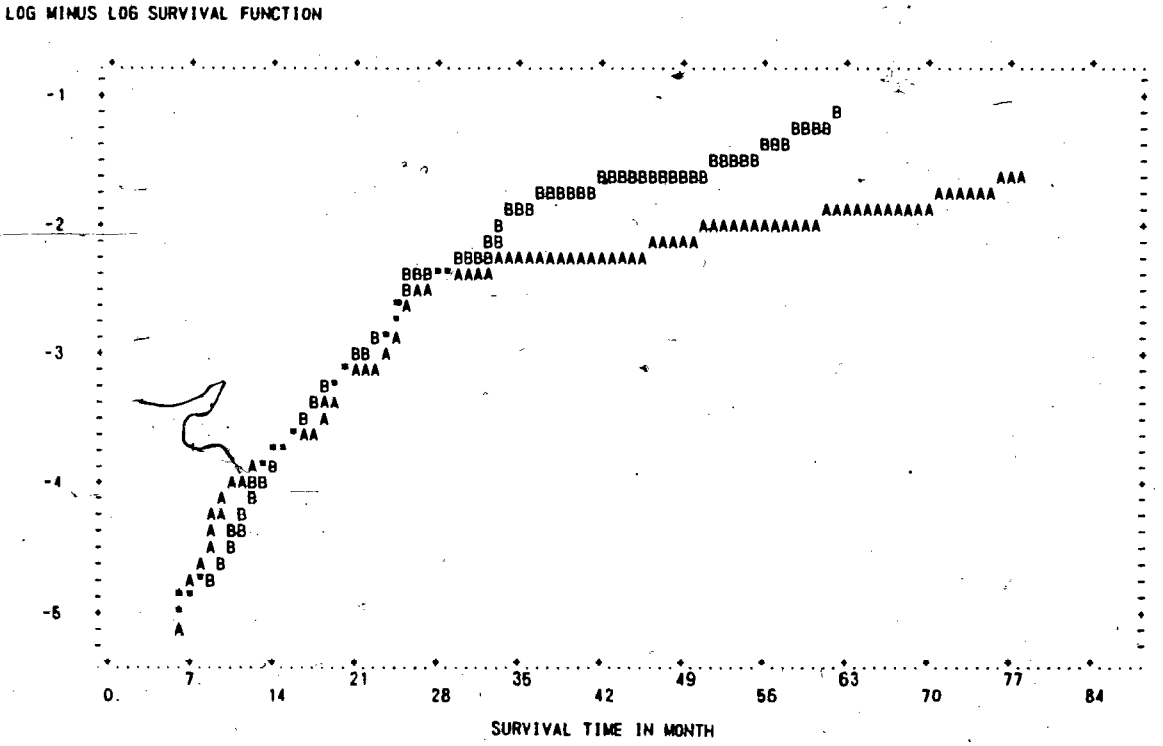
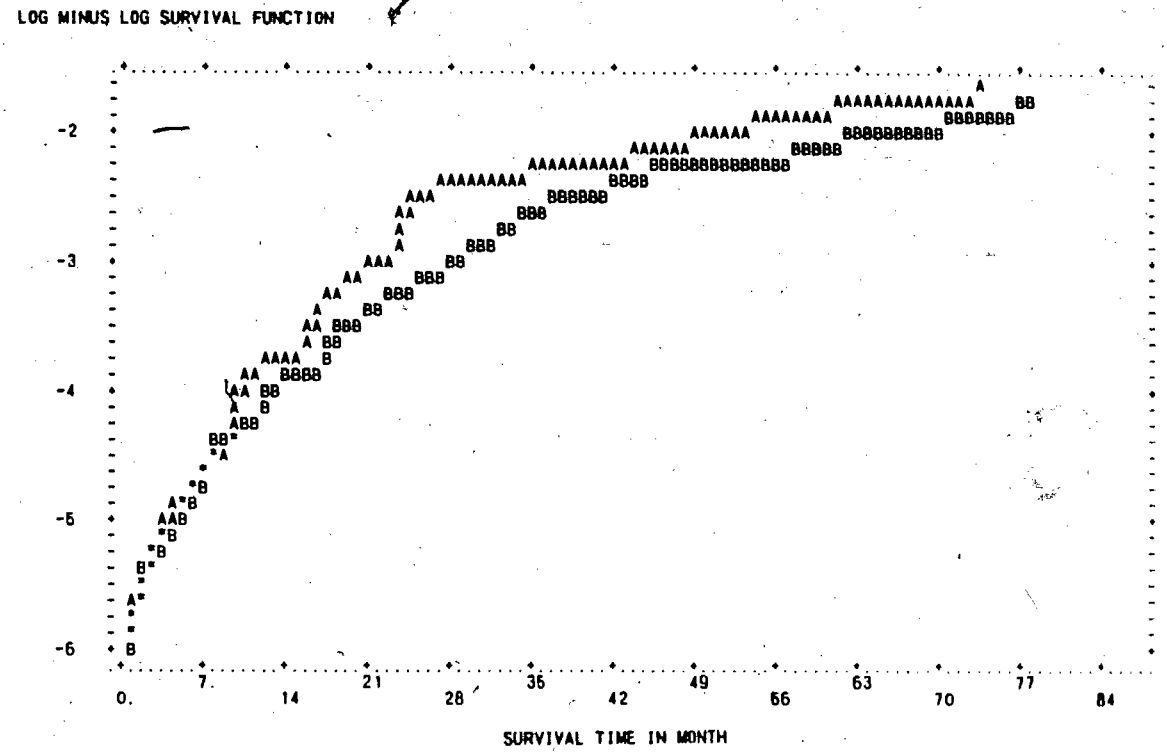
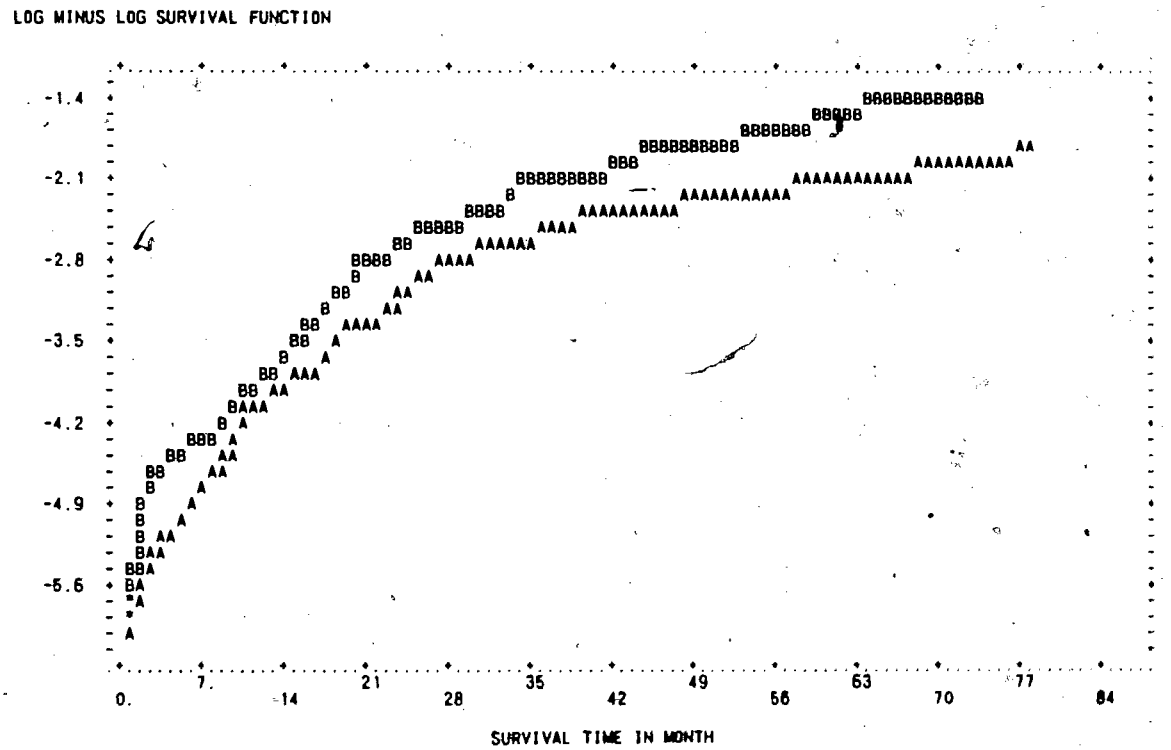


Figure 3.4.7 continued (under Clark's and range III: upper plot for site, lower plot for celltype).





The proportionality assumption seems to hold for all but two cases. Six backward selections are then carried out (Table 3.4.15). Differentiation can not be included into the model for the first two depth ranges because its inclusion causes the so-called monotonicity problem ([10]). Therefore, for depth range I and depth range II differentiation is not used.

Table 3.4.15 Backward selections for the three depth ranges.

When DEPTH is used		When CLARK'S LEVELS are used	
Covariates kept in	P-values	covariates kept in	P-values
	Range I		
none	/	none	/
	Range II		
depth	0.0000	sex	0.0016
sex	0.0041	cell	0.0472
	Range III		
depth	0.0	Clark's	0.0054
age	0.0007	age	0.0012
sex	0.0000	sex	0.0000

Interesting enough, no factor is significant for the first depth range. In fact, the global chi-square test for model fit had never been significant once even when depth alone was included into the model. There are two explanations. The first is that the exponential form of the Cox regression model is not

suitable for this depth range. This is possible, but is unlikely because Cox's model does not specify the underlying distribution and thus is pretty robust and flexible. The second explanation is that there really does not exist a significant factor in the first range, at least BANS is not significant. This second explanation is prefered because it is what the global chi-square test says. In addition, BANS is not kept in in the second and the third depth range analyses.

A few comments about the depth cut points used so far seem to be necessary. By reading through the literature, one can find out that cutpoints like 0.85, 1.70 and 3.65, which have been cited and used very often, were first chosen by Day and another two co-authors in a paper titled " The Natural Break Points for Primary-Tumor Depth in Clinical Stage I Melanoma." The main approach Day *et al.* used to reach the above cutpoints was a computer-based searching. As Day *et al.*, wrote: " A separate dichotomous variable was created for each depth value at 0.05 mm increments from 0.05 to 10.0 mm. These depth variables were then simultaneously tested with logistic regression. The 95 per cent confidence intervals were generated with the jackknife technique." "... the risk increases with depth in quantum jumps, analogous to a rising staircase, with four 'stair steps' or categories with the following boundaries: < 0.85 mm, 0.85 mm through 1.69 mm, 1.70 mm through 3.60 mm and ≥ 3.60 mm."

Without the intention to go into the detail any further, only the other side of the coin is mentioned here. It has been

noticed that important depth ranges vary from place to place ([12],[29],[61]). An unresolved important issue is how to get biological evidence to support or disprove certain cutpoint values. Incidently, it should be mentioned that the word, "Natural", used by Day *et al*. in their paper does not have any natural biological meaning.

In conclusion, the above analyses provide a negative case study on the BANS concept.

3.5 Fitting Cox's Regression Model and Exploring New Hypotheses

3.5.1 New Hypothesis 1

In the course of looking for significant prognostic factors of malignant melanoma, various trials have been made (see [12] for a review). A central issue involved was how to characterize this deadly neoplasm on objective grounds so that the results thus obtained could be used by clinicians to design and choose appropriate treatments. Many classifications were suggested but none of them was accepted outside the centre from which it originated. This situation lasted for about thirty years until in 1969, Clark *et al*. classified malignant melanoma into five levels of invasion, which were found to be of practical relevance and later became internationally accepted.

After Clark *et al*.'s work, the field changed rapidly. In 1970, V.J. Mcgovern stressed the importance of the growth patterns of primary melanoma (or cell types), and in the same

year, A. Breslow claimed the importance of the measurement of maximum depth of primary melanoma. In 1978, the first study comparing Clark's levels of invasion and Breslow's tumour depth was published, and the superiority of tumour depth over levels of invasion was claimed in terms of being more objective, more reliable, and reproducible ([12]). Today Breslow's tumour depth has become accepted as the most important prognostic criterion of stage I malignant melanoma.

In this section, the details of how the superiority of tumour depth over levels of invasion was established are reviewed. Following this, the Cox model is fitted to the data to corroborate the superiority. Tumour depth and Clark's levels of invasion are studied in the same model for the first time in this project.

In one of the first two papers comparing tumour depth with levels of invasion[2], C.M. Balch *et al*. (1978) proceeded as below.

1. Kaplan and Meier's method (or product-limit method) was used to estimate five year survival. First, levels of invasion were fixed and within each level depth was divided into two to three ranges. Secondly, four depth ranges were fixed and within each range different levels of invasion were compared. The results are summarized in Figure 3.5.1.

---

[2] Another paper is in Italian. It has been arranged to get a copy in English.

