INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17” x 23” black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Bivariate survival methods for epidemiology: An application to the Framingham Heart Study of risk factors for cardiovascular disease

Wassell, James Terrence, Ph.D.

The Ohio State University, 1989

Copyright ©1989 by Wassell, James Terrence. All rights reserved.
Bivariate Survival Methods for Epidemiology:
An Application to the Framingham Heart Study
of Risk Factors for Cardiovascular Disease

Dissertation

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

* * * * *

The Ohio State University
1989

Dissertation Committee:
Richard R. Lanese, Ph.D.
Martin D. Keller, M.D., Ph.D.

Approved by
Melvin L. Moeschberger, Ph.D.
Department of Preventive Medicine.
This paper uses data supplied by the National Heart, Lung, and Blood Institute, NIH, DHHS from the Framingham Heart Study. The views expressed in this paper are those of the author and do not necessarily reflect the views of the National Heart, Lung, and Blood Institute or of the Framingham Study.
ACKNOWLEDGEMENTS

I am grateful to Dr. Melvin L. Moeschberger, my adviser, for his direction and guidance throughout this project. He has devoted many hours to this project and has helped me to overcome numerous difficulties encountered during the course of this research.

The other members of my dissertation committee have actively participated in this research and I am grateful for their valuable contributions. Dr. Richard R. Lanese raised several important questions which resulted in the refinement of this study and also suggested obtaining the Framingham Heart Study Data for this application. Dr. Martin D. Keller has provided many insightful suggestions to improve the quality of this dissertation. Dr. John P. Klein has provided recommendations for the improvement and future development of this investigation.

I am grateful to Robert J. Garrison and to the Framingham Study Policy Committee of the National Heart, Lung, and Blood Institute for supporting my request for access to data.

There are two additional people who have made many sacrifices in order to see this work through to completion. I express special thanks to my wife Barbara Ann and my son Terrence Stephen for their support and encouragement throughout the long course of my graduate studies.
VITA


1976 . . . . . . . . . B.S., Biology, University of Pittsburgh, Pittsburgh, Pennsylvania.


1981 . . . . . . . . . M.S., Zoology, Louisiana State University, Baton Rouge, Louisiana.


1982 - 1983 . . . . . Graduate Teaching Assistant, Ohio State University, Department of Statistics, Columbus, Ohio.

1983 - 1988 . . . . . Graduate Teaching Assistant, Ohio State University, Department of Preventive Medicine, Columbus, Ohio.
PUBLICATION


FIELDS OF STUDY

Major Field: Preventive Medicine

Studies in Quantitative Methods:

Professor Melvin L. Moeschberger

Studies in Epidemiology:

Professors Martin D. Keller and Richard R. Lanese
TABLE OF CONTENTS

PAGE

Acknowledgements ................................................................. ii
VITA ............................................................................................. iii
LIST OF TABLES ........................................................................ vii
LIST OF FIGURES ....................................................................... ix

CHAPTER

I. INTRODUCTION AND OVERVIEW ............................. 1

Significance ................................................................. 3
Methodology ................................................................. 5
Data Investigated .............................................................. 6
Expected Results ............................................................... 7
A Review of the Literature .................................................. 8
  Bivariate Survival Methods ............................................. 8
  Methods Used to Investigate Selected Risk Factors in Framingham ........................................... 12

II. A DESCRIPTION OF THE CLAYTON-OAKES MODEL . 20

Formulation of the Likelihood Function ......................... 21
A Likelihood Ratio Test for Independence ...................... 27
Conditional Expectation ................................................... 30

III. A DESCRIPTION OF THE LEE-KLEIN MODEL ......... 31

Formulation of the Likelihood Function ......................... 32
A Likelihood Ratio Test for Independence ...................... 35
IV. DEFINITIONS AND PRELIMINARY ANALYSES .......... 37