Figure 3.5.1 Comparison: Part A

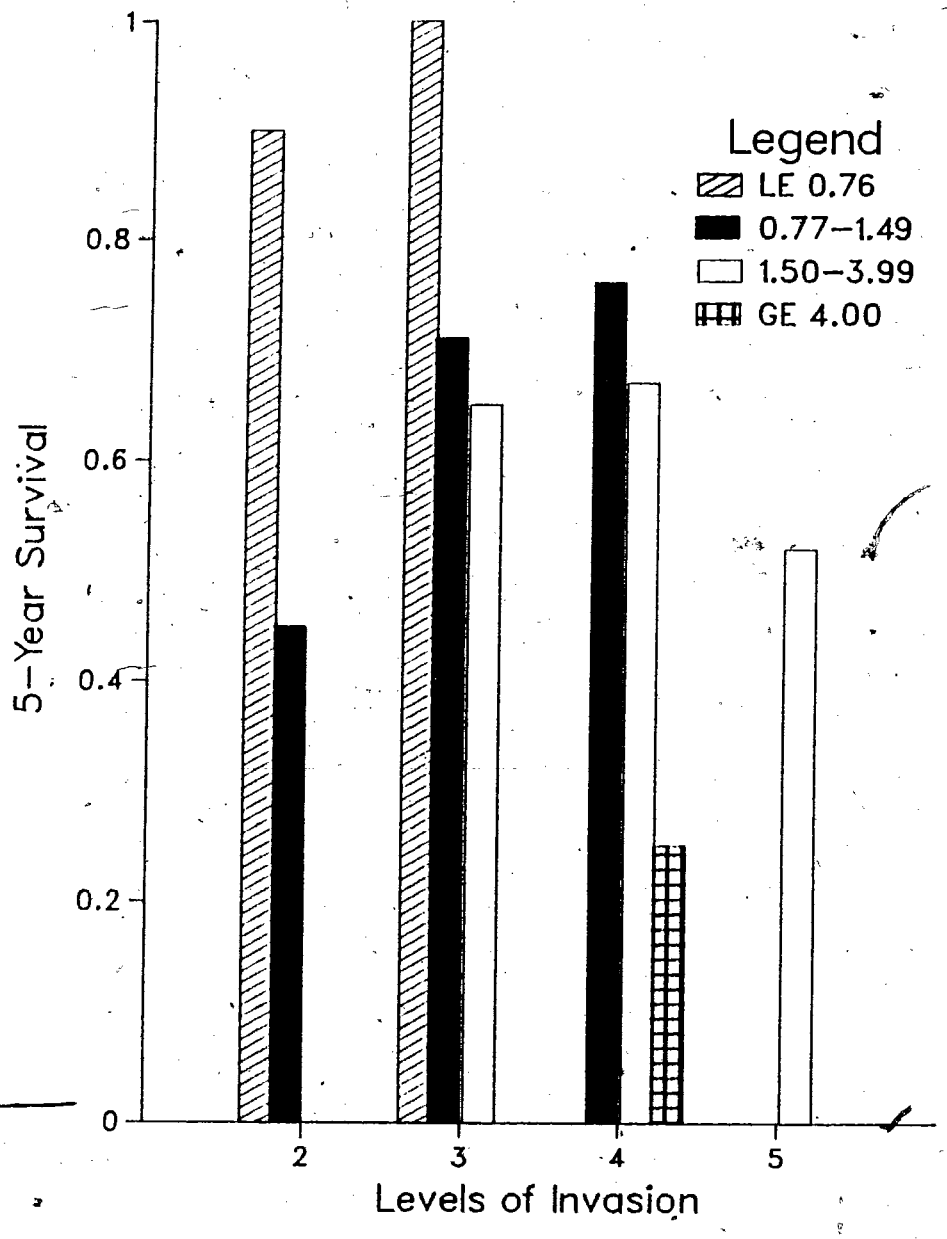
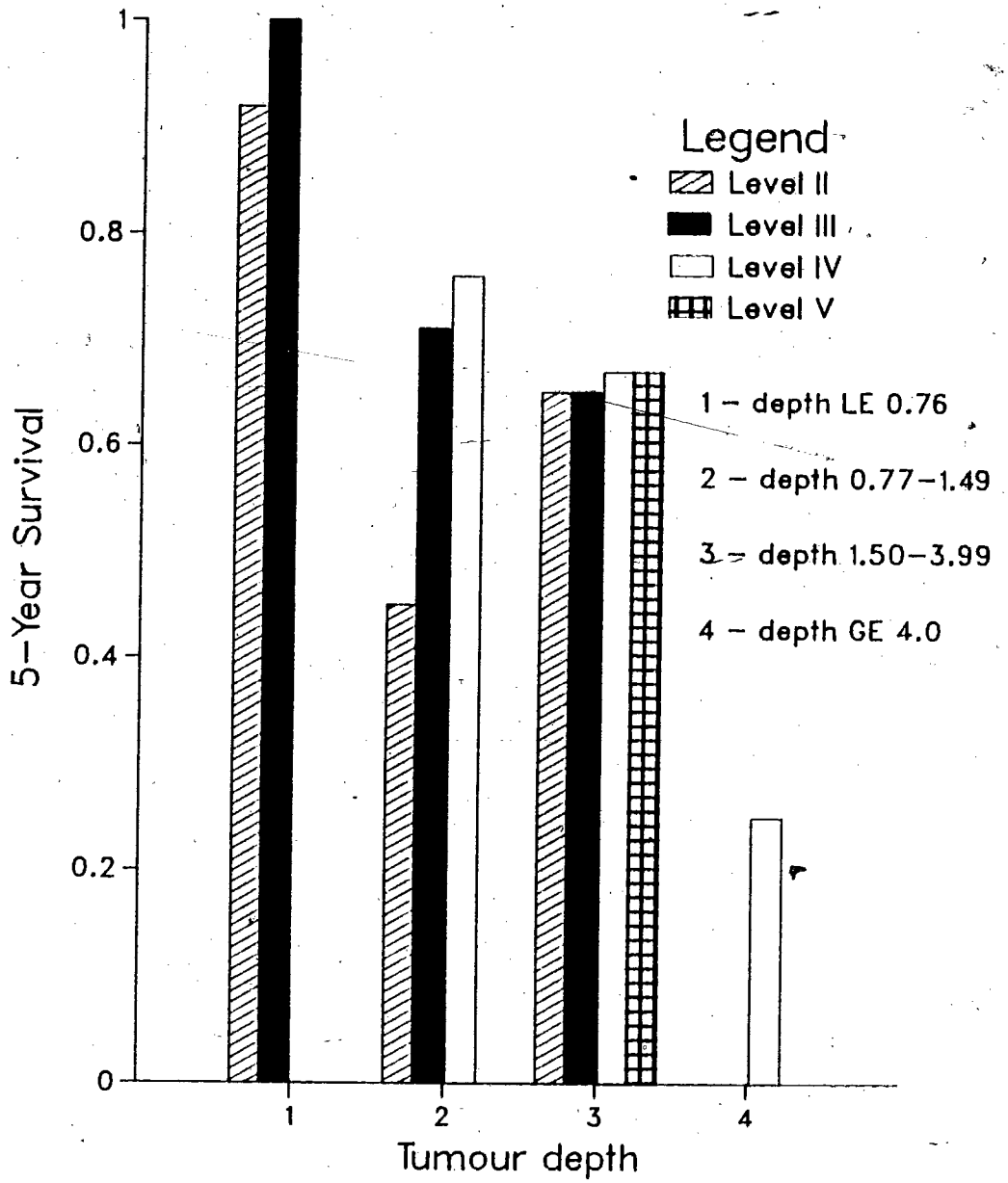


Figure 3.5.1 Comparison: Part B



As Balch *et al*. pointed out, within three Clark's levels there were gradations of depth that influenced survival (Part A), but converse relationships were not observed when the roles of depth and Clark's levels were exchanged (Part B).

2. Thirteen potential prognostic factors were studied first by univariate method (Breslow's test.) Six factors were significant at 0.05 level. A forward multiple logistic regression procedure was then used to analyze the thirteen factors simultaneously. Tumour depth was chosen to be one of the five significant factors, while levels of invasion belonged to the group of insignificant factors.
3. A comparison was made between the observed number of five year survivors and the number of predicted five year survivors based on the logistic model containing the five significant factors for evenly spaced survival probability intervals. A good fit was claimed.

A few comments on Balch *et al*.'s paper are given below. First, a good fit means a comprehensive agreement between data and the proposed model. Logistic models deal with "dead" or "alive" only, so the information contained in the length of survival time was not fully used, leaving possibility and need for further investigation. (Recent studies do use the length of survival time, i.e., Cox's regression model, but the ideas and the approach are the same as those of Balch *et al*.'s [28]).

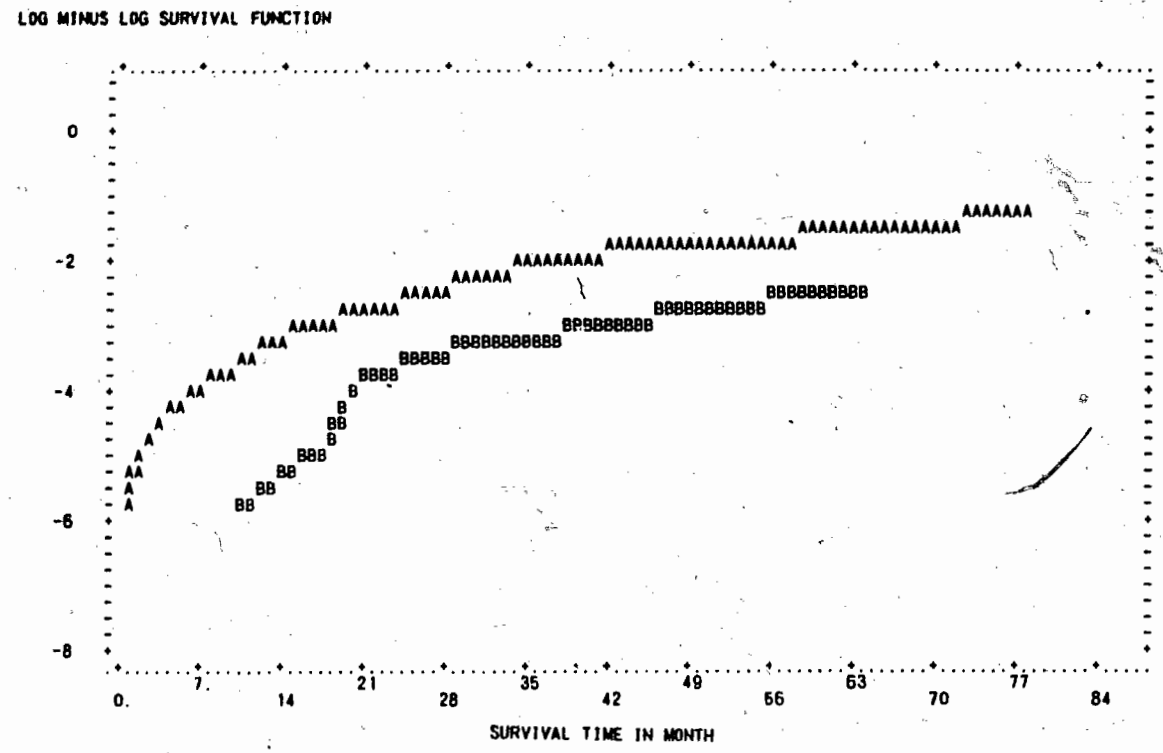
Secondly, levels of invasion are based on the structure of human integument and are essentially qualitative, while tumour depth is the direct measurement of maximum invasion regardless of skin structure and is basically quantitative. It is certainly worthwhile to look for gradations that influence the survival within each level of invasion, and for changes of survival accompanied with the change of levels of invasion within a certain chosen depth range. But the results thus obtained should be used with caution because the ranges of depth used are just a tiny portion of infinitely many possible ranges. In this sense the comparison done by Balch *et al*. was not fair nor complete.

Thirdly, as mentioned by the authors, interaction might exist and change the results of analysis. Therefore, the results of forward selection should at least be compared with those of backward selection, and when there are signs of interactions, one should persue in the indicated directions.

With the above review and analysis, the Cox model is fitted to the data and forward as well as backward selections of significant prognostic factors are performed. Still pigmentation and regression are not considered in the following multivariate analyses.

As before, the proportionality assumption is checked first. In Figure 3.5.2, the upper plots correspond to the original codes (Table 2.2); the lower plots correspond to the new codes in Table 3.4.1 with the exception that celltype is recoded as

Figure 3.5.2 Checking proportionality when depth and Clark's levels are used (sex: upper plot for original codes, lower plot for new codes).



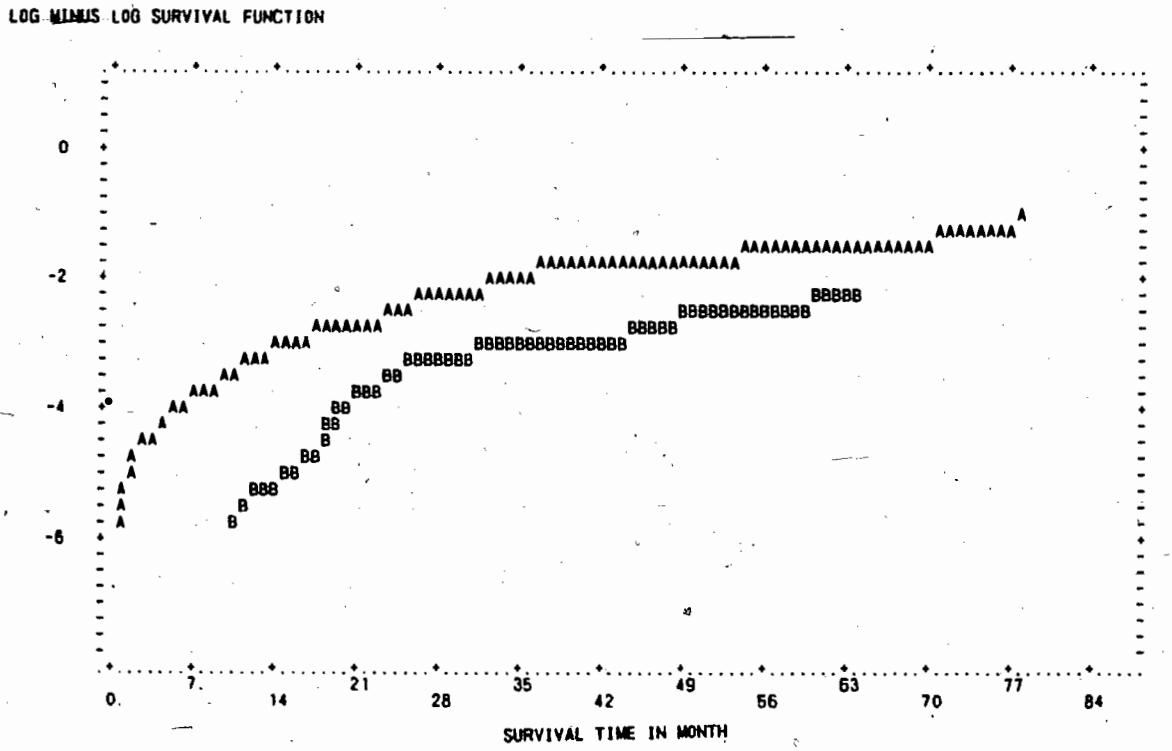
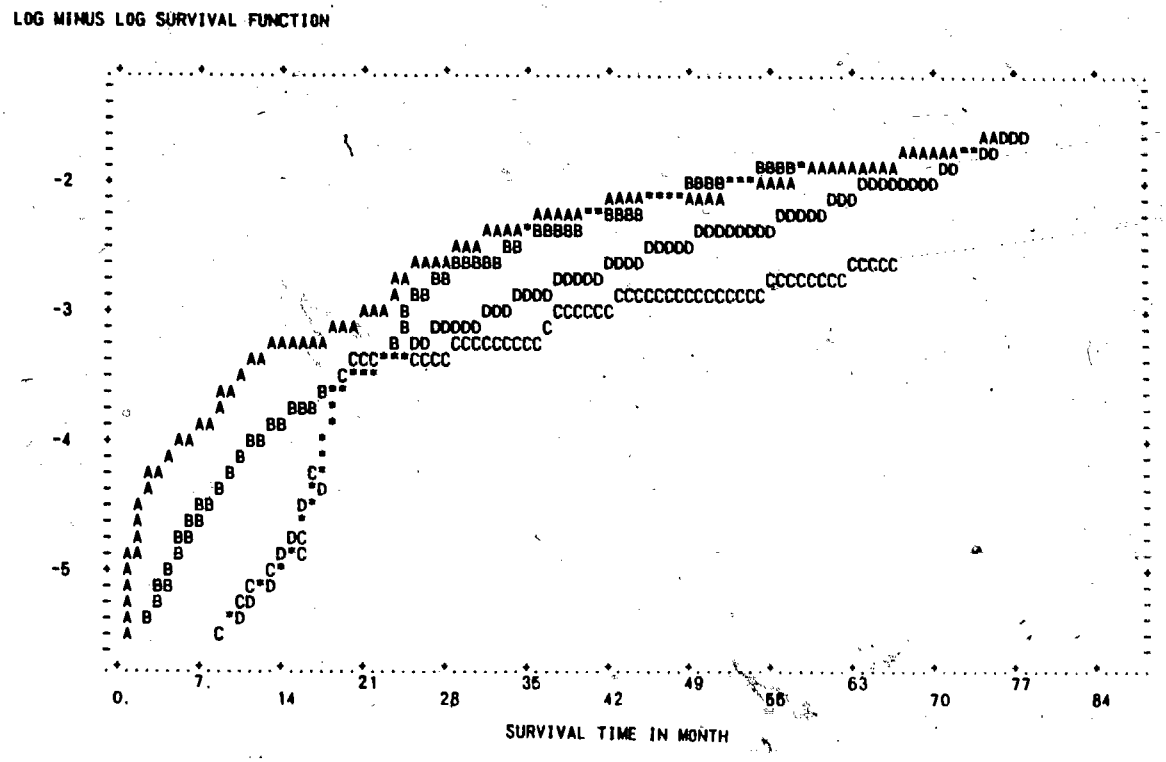


Figure 3.5.2 continued (site: upper plot for original codes, lower plot for new codes).



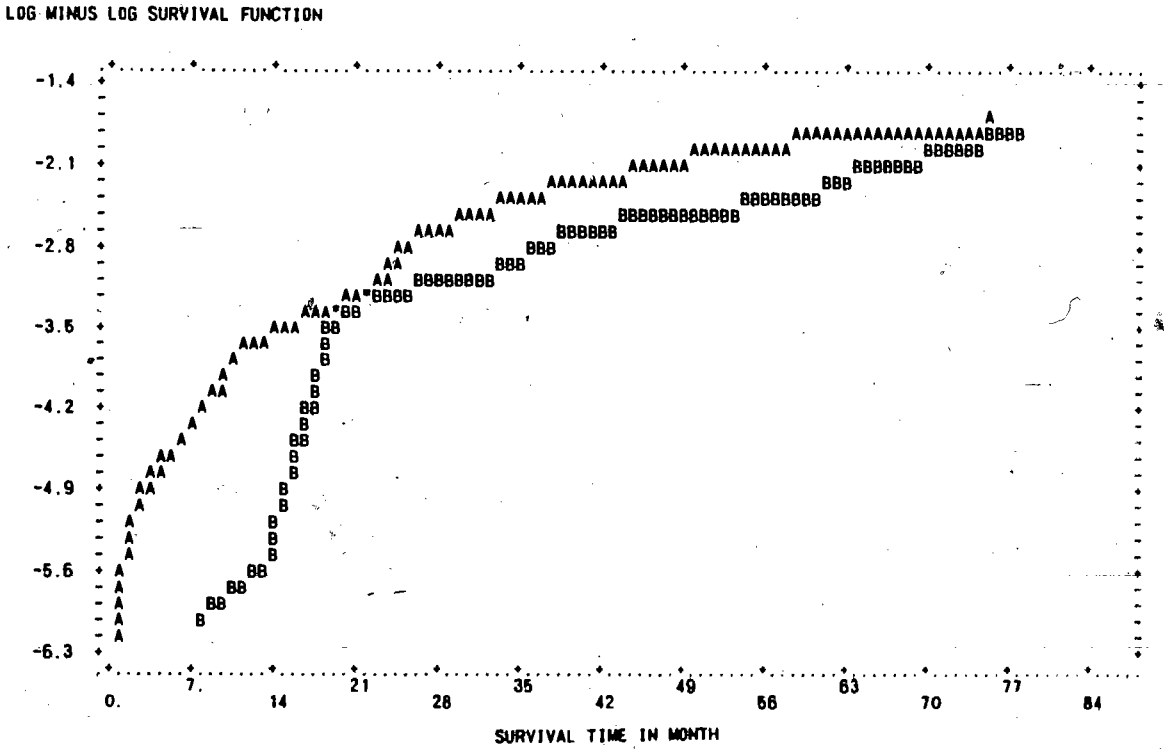
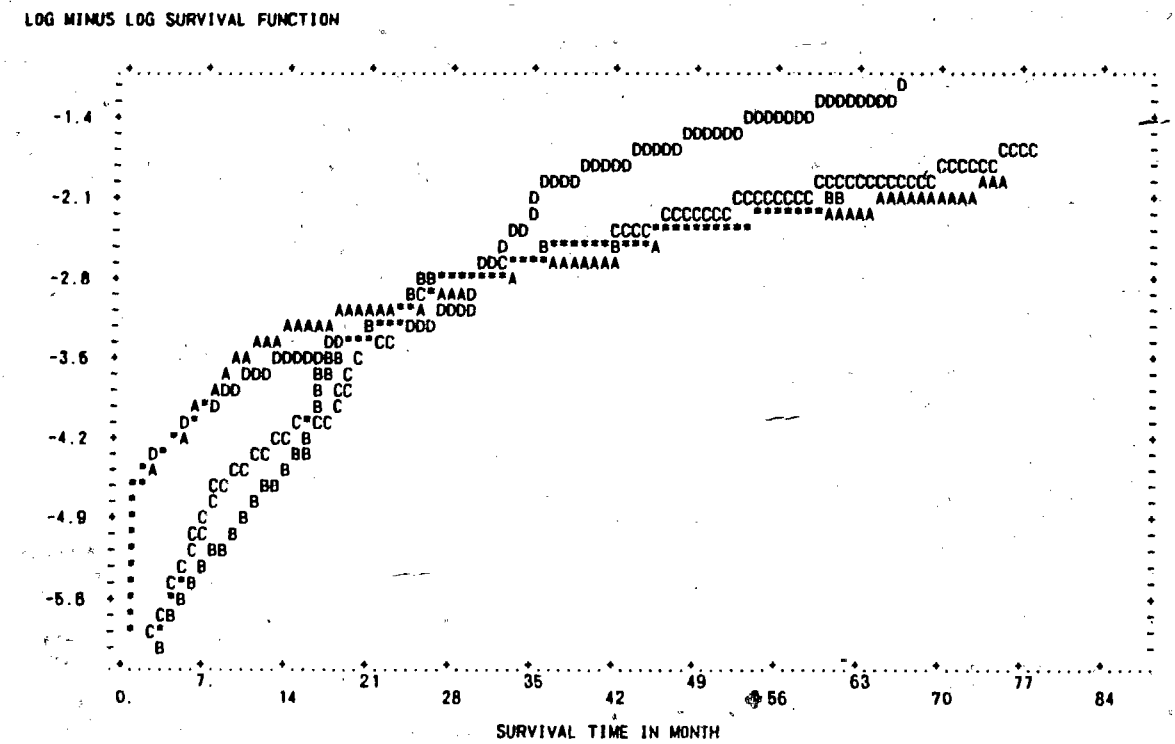


Figure 3.5.2 continued (Clark's: upper plot for original codes, lower plot for new codes).