Exclusion Criteria ......................................................... 37
A Definition of Cardiovascular Disease ......................... 37
A Definition of Hypertension ........................................... 42
Anti-hypertensive Medication Use .................................... 47
Smoking ......................................................................... 49
Gender .......................................................................... 52
Serum Cholesterol .......................................................... 55
Body Mass Index ............................................................ 58
Logistic Regression Analyses .......................................... 61
Cox Regression Analyses ................................................ 63
Results .......................................................................... 63
Discussion ...................................................................... 66

V. CLAYTON-OAKES MODEL ANALYSES ......................... 71

Results and Discussion .................................................. 71

VI. LEE-KLEIN MODEL ANALYSES ................................. 85

Results and Discussion .................................................. 85

VII. SUMMARY AND CONCLUSIONS ............................... 97

Appendix A: Log Likelihood Function Derivatives,
Clayton-Oakes Model ..................................................... 100

Appendix B: Log Likelihood Function Derivatives, Lee-
Klein Model ................................................................. 110

Appendix C: FORTRAN Programs ................................. 117

Bibliography .................................................................. 149
### LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Numbers of Framingham Participants Excluded from Analysis and Reasons for Exclusion.</td>
<td>38</td>
</tr>
<tr>
<td>2. Type and Frequency of Cardiovascular Events.</td>
<td>41</td>
</tr>
<tr>
<td>3. Frequency of Various Temporal Relationships of First Cardiovascular Disease Event and First Detection of Hypertension.</td>
<td>46</td>
</tr>
<tr>
<td>4. Frequency of Cardiovascular Disease Among Normotensives, Hypertensives on Medication and Hypertensives Not on Medication.</td>
<td>48</td>
</tr>
<tr>
<td>5. Frequency of Cardiovascular Disease by Smoking Status Among Normotensives and Hypertensives.</td>
<td>50</td>
</tr>
<tr>
<td>6. Frequency of Cardiovascular Disease by Gender Among Normotensives and Hypertensives.</td>
<td>53</td>
</tr>
<tr>
<td>7. Frequency of Cardiovascular Disease by Serum Cholesterol Status Among Normotensives and Hypertensives.</td>
<td>56</td>
</tr>
<tr>
<td>8. Frequency of Cardiovascular Disease by Body Mass Index Group Among Normotensives and Hypertensives.</td>
<td>59</td>
</tr>
<tr>
<td>9. Logistic Regression: Two Variable Analyses Controlling for Hypertension.</td>
<td>62</td>
</tr>
<tr>
<td>10. Cox Regression: Two Variable Analyses Controlling for Hypertension.</td>
<td>68</td>
</tr>
<tr>
<td>11. Cox Regression: Multivariate Analysis Controlling for Hypertension.</td>
<td>69</td>
</tr>
<tr>
<td>12. Cox Regression, Hypertensives: Two Variable and Multivariate Analysis for Individuals with Prior Hypertension (N=1087).</td>
<td>70</td>
</tr>
<tr>
<td>13. Clayton-Oakes Model: Univariate Analysis.</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Clayton-Oakes Model: Multivariate Analysis.</td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>14</td>
<td>Clayton-Oakes Model: Multivariate Analysis with Interactions.</td>
</tr>
<tr>
<td>15</td>
<td>Clayton-Oakes Model, Hypertensives: Univariate and Multivariate Analyses for Individuals Ever Hypertensive (N=1192).</td>
</tr>
<tr>
<td>16</td>
<td>Clayton-Oakes Model, Hypertensives: Multivariate Analysis with Interactions for Individuals Ever Hypertensive (N=1192).</td>
</tr>
<tr>
<td>17</td>
<td>Lee-Klein Model: Univariate Analysis.</td>
</tr>
<tr>
<td>18</td>
<td>Lee-Klein Model: Multivariate Analysis.</td>
</tr>
<tr>
<td>19</td>
<td>Lee-Klein Model: Multivariate Analysis with Interactions.</td>
</tr>
<tr>
<td>20</td>
<td>Lee-Klein Model, Hypertensives: Univariate and Multivariate Analyses for Individuals Ever Hypertensive (N=1192).</td>
</tr>
<tr>
<td>21</td>
<td>Lee-Klein Model, Hypertensives: Multivariate Analysis with Interactions for Individuals Ever Hypertensive (N=1192).</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Probability of Survival Free of Cardiovascular Disease by Smoking Status.</td>
<td>51</td>
</tr>
<tr>
<td>2. Probability of Survival Free of Cardiovascular Disease by Gender.</td>
<td>54</td>
</tr>
<tr>
<td>3. Probability of Survival Free of Cardiovascular Disease by Serum Cholesterol Status.</td>
<td>57</td>
</tr>
<tr>
<td>4. Probability of Survival Free of Cardiovascular Disease by Body Mass Index Group.</td>
<td>60</td>
</tr>
<tr>
<td>5. Clayton-Oakes Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, Gender and Anti-hypertensive Medication Status.</td>
<td>79</td>
</tr>
<tr>
<td>6. Clayton-Oakes Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, Gender and Smoking Status.</td>
<td>80</td>
</tr>
<tr>
<td>7. Clayton-Oakes Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, and Serum Cholesterol Status.</td>
<td>81</td>
</tr>
<tr>
<td>8. Clayton-Oakes Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, and Body Mass Index Group.</td>
<td>82</td>
</tr>
<tr>
<td>9. Lee-Klein Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, Gender and Anti-hypertensive Medication Status.</td>
<td>91</td>
</tr>
<tr>
<td>10. Lee-Klein Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, Gender and Smoking Status.</td>
<td>92</td>
</tr>
<tr>
<td>11. Lee-Klein Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of</td>
<td></td>
</tr>
</tbody>
</table>
12. Lee-Klein Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, and Body Mass Index Group.
CHAPTER I
INTRODUCTION AND OVERVIEW