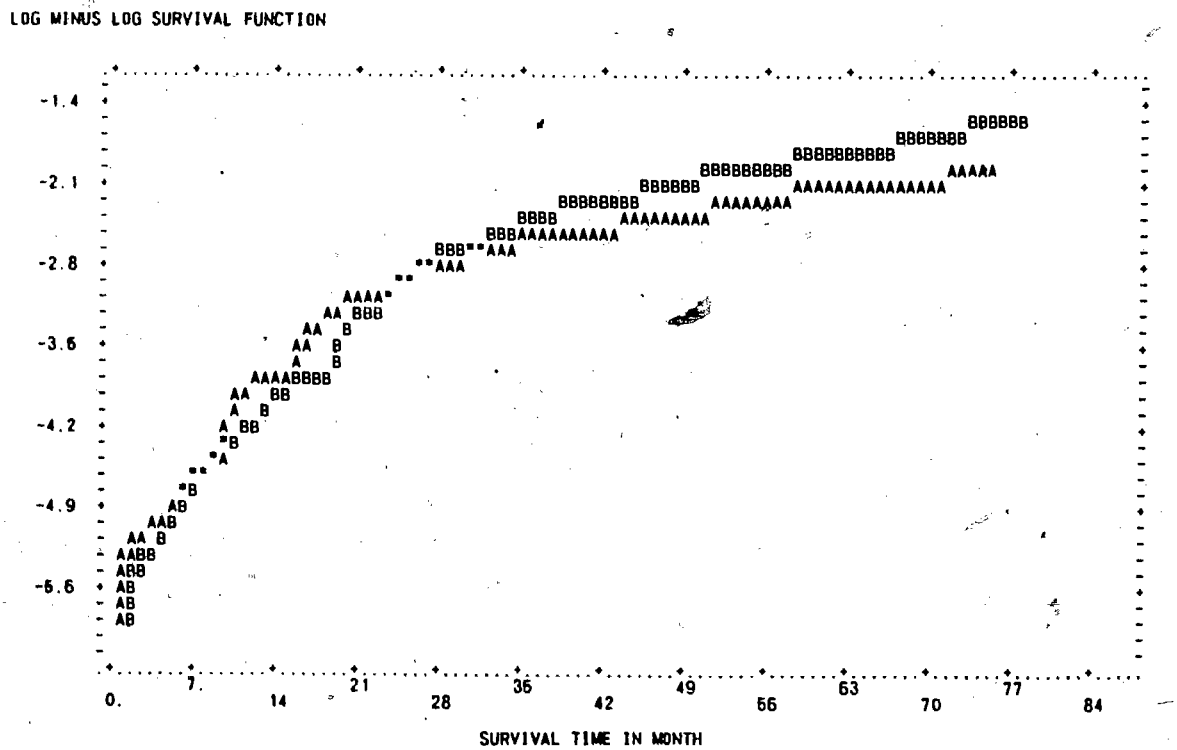
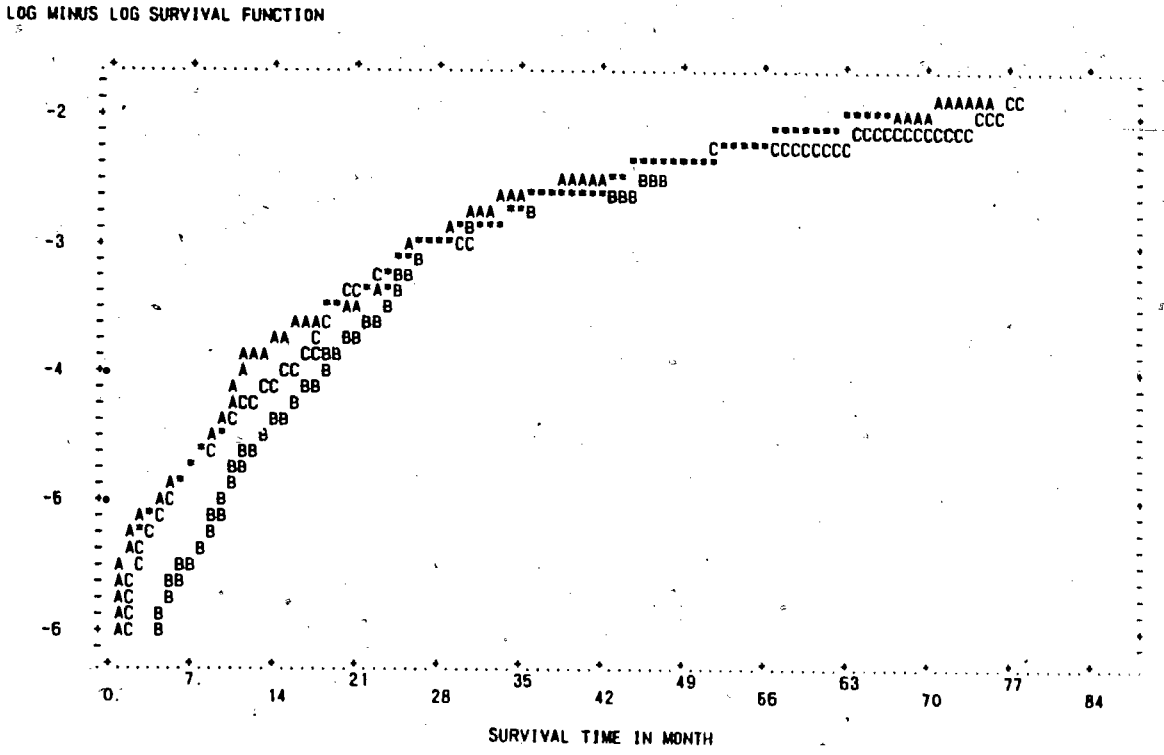


Figure 3.5.2 continued (mitoses: upper plot for original codes, lower plot for new codes).



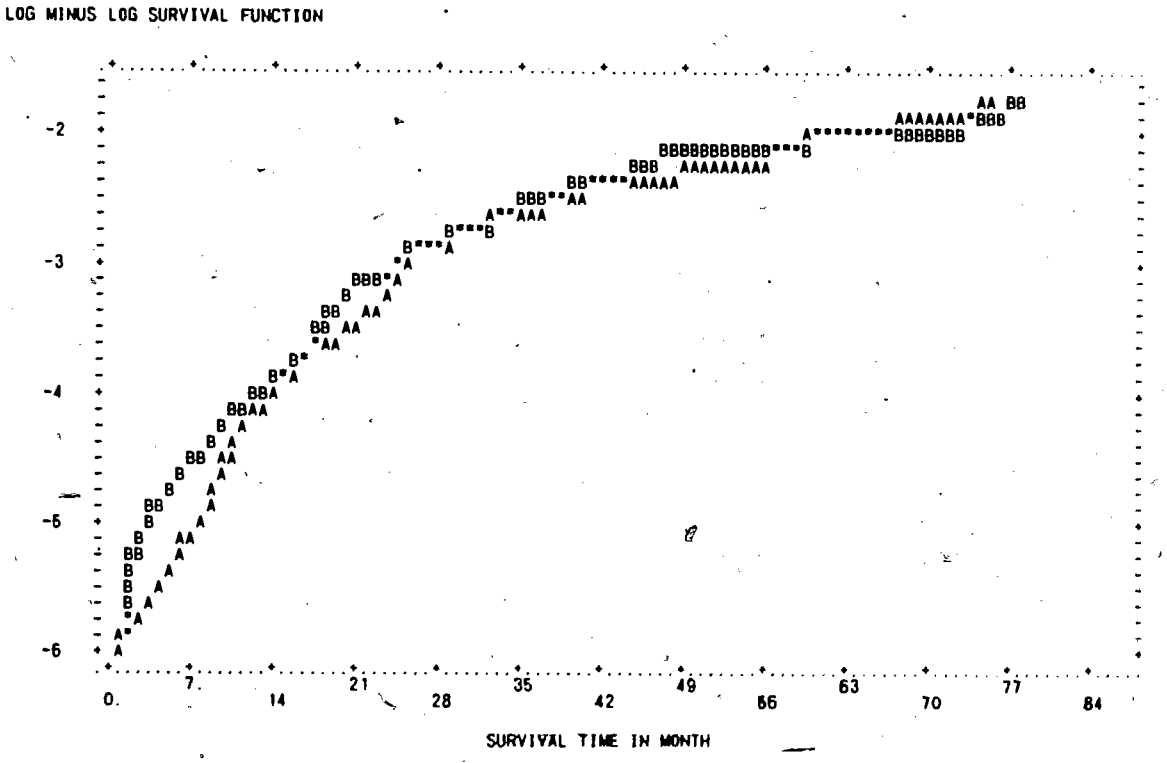
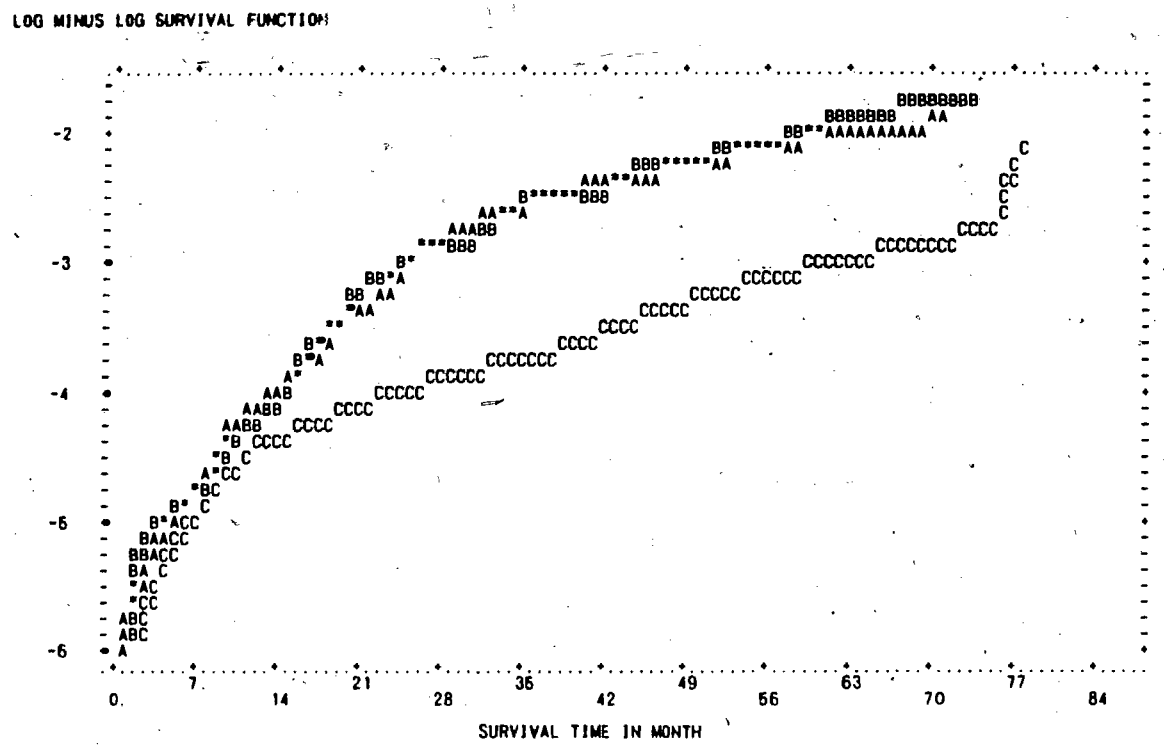


Figure 3.5.2 continued (celltype: upper plot for original codes, lower plot for new codes).



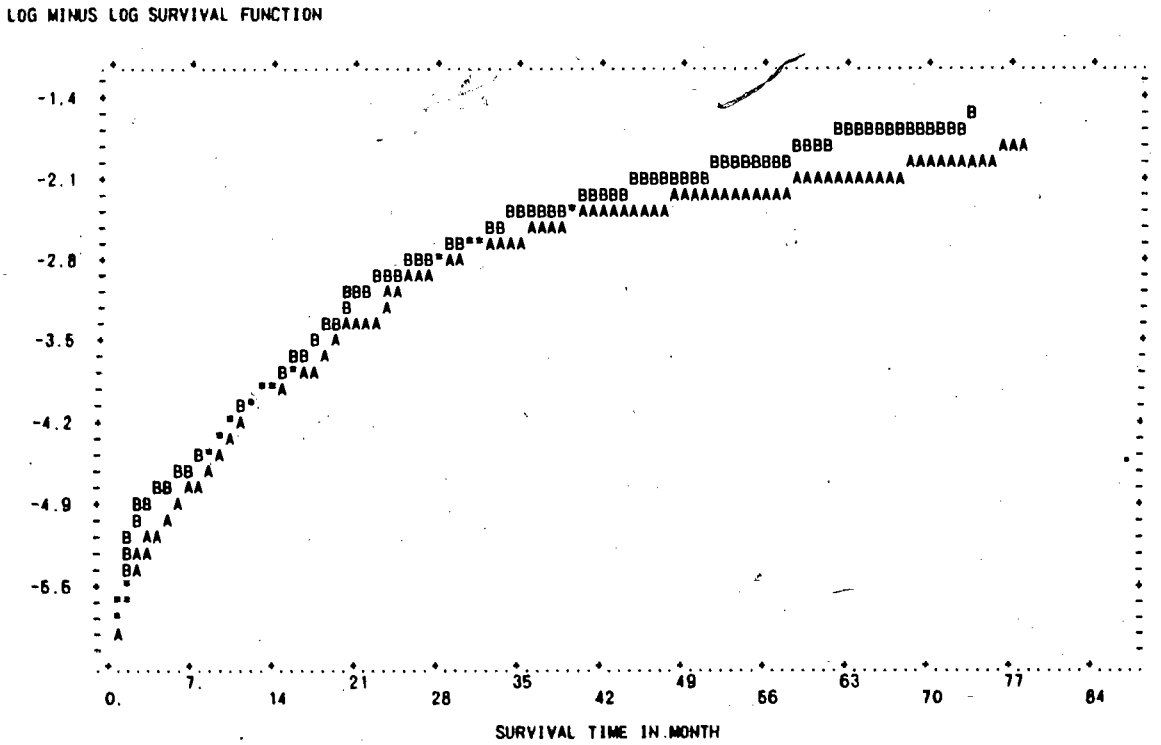
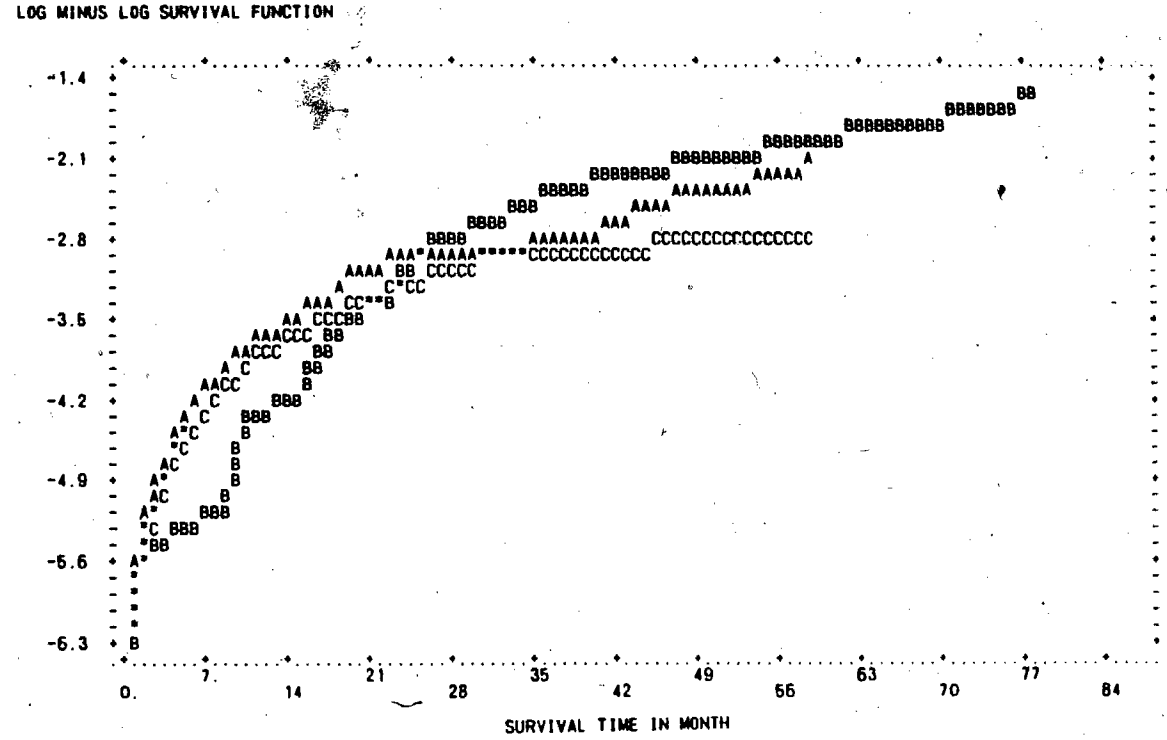


Figure 3.5.2 continued (differentition: upper plot for original codes, lower plot for new codes).



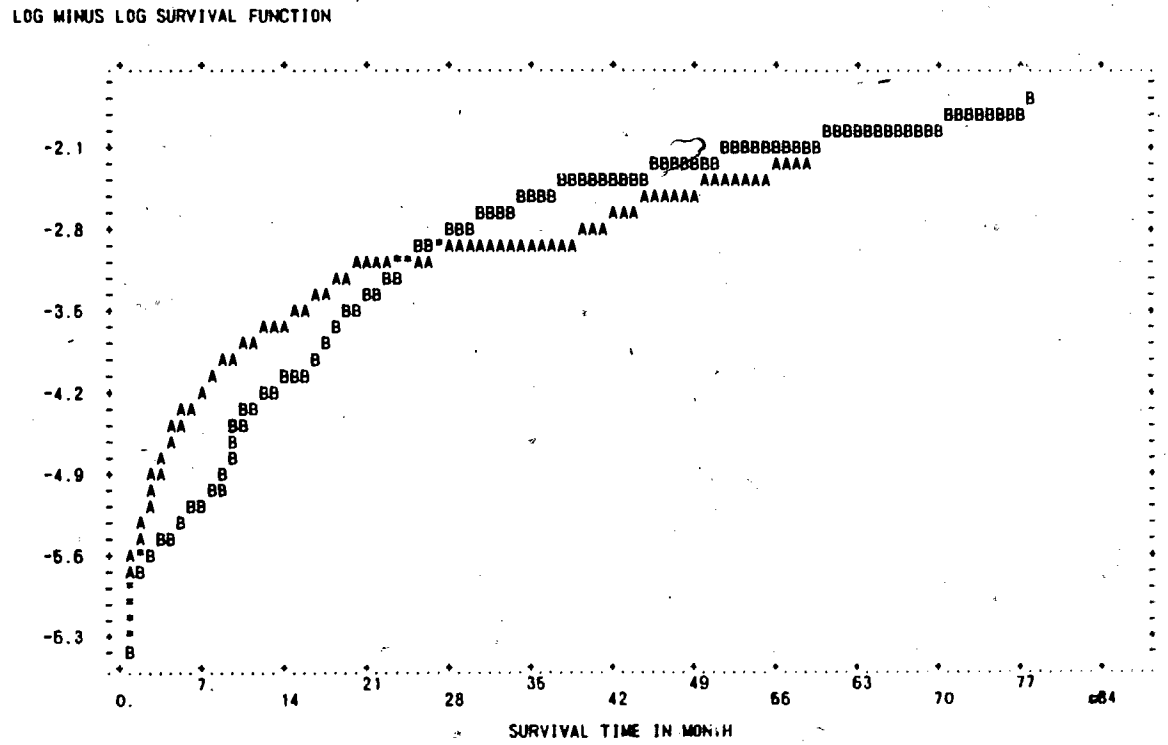
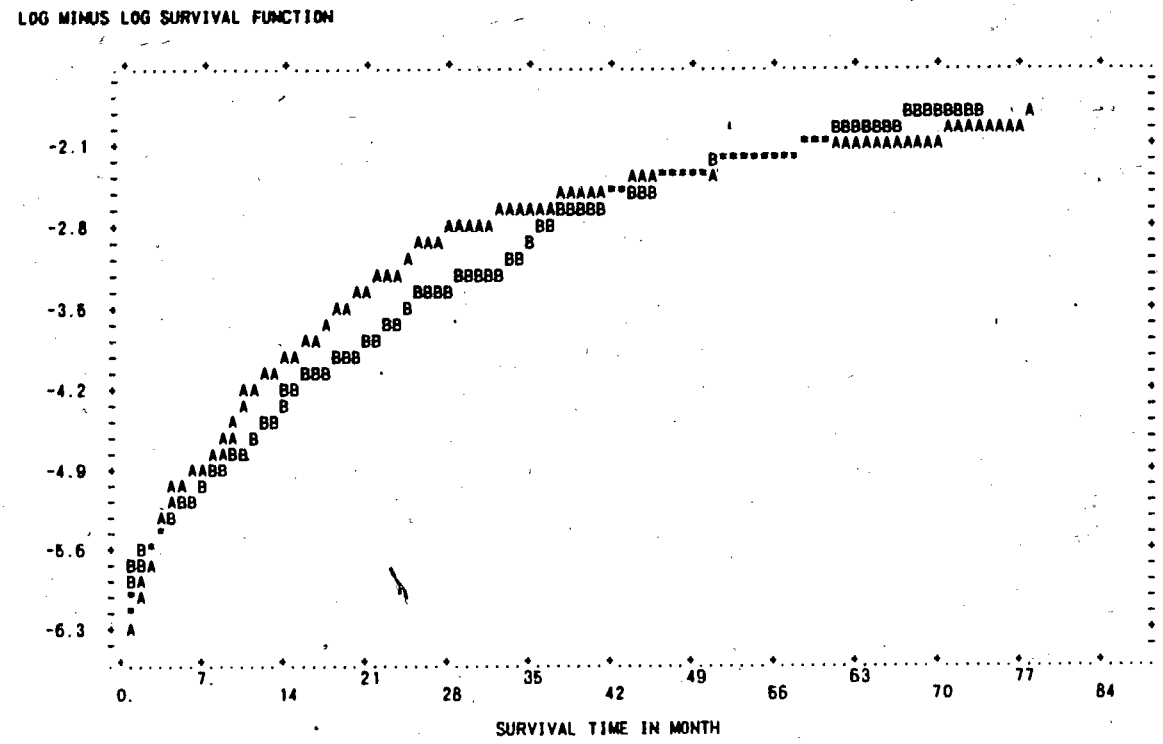


Figure 3.5.2 continued (ulceration: upper plot for original codes, lower plot for new codes).