The purpose of this dissertation is the application of two models of bivariate survival analysis to the relationship between the age at onset of hypertension and the age at onset of cardiovascular disease. The Framingham Heart Study which monitored a large number of people for hypertension and cardiovascular disease provided an ideal dataset for this investigation. After 10 biennial examinations, there were 1192 detections of hypertension and 727 cardiovascular events in a group of people who were previously disease-free. The age at which the first episode of cardiovascular disease occurred was recorded (generically, this may be called an "observed time" or a "failure time"). For those people free of cardiovascular disease at the 18 year follow-up, the age at which they were last known to be disease-free was recorded (generically called a "censored time" because cardiovascular disease was not detected). With respect to hypertensive status, an observed time was recorded for the age at first detection of hypertension or a censored time for the age at which a person was last known to be normotensive. Because the onset of hypertension is usually asymptomatic and unrecognized, the age at first detection of hypertension is used as a proxy for the age at onset in this investigation.
Logistic regression methods have most recently been the main statistical tools for the analysis of The Framingham Heart Study data. However, logistic regression does not utilize information about the specific age at which cardiovascular disease occurred. Survival analysis methods, however, were developed for the analysis of data which represent the time (or age) at which some event, such as cardiovascular disease, occurred.

Standard survival analysis methods such as proportional hazards regression or Cox regression (Cox, D. R., 1972) have been developed to accommodate the analysis of studies in which there is a single event of interest. Proportional hazards regression could be used in an analysis of the age at first cardiovascular disease or its censored counterpart, the age at last follow-up.

In addition to the information about the age at first evidence of cardiovascular disease or last follow-up, bivariate methods can be used to incorporate information about the age at first detection of hypertension or, its censored counterpart for normotensives, the age at last normal blood pressure determination. In this way, the bivariate methods investigated in this dissertation utilize information about two events, either or both being possibly censored.

This approach illustrates how the use of detailed information about hypertension onset may offer better understanding of the relationship between hypertension and cardiovascular disease. The new methods will provide an estimate of the dependence or correlation between the age at first cardiovascular disease and the age at first detected hypertension.
Based on this estimate of dependence, it is possible to investigate the increased risk of cardiovascular disease associated with hypertension.