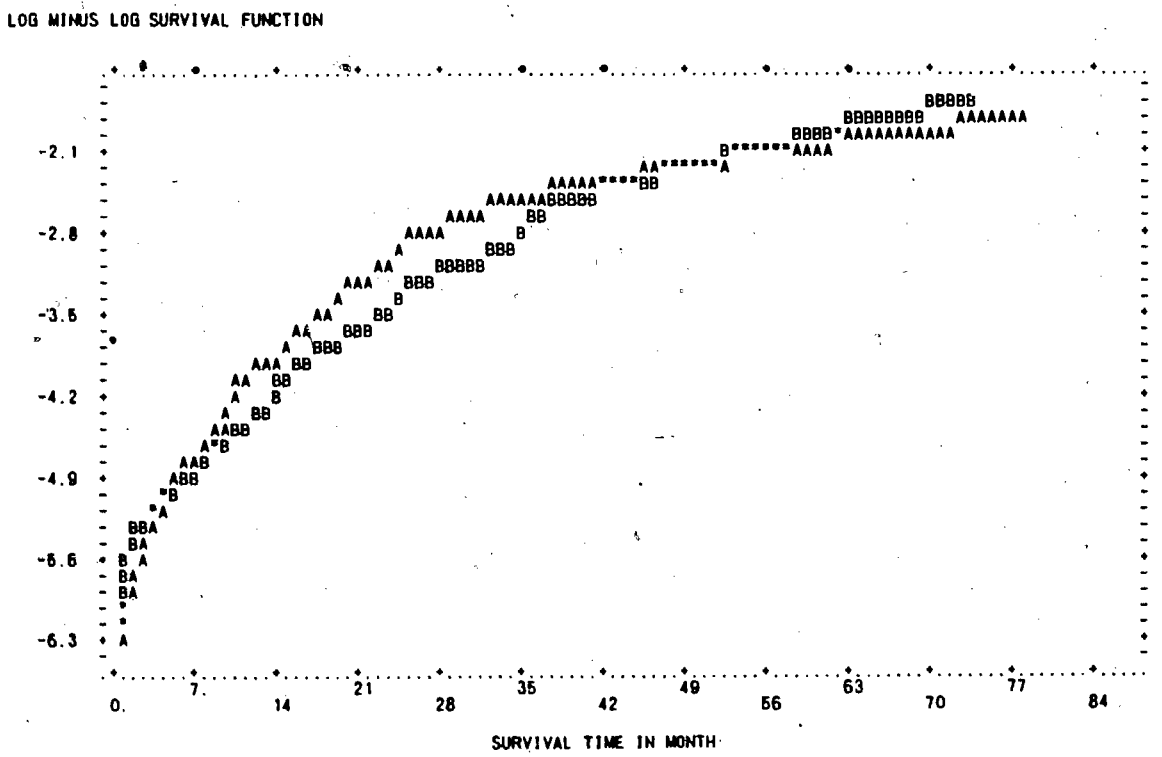
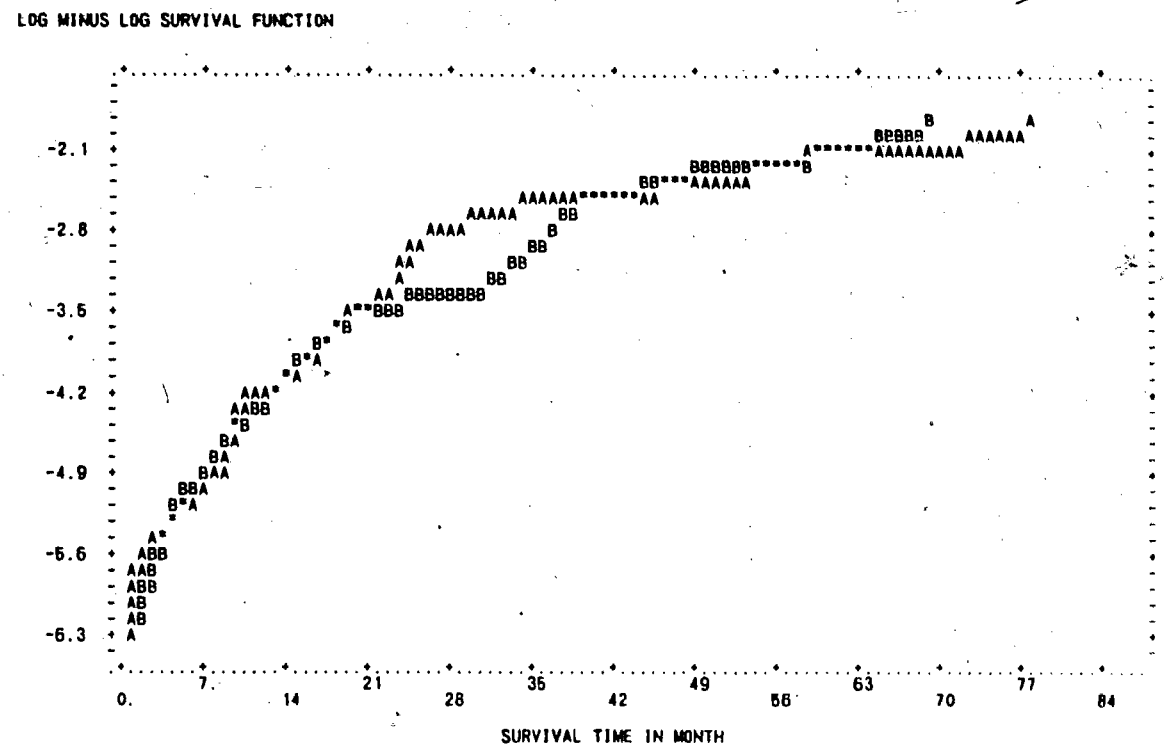
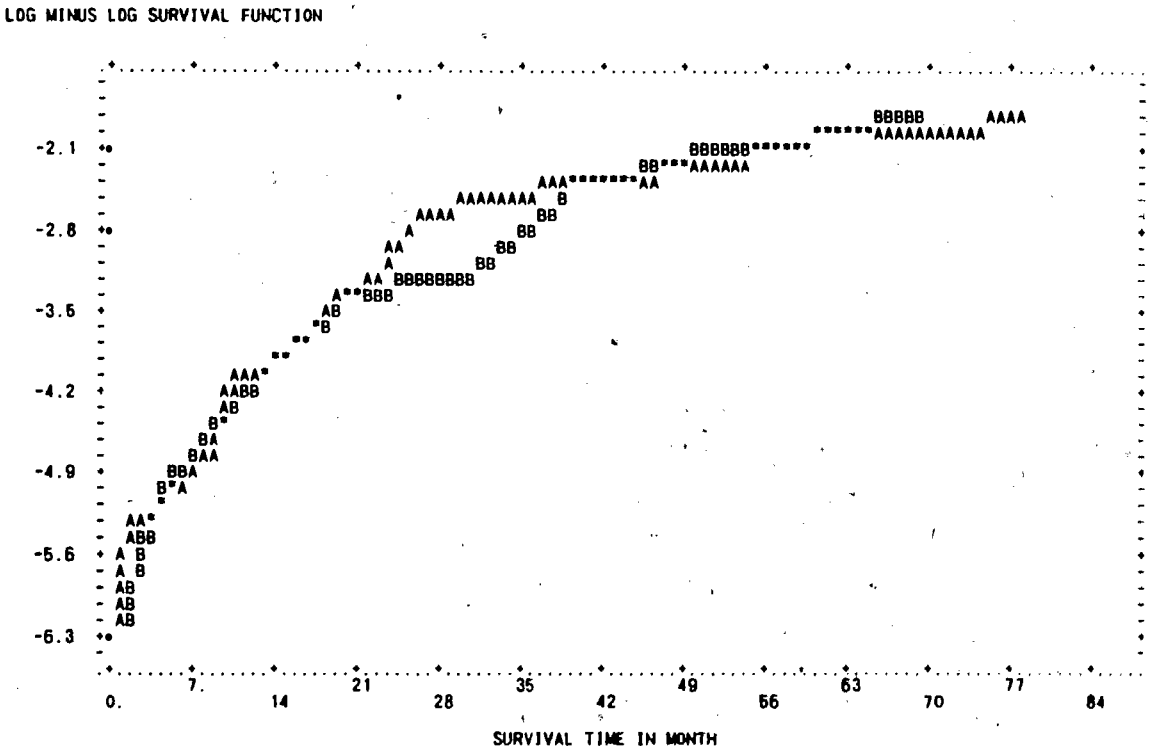


Figure 3.5.2 continued (lymphatic inflammation: upper plot for original codes, lower plot for new codes).





lentigo plus superficial spreading vs nodular. This change is suggested by the analyses done in section 3.4.3.

After recoding the data, the proportionality assumption appears to hold to a reasonable degree. Backward and forward selections are then carried out, giving results shown in Table 3.5.1 and Table 3.5.2.

Clark's levels of invasion is not selected in the above two covariate selections. Since the selected covariates are the same as those obtained in section 3.4.1, no further checking and testing need to be done and the superiority of tumour depth over Clark's levels of invasion is corroborated.

Table 3.5.1 Backward selection of significant prognostic factors.

Step	Covariates	Chi-square to			Global test
no.	left out	leave out	df	P-values	P-values
1	ulceration	0.000	1	0.9929	0.0000
2	mitoses	0.004	1	0.9449	0.0000
3	lymp	0.027	1	0.8685	0.0000
4	diff	0.357	1	0.5504	0.0000
5	celltype	0.617	1	0.4322	0.0
6	site	1.410	1	0.2349	0.0000
7	Clark's	2.498	1	0.1140	0.0
8	no covariate to be left out.				

The kept-in covariates are:

Name	P-values
depth	0.0
age	0.0007
sex	0.0000

The estimation of the kept-in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.
depth	0.3309	0.0502	6.5881
age	0.0231	0.0069	3.3360
sex(female)	-1.1655	0.2420	-4.8162

Table 3.5.2 Forward selection of significant prognostic factors.

Step no.	Covariates added in	Chi-square to add in	df	P-values	Global test P-values
1	depth	39.103	1	0.0	0.0
2	sex	21.813	1	0.0000	0.0
3	age	11.498	1	0.0007	0.0
4	no covariate to be added in.				

The estimation of the added in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.
depth	0.3309	0.0502	6.5881
age	0.0231	0.0069	3.3360
sex(female)	-1.1655	0.2420	-4.8162

3.5.2 New Hypothesis 2

As has been mentioned before, tumour depth is at present the best prognostic factor of stage I malignant melanoma. However, new prognostic factors are proposed from time to time. In 1978, Schmoeckel and Braun-Falco reported to have found an even better prognostic factor (they called it Prognostic Index, thus the name PI [55]). By definition, PI equals the product of tumour depth and the number of mitoses per square millimeter on standard histologic sections of the tumour.

However, if one reads Schmoeckel and Braun-Falco's paper, one can soon find out that the authors did not provide a convincing analysis to support their PI proposal. In a very recent paper by A.M. Kopf *et al.* (1987), the authors programmed their computer to seek significant cut points of the PI range and then used the cut points thus obtained to sub-divide evenly spaced depth ranges (1.50 to 2.49, 2.50 to 3.49, greater than 3.50) and showed that within each depth range, patients with different PI values (< 19 vs ≥ 19) had different survival. Also logistic regression was employed to show that cut point 19 was more significant than depth when both PI and depth were in the regression model.

For similar reasons to those given in section 3.4.1 about the comparison of Clark's levels of invasion with tumour depth, the comparison done by Kopf *et al.* can be improved. (By the way, Schmoeckel and Braun-Falco used the Student's t-test only). In this section, the prognostic value of PI is studied.

A word must be said about the definition of PI to be used below. When the data were collected, the information on the number of mitoses per square millimeter was not recorded. Instead, the actual number of mitoses were counted. Therefore the definition of PI to be used is

PI = (tumour depth) times(number of mitoses) = Index

For this reason, this section is named a new hypothesis.

As in section 3.4.1, tumour depth and Clark's levels are used to lead two parallel analyses, in each of which index (or PI) is included.

Graphical checks of the proportionality assumption produce plots quite similar to those displayed in section 3.4.1 and section 3.5.1 and thus are not presented. The proportionality assumption comes to hold in general after the data are recoded in the way shown in Table 3.4.1 except that celltype is recoded as lentigo plus superficial spreading vs nodular.

Four covariate selections are carried out. When depth is used, the forward and the backward selections select the same covariates, but index is not selected; when Clark's levels are used, the forward and the backward selections still select the same covariates, and index is selected and has smaller P-value than that of Clark's levels.

Table 3.5.3 Backward selection when depth is used.

Step	Covariates	Chi-square to	Global test

no.	left out	leave out	df	P-values	P-values
1	lymp	0.000	1	0.9964	0.0000
2	ulceration	0.024	1	0.8761	0.0000
3	diff	0.207	1	0.6495	0.0000
4	celltype	0.786	1	0.3754	0.0000
5	site	1.504	1	0.2200	0.0000
6	mitoses	1.831	1	0.1760	0.0000
7	index	1.308	1	0.2528	0.0
8	no covariate to be left out.				

The kept-in covariates are:

Name	P-values
depth	0.0
age	0.0007
sex	0.0000

The estimation of the kept-in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.
depth	0.3309	0.0502	6.5881
age	0.0231	0.0069	3.3360
sex(female)	-1.1655	0.2420	-4.8162

Table 3.5.4 Forward selection when depth is used.

Step no.	Covariates added in	Chi-square to add in	df	P-values	Global test P-values
1	depth	39.103	1	0.0	0.0
2	sex	21.813	1	0.0000	0.0

3	age	11.498	1	0.0007	0.0
4	no covariate to be added in.				

The estimation of the added-in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.
depth	0.3309	0.0502	6.5881
age	0.0231	0.0069	3.3360
sex(female)	-1.1655	0.2420	-4.8162

Table 3.5.5 Backward selection when Clark's levels are used.

Step no.	Covariates left out	Chi-square to leave out	df	P-values	Global test P-values
1	ulceration	0.000	1	0.9964	0.0000
2	lymp	0.024	1	0.8761	0.0000
3	diff	0.207	1	0.6495	0.0000
4	mitoses	0.786	1	0.3754	0.0000
5	celltype	1.504	1	0.2200	0.0000
6	site	1.831	1	0.1760	0.0000
7	no covariate to be left out.				

The kept-in covariates are:

Name	P-values
Clark's	0.0322
age	0.0024
sex	0.0000
index	0.0001

The estimation of the kept-in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.
Clark's (IV,V)	0.5606	0.2644	2.1201
age	0.0210	0.0070	2.9981
sex (female)	-1.0994	0.2400	-4.5812
index	0.0803	0.0185	4.3519

Table 3.5.6 Forward selection when Clark's levels are used.

Step no.	Covariates added in	Chi-square to add in	df	P-values	Global test P-values
1	index	32.595	1	0.0	0.0
2	sex	20.840	1	0.0000	0.0
3	age	11.512	1	0.0007	0.0
4	Clark's	4.585	1	0.0322	0.0000
5	no covariate to be added in.				

The estimation of the added-in covariates is:

Covariates	Coefficients	Standard Error	Coeff./s.e.
Clark's (IV,V)	0.5606	0.2644	2.1201
age	0.0210	0.0070	2.9981
sex (female)	-1.0994	0.2400	-4.5812
index	0.0803	0.0185	4.3519

The above results suggest that index stands in between depth and Clark's levels. Considering the fact that Clark's level is not selected when it is analyzed together with depth, and that there is no natural meaning attached to index, index is not

recommended as a significant prognostic factor.

In fact, since the absolute counts of mitoses are used, one may argue that some adjustment is needed to get rid of the effect of presumed evenly spaced ranges of prognosis, which is unlikely to be true because depth seems to measure the effect of duration and cumulated invasions, thus is a long term effect, while mitosis measures growth rate of the tumour and is more dynamic. The use of *the number of mitoses per square millimeter* seems to be a step in the right direction.

3.5.3 New Hypothesis 3

During the course of analyzing the data, some signs of interactions among the covariates were observed. In this section, an effort is made to look for interactions. Two methods are supplied and one of the two methods is demonstrated using the data.

The first method is most suitable for searching interactions among covariates. Suppose a covariate has k levels (for a continuous covariate, it is discretized into k levels). A separate model is fitted to each level and a covariate selection procedure is carried out. If different covariates are selected as being significant for different levels, the implication is that an interaction may exist between the chosen covariate and the other covariates. ([31]).

The second method is formal and is based on the likelihood ratio test. It is suitable for both searching and confirming

existence of interactions. As above, for a chosen covariate with $k+1$ levels, let $\alpha_1, \alpha_2, \cdots, \alpha_k$ be the k corresponding effects on survival. For each level i, let $\beta_i = (\beta_{i1} \cdots, \beta_{is})^T$ be the regression coefficients associated with the covariates $z_1, z_2, \cdots, z_s$, $i = 1, 2, \cdots, k$. Then the following null hypothesis

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k$$

means that there is no interaction between the chosen covariate and other covariates.

Denote by $L_0(k+s)$ the maximized partial likelihood under H_0 and by $L((k+1)s)$ the maximized partial likelihood under no restrictions. If H_0 holds, then

$$X^2 = -2 \ln \left(\frac{L_0(k+s)}{L((k+1)s)}\right)$$

will distributed asymptoticlly as chi-square with $k(s-1)$ degrees of freedom ([31]).

Since BMDP does not support the second method easily, in the following demonstration, only the first method is used.

This section is mostly an exlporatory section, especially it is hoped to find (significant) interaction(s) among the insignificant factors according to the previous analyses. In order to search as far as possible, pigmentation and regression are also considered, but of depth and Clark's levels, only depth is used.

To ease the application of the first method, the data are recoded. For technical reasons, some factors can not be included into the model desired to be fitted ([10]). This is made clear

when the results are presented. The backward selection procedure is used to obtain clues of interactions.

Table 3.5.7 Recoding information for finding interactions.

Covariate	Levels	Codes	Names	Base-level
sex	male	1	sex1	sex1
	female	2	sex2	
site	non-limbs	2	site1	site1
	limbs	4	site4	
depth	continuous	/	depth	/
mito	level I and II	1	mito1	mito1
	level III	3	mito3	
cell	lentigo and superficial spreading	1	cell1	cell1
	nodular	3	cell3	
diff	level I	1	diff1	
	level II and III	3	diff3	diff3
pigm	level 0 and I	1	pigm1	pigm1
	level II and III	3	pigm3	
ulce	absent	0	ulce0	ulce0
	present	1	ulce1	
lymp	absent	0	lymp0	lymp0
	present	1	lymp1	
regr	absent	0	regr0	regr0
	present	1	regr1	
age	continuous	/	age	/

Table 3.5.8 Backward selections for clues of interaction.

Factors selected	Under Male	Factors not in Model
depth age diff3		none
	Under female	
depth age		none
	Under site1	
depth sex2 age pigm3		diff3
	Under site3	
depth		diff3
	Under mito1	
depth sex2 age		diff3
	Under mito3	
depth sex2 age pigm3 regr1		diff3
	Under cell1	
depth sex2 age		diff3 Site4 mito3 pigm3
	Under cell3	
Sex2 age		diff3 Site4 mito3 pigm3
	Under diff1	
depth sex2 age pigm3		ulce1 mito3 regr1
	Under diff3	
depth		ulce1 mito3 regr1

	Under pigm1	
depth sex2 age		diff3
	Under pigm3	
depth sex2		diff
	Under ulce0	
depth sex2 age		none
diff3 pigm3		
	Under ulce1	
lymp1 sex2 age		none
	Under lymp0	
depth sex2 age		none
pigm3		
	Under lymp1	
depth sex2		none
	Under regr0	
depth sex2 age		diff3
	Under regr1	
depth sex2 age		diff3
mito3		

Based on the clues of Table 3.5.8, formal statistical tests are performed by introducing cross-product terms. The factors always in the model when doing any test are depth, sex and age. In every single test to be performed, only one cross-product term is tested. Table 3.5.9 has the summary for those cross-product terms in which at least one element is a

significant factor for the whole data set.

Table 3.5.9 Testing cross-product terms I.

Cross-products tested	P-values	
sex2 by diff1	0.3649	
site3 by sex2	0.5946	(P-values are given by
mito3 by sex2	0.4679	the likelihood ratio
pigm3 by sex2	0.4309	test)
ulce1 by sex2	0.7485	
lymp1 by sex2	0.2633	

None of the cross-product terms are significant.

The cross-product terms in which none of the elements are significant factors are tested below. Note that "0" means not significant, "1" means significant. Significance level is α = 0.05.

One significant cross-product term is found, namely, mitoses by regression. Is it legitimate to report this cross-product term as a significant factor ? Should the significance level be controlled because of what people call the "multiple comparison" effect ?

These two questions are connected. The key point is: What does it mean by being significant ? Twenty-seven tests at α = 0.05 level are carried out, then the probability of having at

Table 3.5.10 Testing cross-product terms II

	site	mito	cell	pigm	ulce	lymp	regr
site							
mito	0						
cell	0	0					
pigm	0	0	0				
ulce	0	0	0	0			
lymp	0	0	0	0	0		
regr	0	1	0	0	0	0	

least one significant result is $1-(1-0.05)^{27} = 0.7497$. This is a high probability. If one controls the above probability at 0.05, α should be set to 0.0019 at the very beginning, then the term mitoses by regression is no longer significant (it has P-value 0.0116.)

Something subtle and philosophical is involved here. However, this section is an exploratory section, so it is recommended to investigate the interaction between mitoses and regression further.

On the whole, looking for significant interactions is not the easiest job in the world. Based on the data, only some signs are seen.

3.5.4 A Goodness-of-Fit Problem

In section 3.3, the statistical background of the Cox regression model was introduced. In particular, it was mentioned how to check the proportionality assumption by plotting or by testing time-dependent covariates. As to the problem of model adequacy, a global chi-square test was mentioned and in section 3.4.1 a brief discussion on some practical problems was offered.

In this section, the problem of checking model adequacy is studied in a little more detail. Instead of describing a solution to the problem, which has not been available practically, a few considerations are given in looking for a solution to this problem..

In the field of survival analysis of biological data, various kinds of randomness exist. But precise information necessary to describe them is usually hard to obtain. A typical problem is: a large number of variables were measured, but they were measured in greatly varing precisions, due to arbitrary coding, complicated censoring, competing risks and unevenly correlated associations. When the data are ready, one usually has an enormous amount of information, but perhaps only a small portion is useful in answering the questions in one's mind. Sometimes one even does not know what portion one needs to take for uses. Intuitively speaking, analysis methods requiring high and uniform precision of data are neither practical nor suitable.

At present, graphical methods are the most commonly used methods. There are some advantages in using graphical techniques, such as their being quick, easily understood and fairly comprehensive. However, when bordline or near bordline cases arise, graphical methods seem to be incompetent. Moreover, all the methods available have "presumed", "approximately" and "may be expected to" etc. in their descriptions. For instance, let $h(t,\underline{z})$ be a hazard rate, $H(t,\underline{z})$ be the cumulative hazard function defined by

$$H(t,\underline{z}) = \int_0^t h(u,\underline{z})du.$$

Then the transformation

$$E_i = H(T_i,\underline{z}) \quad i = 1, 2, \cdots, n$$

will transform survival times $\{T_i\}$ to a (censored) sample from the unit exponential distribution <u>if one knows the underlying theoretical $H(t,\underline{z})$.</u> However, one usually does not know and can only estimate $H(t,\underline{z})$ by, say, $\hat{H}(t,\underline{z})$. When this estimate is used to calculate

$$e_i = \hat{H}(t_i,\underline{z}),$$

$\{e_i\}$ will have their own statistical properties which one does not know exactly. The hope, which is the basis of some graphical methods, is that $\{e_i\}$ will behave <u>somehow</u> like a genuine sample from the unit exponential distribution. But at present not much is known about the impact of this assumption on the results thus obtained.

Note that the above question is a goodness-of-fit question, but it is not in the usual framework. Instead of estimating a

few parameters, one first estimates the whole cumulative hazard function $H(t,\underline{z})$ and the uses this estimate to obtain what people nowdays call the generalized residuals $\{e_i\}$ ([17]).

In the case of fitting the Cox model, one does not specify $h(t,\underline{z})$ (thus $H(t,\underline{z})$) at all, so there even does not exist an object to make comparison to. An approach has been suggested to treat the base line hazard rate function $h_0(t)$ in the Cox model as constant or piecewise constant ([1],[57]). Perhaps a method with certain upper and lower bounds when estimating $h_0(t)$ might be more appropriate for the reason similar to that of providing a confidence interval for a point estimate.

It seems to be unusual for a project to bring about a problem without solving it, at least partially. However, in the present case, solving the problem itself would be another project or even a thesis. Nevertheless, there is another purpose to discuss the above problem here, namely, to make the following point clear from a practical point of view: care should be taken when fitting the Cox survival regression model because powerful techniques for checking model adequacy are not practically available.

3.5.5 Prediction of Five-year Survival

One purpose of fitting a regression model is to do prediction. In this section, the model containing tumour depth, sex, age and cell type (LMM vs non-LMM) determined in section 3.4.2 is utilized to predict five-year survival of stage I

malignant melanoma patients covered by the data.

Predictions are based on the following formula. According to section 3.3,

$$h(t;\underline{z}) = h_0(t)\exp(\beta^T\underline{z}).$$

Thus,

$$S(t;\underline{z}) = \exp(-\int_0^t h(u;\underline{z})du)$$

$$= [S_0(t)]^{\exp(\beta^T\underline{z})},$$

where $S_0(t)$ is the base line survival function corresponding to $\underline{z} = 0$ ([40],[46]).

In the present context, $\underline{z} = (z_1,z_2,z_3,z_4)^T$, where z_1 = depth, z_2 = sex (= 0 if male, = 1 if female), z_3 = age and z_4 = celltype (=0 if non-LMM, = 1 if LMM). The depth range is divided into four intervals: LE 0.74 mm, 0.75-1.49 mm, 1.50-3.49 mm, GE 3.50 mm. In each interval, the mean depth is used when doing prediction. Similarly, age is divided into under 64 and over 65 years old and the mean age in each interval is used to do prediction. S_0(five years) = 83.04% is estimated by the product-limit method.

All together, 32 predictions of 5-year survival are made. The following table contains these predictions.

Table 3.5.11 Predictions of five-year survival.

Depth	Sex	Age	Celltype	Five-year survival
LE 0.74	male	under 64	LMM	96.0 %

LE 0.74	male	under 64	non-LMM	89.9 %
LE 0.74	male	over 65	LMM	92.1 %
LE 0.74	male	over 65	non-LMM	80.5 %

Depth	Sex	Age	Celltype	Five-year survival
LE 0.74	female	under 64	LMM	98.9 %
LE 0.74	female	under 64	non-LMM	97.0 %
LE 0.74	female	over 65	LMM	97.7 %
LE 0.74	female	over 65	non-LMM	93.9 %

Depth	Sex	Age	Celltype	Five-year survival
0.75-1.49	male	under 64	LMM	95.1 %
0.75-1.49	male	under 64	non-LMM	87.6 %
0.75-1.49	male	over 65	LMM	89.5 %
0.75-1.49	male	over 65	non-LMM	74.5 %

Depth	Sex	Age	Celltype	Five-year survival
0.75-1.49	female	under 64	LMM	98.6 %
0.75-1.49	female	under 64	non-LMM	96.3 %
0.75-1.49	female	over 65	LMM	97.1 %
0.75-1.49	female	over 65	non-LMM	92.4 %

Depth	Sex	Age	Celltype	Five-year survival
1.50-3.49	male	under 64	LMM	92.7 %
1.50-3.49	male	under 64	non-LMM	81.9 %
1.50-3.49	male	over 65	LMM	83.8 %
1.50-3.49	male	over 65	non-LMM	62.7 %

Depth	Sex	Age	Celltype	Five-year survival
1.50-3.49	female	under 64	LMM	97.9 %
1.50-3.49	female	under 64	non-LMM	94.6 %
1.50-3.49	female	over 65	LMM	94.7 %
1.50-3.49	female	over 65	non-LMM	86.6 %

Depth	Sex	Age	Celltype	Five-year survival
GE 3.50	male	under 64	LMM	85.3 %
GE 3.50	male	under 64	non-LMM	65.7 %
GE 3.50	male	over 65	LMM	67.2 %
GE 3.50	male	over 65	non-LMM	35.0 %

Depth	Sex	Age	Celltype	Five-year survival
GE 3.50	female	under 64	LMM	94.1 %
GE 3.50	female	under 64	non-LMM	85.0 %
GE 3.50	female	over 65	LMM	86.3 %
GE 3.50	female	over 65	non-LMM	67.8 %

The worst situation involves a male patient over 65 years old and with non-LMM lesion more than 3.50 mm thick; the best situation happens to a female patient under 64 years old and with LMM lesion less than 0.74 mm in depth.

CHAPTER IV

CONCLUSIONS

Towards the main objective of identifying significant prognostic factors of stage I malignant melanoma, both univariate analysis (Mantel's test and Breslow's test, etc.) and multivariate analysis (Cox's proportional hazards regression model) were carried out to analyze the data set from the Western Canada Melanoma Study. Among the thirteen potential prognostic factors covered by the data, melanoma tumour depth, sex and age of the melanoma patients were found to be important prognostic factors, that is, the deeper the tumour depth or the older the patients, the worse the prognosis; the female patients generally have better prognosis than the male patients.