In order to provide meaningful results from these analyses, some provision must be made for the estimation and control of the effects of 'traditional' risk factors. Differences in the ages of the first manifestation of cardiovascular disease for men and women and the effects of smoking, obesity, elevated cholesterol and anti-hypertensive medication are documented in epidemiologic literature. The survival analysis methods investigated in this dissertation have been adapted to include covariates that allow adjustment for these effects. Also, it is possible to estimate coefficients and obtain significance tests for these other risk factors while controlling for the age at first detection of hypertension.

These bivariate survival analysis methods allow estimation of the age at first evidence of cardiovascular disease conditional on knowing the age at first detection of hypertension and other risk factors. Consequently, the analytical results can be given an interpretation analogous to ordinary regression, even though the methodology is quite different, considerably more complex, and allows the use of more information.

**SIGNIFICANCE**

Censoring occurs in clinical applications of survival time studies as the result of a loss of follow-up on patients or as a result of limitations on the duration of a study. Those individuals who are
observed for a portion of the study, and who do not experience an event by the end of the study, are censored. A primary goal of this dissertation is to use information about these censored individuals in the estimation of the correlation between the age at first cardiovascular disease and the age at first detection of hypertension and, consequently, to determine an estimate of age at first cardiovascular disease conditional on knowing the age at first detection of hypertension and the status of other known risk factors.

In contrast to the two groups of observations (observed times and censored times) which result from studies in which the time for only one event is the variable of interest, there are four groups of observations which occurred in the Framingham bivariate survival data.

3. Hypertensives who are free of cardiovascular disease at the end of the study.
4. Normotensives who are free of cardiovascular disease at the end of the study.

In Chapters II and III, a statistical methodology is developed from a likelihood function consisting of four components which correspond to these four groups of observations and which incorporate covariates associated with the individuals in the study.

One major contribution of this research is to make available to other researchers at other institutions a computer program which can be used to analyze bivariate survival times. At the present time, none of
the statistical packages include programs which are suitable for the estimation of correlation and covariate effects in the analysis of bivariate possibly censored times.

**METHODOLOGY**

The two bivariate survival models which will be adapted for censored observations and covariates are based on recent research (Clayton, D. G., 1978; Oakes, D., 1982a; Lee, S., 1986; Lee, S., and J. P. Klein, 1986). These models will be referred to as the Clayton-Oakes Model and the Lee-Klein Model throughout this text. For the purposes of this dissertation, these two bivariate models of association differ with respect to the estimation of expected age of cardiovascular disease. This can be attributed to differences in their conditional survival functions.

For the models considered in this dissertation, the Weibull distribution has been chosen as the underlying survival function. It provides a good fit to the age distribution of chronic diseases in man and has considerable flexibility in modeling times to event (Cook, P. J., et al., 1969). The Weibull distribution can be used in situations where the hazard rate is increasing with time, constant over time, or decreasing with time.

The bivariate survival models considered in this dissertation are based on an assumption which requires explanation. The proportional hazards assumption of these bivariate survival models (not to be confused with the proportional hazards assumption of the Cox regression model) can be given the following interpretation in terms of the
application investigated here: the hazard of cardiovascular disease among hypertensives relative to the hazard of cardiovascular disease among normotensives is a constant at all ages. Risk ratio estimates obtained from these models must be considered as a measure of the average risk over all ages.

Iterative maximum likelihood procedures have been derived and written to obtain estimates of coefficients through FORTRAN programs using IMSL subroutines. An optimization scheme known as Marquart's Compromise (David, H. A., and M. L. Moeschberger, 1978) is used to maximize the likelihood function. A listing of these programs is included in Appendix C.

Programs for management of the data and for descriptive results were written using the Statistical Analysis System (SAS). Procedures including Kaplan-Meier survival curves, logistic regression, and Cox proportional hazard regression were also conducted using SAS. The Cox proportional hazard model with time-dependent covariates was analyzed using the Biomedical Data Programs (BMDP).

DATA INVESTIGATED

The Framingham Study Policy Committee of the National Heart, Lung, and Blood Institute of the National Institutes of Health has provided to this investigator the eighteen year follow-up data from the Framingham Longitudinal Study of cardiovascular disease risk factors. The age at first evidence of cardiovascular disease and detailed information regarding blood pressure measurements are available in this
dataset. A total of 5209 individuals (median age of 42 years) were enrolled in this study. After excluding individuals with evidence of pre-existing cardiovascular disease, hypertension, and missing covariate information, a total of 4531 individuals were represented in this analysis.

EXPECTED RESULTS

The results of an application of the bivariate survival methods may be given a 'regression-like' interpretation which offers many advantages. It will be possible to estimate the age at first evidence of cardiovascular disease controlling for the age at first detection of hypertension and other risk factors. Information of this nature would be of considerable use to clinicians as well as health researchers and administrators.

Recent studies have demonstrated that the failure to model dependency in survival times when such dependency exists will result in biased estimates (Moeschberger, M. L., and J. P. Klein, 1984; Klein, J. P., and M. L. Moeschberger, 1988) The estimates of covariate effects will be influenced by controlling for a time variable which may be subject to censoring. The evaluation of risk factors for cardiovascular disease is anticipated to differ from results based on standard statistical methods as a result of incorporating information about the age at first detection of hypertension in a precise manner.
A REVIEW OF THE LITERATURE

Bivariate Survival Methods

Partially nonparametric approaches to the analysis of paired survival times have been proposed, but these models have some limitations. The proportional hazard model (Cox, D. R., 1972) can accommodate covariates which change through time and precede the event of interest. The time-dependent covariate of the Cox regression model provides no estimate of the degree of dependence between the two times of interest. More importantly, the proportional hazards model does not permit the use of a censored time as a covariate.

A bivariate model of association was first proposed as a tool for investigation of the genetic predisposition to chronic diseases (Clayton, D. G., 1978). Specifically, an application to the study of heart disease among fathers and their sons was proposed. The association between the survival of fathers and sons is postulated to arise from an unobserved common environmental or genetic effect. This type of process which gives rise to dependent survival times has been termed a 'random environmental effects' process. The unobserved common factor acts as an exponent on the marginal survival probabilities of the father and the son, resulting in a common improvement or degradation of survival.

Clayton recognized that the problems of statistical analysis are compounded by the censored nature of the data. Methods for analyzing univariate censored survival data have been well established; however, in bivariate data, censoring creates additional complications. Among
father-son pairs, often only one member of the pair will have evidence of heart disease, while the other member of the pair is disease-free. Both members of the pair may be observed for varying lengths of time, depending on the duration of the study, without having shown any evidence of disease.

In keeping with the spirit of the proportional hazards model (Cox, D. R., 1972), the model proposed by Clayton is based on proportional conditional hazards which are constant over all times. This constant conditional hazard ratio may be interpreted as a relative risk measure in some applications. The numerator of the ratio is conditional on the failure of the second member of the pair, while the denominator of the ratio is conditional on the survival of the second member of the pair.

In Clayton's proposed application, the hazard for sons whose fathers die at some age = t is a constant multiple of the hazard for sons whose fathers have survived until age = t. This hazard ratio is constant for all ages of the sons and for all ages of the fathers. This same constant ratio of hazards described for the sons is equal to a similarly defined hazard ratio which could be described for the fathers.

The importance of controlling for observed covariates in the analysis of familial tendency in heart disease was stressed in Clayton's work. Specifically, is there an association between disease incidence in fathers and disease incidence in sons if the effect of height is removed? The Weibull distribution was recommended as a model for survival times with covariates, and a general form of a likelihood function for such a model was considered.
A nonparametric concordance measure known as 'Kendall's tau' (Kendall, M. G., 1938) is a widely accepted statistic for estimating bivariate correlation. In a comparison of any two pairs, if both members of one pair have values larger (or smaller) than the corresponding members of another pair, the pairs are said to be concordant. Otherwise, the pairs are discordant. In comparing all possible pairs, if the number of concordant pairs exceeds the number of discordant pairs, a positive correlation is indicated. Kendall's tau is a nonparametric measure because it uses information about the relative rankings of the members of pairs rather than the actual observed values.