Two current research problems were also investigated. In the first problem, some researchers claim that there is no difference in prognosis between stage I lentigo melanoma (LMM) and stage I superficial spreading and nodular melanomas (non-LMM) after controlling for tumour depth and location ([38]). The analyses of the data in this project show that LMM has a better prognosis than non-LMM, a conclusion which has been reported in [12] and [54]. In the second problem, the BANS concept (that is, tumours located on the upper Back, posterior Arms, posterior Neck and posterior Scalp have worse prognosis than tumours located elsewhere) was shown to be invalid based on the data, adding one more negative case study to [13], [55] and

[62].

The review of the development of two important prognostic factors of stage I melanoma, namely Breslow's tumour depth and Clark's levels of tumour invasion, reveals that the former is a quantitative measurement, the latter is a qualitative measurement, of the same characteristic of melanoma tumours, and therefore the two factors are highly correlated. For the study of stage I melanoma, the present fashion is to analyze tumour depth and Clark's levels of invasion in one model, which often leads to such conclusions as that both tumour depth and Clark's levels are barely significant and either sex or age is much more significant ([28]). From a statistical analysis point of view, including highly correlated covariates into one model is not recommended in general; from a biological interpretation point of view, it is more appropriate to analyze tumour depth and Clark's levels of invasion separately. Using this strategy, Clark's levels of invasion, sex, age and melanoma cell types are found to be another set of significant prognostic factors.

Based on the data, a new prognostic index, i.e., the product of tumour depth and the number of mitoses, is studied. This index is shown to be not as powerful as tumour depth alone in predicting prognosis.

A systematic search for interactions between the potential prognostic factors is performed. Mitoses seems to interact on regression. This interaction is recommended for further

investigation.

For doctors to use the results of the analyses, a series of predictions of five-year survival of the melanoma patients are supplied. The method of predicting any number of years of survival is also explained and the formula provided.

There are several areas where further analyses can be of interest.

Throughout this project nonparametric techniques are utilized. The advantage of using nonparametric techniques is that they are pretty robust (for example, to outliers). But fitting parametric models such as Weibull's may provide alternative descriptions to the same data.

Most melanoma patients are old people and naturally some of them are patients with some other diseases, too. This means that there are competing risks among the melanoma patients. It should be interesting to see how the competing risks influence the health of melanoma patients; that is, one may carry out a competing risks analysis.

Since British Columbia has more sunshine and beautiful beaches than the other three provinces and in general each province has its own characteristics, an analysis aimed at the difference(s) among the four provinces may provide additional knowledge about the behaviours of melanoma in Western Canada. This can be done when the whole data set has been cleaned.

As for the statistical analysis, there is still some work to be done to improve the present modeling techniques, especially, to provide more powerful techniques to check model adequacy. This includes improving the graphical methods in use now and developing easy-to-use analytical methods.

Finally, some general comments which provide additional information about the statistical techniques used will conclude this project.

The life-table method and the product-limit method usually give very close estimates of survival function. A good reference for the underlying theory is [8].

There are at least three different ways of looking at the Mantel test and the Breslow test. One way is to consider the difference between the observed number of deaths and the expected number of deaths; another way is by means of dummy variables and the Cox regression model; the third way is to use rank statistical argument. The first way is intuitively appealing, the second way can easily incorporate other regressor variables into the analysis, and the third way supplies a unified approach and in some case produces more efficient procedures. See [41], [43], [52] and the references contained.

The Cox regression model is based on the idea of partial likelihood. Justifications of the use of partial likelihood developed as below. Kalbfleisch and Prentice (1973) derived the partial likelihood as the marginal likelihood based on the rank

statistic for the data; Cox (1975) first used the name "partial likelihood" and outlined the asymptotical theory; Efron (1977) used the idea of "overall" hazard function and explained what was ignored from the full likelihood by using the partial likelihood. There were many other works along this line, and an important step achieved was Tsiatis' work (1981) which established the consistency and asymptotical normality of the partial likelihood estimator under a random independent censoring assumption. For an advanced and detailed treatment of the above outline, see [35].

The following problem of practical importance is unsolved so far. When a Cox model is built for a data set, one has an estimator for the regression parameter β and the related variance-covariance information. One can go on to estimate the base line survival function $S_0(t)$ and attach a variance to the estimator. The problem is: How can one supply a reasonable confidence interval when one predicts survival probability at any fixed time point using the above estimators?

REFERENCES

1. Andersen, P.K. (1982) Testing goodness of fit of Cox's regression and life model. Biometrics. 38:67-77.

2. Balch, C.M., *et al*. (1978) A multifactorial analysis of melanoma: prognostic histopathological features comparing Clark's and Breslow's staging methods. Annals of Surgery. 188:732-742.

3. Balch, C.M., *et al*. (1979) A multifactorial analysis of melanoma. II. Prognostic factors in patients with stage I (localized) melanoma. Surgery. 86:343-350.

4. Breslow, A. (1970) Thickness, cross-sectional area and depth of invasion in the prognosis of cutaneous melanoma. Annals of surgery. 172:902-908.

5. Breslow, A. (1979) Prognostic factors in the treatment of cutaneous melanoma. Journal of Cutaneous Pathology.6:208-212.

6. Breslow, N. (1970) A generalized Kruskal-Wallis test for comparing k samples subject to unequal patterns of censorship. Biometrika. 57:579-584.

7. Breslow, N. (1974) Covariance analysis of censored survival data. Biometrics. 30:89-99.

8. Breslow, N. and Crowley , J. (1974) A large sample study of the life table and product limit estimates under random censorship. The Annals of Statistics. 2:437-453.

9. Brown, M.B. and Benedetti, J.K. (1977) Sampling behavior of tests for correlation in two-way contingency tables. Journal of American Statistic Association. 72:609-315.

10. Bryson, M.C. and Johnson, M. (1981) The incidence of monotone likelihood in the Cox model. Technometrics. Vol.3, No 4:381-383.

11. Ca—A cancer journal for clinicians. (1979) Vol.29,No.4. American Cancer Society, Inc., New York.

12. Cascinelli, N., *et al*. (1985) Prognostic factors for stage I melanoma of the skin: A review. Statistics in Medicine. 4:265-278.

13. Cascinelli, N., *et al*. (1986) BANS A cutaneous region with no prognostic significance in patients with melanoma. Cancer 57:441-444.

14. Cochran, W.G. (1983) Planning and analysis of observational studies. John Wiley, New York.

15. Cramer, H. (1946) Mathematical methods of statistics. Princeton University Press, Princeton.

16. Conover, W.J. (1980) Practical nonparametric statistics. John Wiley, New York.

17. Cox, D.R. and Snell, E.J. (1968) A general definition of residuals. (with discussion) Journal of the Royal Statistical Society. B 30:248-275.

18. Cox, D.R. (1972) Regression models and [illegible] tables. (with discussions) Journal of the Royal Statistical Society. B 34:187-202.

19. Cox, D.R. (1975) Partial likelihood. Biometrika. 62:269-276.

20. Cox, D.R. and Oakes, D. (1984) Analysis of survival data. Chapman and Hall, Britain.

21. Crowley, J. and Storer, B.E. (1983) Comment. Journal of American Statistical Association. 78:277-281.

22. Day, C.L., *et al*. (1981) The natural break points for primary tumor thickness in clinical stage I melanoma. New England Journal Of medicine. 305:1155.

23. Day, C.L., *et al*. (1982) prognostic factors for melanoma patients with lesions 0.76-1.69 mm in thickness. Annals of Surgery. 195:30-34.

24. Day, C.L., *et al*. (1983) Malignant melanoma prognostic factors 2: the subsite concept. Journal of Dermatology, Surgery, Oncology. 9:525-526.

25. D'Agostino, R.B. and Stephens, M.A. edited (1986) Goodness-of-fit techniques. Marcel Dekker, Inc.

26. Dixon, W.J., *et al*. (1985) BMDP manual. University of California Press, Berkeley.

27. Draper, N.R. and Smith, H. (198') Applied regression analysis. 2ed. John Wiley, New York.

28. Drzewiecki, K.T., *et al*. (1982) Survival with malignant melanoma: a regression analysis of prognostic factors. Cancer. 49:2414-2419.

29. Eastwood, J. and Baker, T.G. (1983) Cutaneous malignant melanoma in West Yorkshire: I prospective study of variables, survival and prognosis. British Journal of Cancer. 51: 553-556.

30. Efron, B. (1977) The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association. 72:557-565.

31. Elandt-Johnson, R.C. and Johnson, N. (1980) Survival models and data analysis. John Wiley, New York.

32. Fitzpatrick, T.B. (1986) Melanoma. American Cancer Society Cancer Book. Edited by Arthur I. Hollbel. Doubleday & Company, Inc. Garden City, New Tork.

33. Gallagher, R.P. *et al*.. (1986) Risk factors for cutaneous malignant melanoma: The Western Canada Melanoma Study. Recent Results in Cancer Reseach. Vol.102:38-55. Springer-Verlag.

34. Kalbfleisch, J.D. and Prentice, R.L. (1973) Marginal likelihoods based on Cox's regression and life model. Biometrika. 60:267-279.

35. Kalbfleisch, J.D. and Prentice, R.L. (1980) The statistical analysis of failure time data. John Wiley, New York.

36. Kaplan, E.L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. Journal of American Statistical Association. 53:457-481.

37. Kay, R. (1977) Proportional hazard regression models and the analysis of censored survival data. Applied statistics. 26:227-237.

38. Koh, H.K., *et al*. (1984) Lentigo maligna melanoma has no better prognosis than other types of melanoma. Journal of Clinical Oncology. Vol.2, No 9:994-1001.

39. Kopf, A.W., *et al*. (1987) Prognostic index for malignant melanoma. Cancer. 59:1236-1241.

40. Lagakos, S.W. (1981) The graphical evaluation of exlanatory variables in proportional hazard regression models. Biometrika. 68:93-98.

41. Lawless, J.F. (1982) Statistical models and methods for lifetime data. John Wiley, New York.

42. Lee, E.T. (1980) Statistical methods for survival data analysis. Lifetime Learning Publications, Belmont, New York.

43. Mantel, N. (1966) Evaluation of survival data and two new order statistics arising in its consideration. Cancer Chemotherapy Reports. 50:163-170.

44. Mantel, N. (1970) Why stepdown preocedures in variable selection ? Technometrics. 12:621-625.

45. Mcbride, M.L. and Gallagher, R.P. (1987) Cancer in British Columbia: Incidence, motality, and trends. B.C. Medical Journal. Vol.29:41-46.

46. McCune, C.S. (1983) Chaper 21 Malignant Melanoma. Clinical oncology. Edited by Philip Rubin. American Cancer Society, Inc.

47. Miké, V. and Stanley, K.E. edited (1982) Statistics in medical research. John Wiley, New York.

48. Miller, R. Jr. (1981) Survival analysis. John Wiley, New York.

49. O'Oherty, C.J., *et al*. (1986) Sex differences in presentation of cutaneous malignant melanoma and in survival from stage I diseses.Cancer.58:788-792.

50. Neter, J., *et al*. (1985) Applied linear statistical models. 2ed. Richard D. Irwin, Inc, Illinois.

51. Peduzzi, P. *et al*. (1980) A stepwise procedure for nonlinear regression models.Miometrics. 36:511-516.

52. Prentice, R.L. (1978) Linear rank tests with right-censored data. Biometrika. 65:167-179.

53. Rao, C.R. (1973) Linear statistical inference and its application. 2ed. ed. John Wiley, New York.

54. Rigel, D.S., *et al*. (1985) Prognosis of malignant melanoma. Dermatologic Clinics. Vol.3, No 2:309-314.

55. Rogers, G.S., *et al*. (1986) Influence of anatomic location on prognosis of malignant melanoma: attempt to verify the BANS model. Journal of American Academy of Dermatology. Vol.15, No 2:231-236.

56. Schmoeckel, C. and Braun-Falco, O. (1978) Prognostic index in malignant melanoma. Arch Dermatol. Vol.114:871-873.

57. Schoenfeld, D. (1980) Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika. 67:145-153.

58. Rosenthal, S. *et al*. (1987) Medical care of the cancer patients W.B. Saunders Company, Philadelphia.

59. Silver, H.K.B., *et al*. (1983) Prediction of malignant melanoma recurrence by serum N-acetylneuraminic acid. International Journal of Cancer. 31:39-43.

60. Tsiatis, A.A. (1981) A large sample study of Cox's regression model. The Annals of Statistics. 9:93-108.

61. Trau, H., *et al*. (1983) Matestases of thin melanoma. Cancer. 31:553-556.

62. Woods, J.E., *et al*. (1985) Is the BANS concept for malignant melanoma valid ? The American Journal of Surgery. 150:452-455.

63. Upton, G.J.G. (1978) The analysis of cross-tabulated data. John Wiley, New York.