Spearman's rho statistic is another nonparametric estimator of correlation which differs from Kendall's tau. Kendall's tau assigns the value of +1 to concordant pairs and the value of -1 to discordant pairs, and is the average of these values over all possible pair comparisons. Spearman's rho assigns greater weights to pairs with greater differences in relative rankings thus using some information about the magnitude of the differences.

Nonparametric approaches to estimating the degree of association in paired data with censoring began with the modification of Kendall's tau. Several estimators of correlation have been compared to the modified Kendall's tau in a power study (Weier, D. R., and A. P. Basu, 1980). The equivalence of Kendall's coefficient and the dependence parameter (conditional hazard ratio) of the bivariate model introduced by Clayton has been demonstrated explicitly (Oakes, D., 1982a). As a
result, the association between bivariate survival times can be given two interpretations: a conditional hazard ratio (relative risk measure) and a bivariate correlation coefficient (Kendall's tau). A general form of the bivariate density and derivatives necessary for maximum likelihood estimation of parameters for complete data has also been provided in this paper.

This bivariate survival function has been presented in several other publications (Oakes, D., 1982b, 1986, 1987; Cox, D. R., and D. Oakes, 1984).

The Lee-Klein Model is based on a different bivariate survival function (Lee, S., 1986; Lee, S., and J. P. Klein, 1986, 1987). The marginal distributions take on the form of a Burr distribution (Johnson, N. L., and S. Kotz, 1970) by choosing a Weibull hazard function for the cumulative within-group hazard function. The dependence structure of this model is demonstrated to be identical to the dependence structure of the Clayton-Oakes Model. The relationship between the dependence parameter of this model and Kendall's tau is shown. The application of this model to the study of reliability in series components systems for industrial design is considered in detail.

There is a limitation in adapting this model to epidemiological problems. Certain situations may arise where an estimate of the marginal expected lifetime is not available. When the Weibull shape parameter is small (the failure rate is decreasing over time), the model is not usable for highly correlated failure times. In the case where the Weibull shape parameter is 1.0 (the exponential model or constant failure rate) the maximal correlation is 0.50.
The Lee-Klein Model has been generalized to groups with more than two individuals all sharing some common characteristic (Wild, C. J., 1983). A general form of the likelihood function was presented which consists of a product of quantities contributed by each pair or group and allows groups of different sizes. With this specification, not only pairs, but triplets and larger groups could all be combined in the same analysis to estimate covariate effects.

A family of multivariate distributions which includes both the Clayton-Oakes and Lee-Klein Models has been described (Cook, R. D., and M. E. Johnson, 1981). This generalization of the two models is based on the form of the marginal distributions, and permits extension to matched groups larger than pairs. A general representation of the conditional expectation and methods for data simulation are presented. A model with the assumption of normal marginal distributions is applied to (uncensored) geological survey data and compared to estimates based on the standard bivariate normal distribution.

**Methods Used to Investigate Selected Risk Factors in Framingham**

This section of the literature review is organized according to the various methodologic approaches to the estimation of covariate effects in the Framingham Heart Study. A vast body of literature has evolved as a result of the Framingham Heart Study. This literature review is limited to the risk factors and effects investigated in the subsequent analysis: hypertension, anti-hypertension medication, smoking, gender, serum cholesterol, and obesity measured by a body mass index.
Most analyses are based on data representing a follow-up period different from the 18 years of follow-up used in this dissertation. The results of these analyses are reviewed but an appreciation of the effects of a longer or shorter follow-up must be recognized in any comparisons with the analyses presented here.

Participants in the Framingham Study consist of a homogeneous group from a limited geographical area which includes no Blacks. However, the results of Framingham analyses have recently been compared with results based on the NHANES I national probability sample of white adults (Leaverton, P. E., et al., 1987). With regard to hypertension, cholesterol and smoking, results based on the Framingham Study are shown to be generalizable to the United States adult white population.

There is one major reference which requires special treatment in this literature review, because it provided some initial direction in the selection of the risk factors investigated in this study. *The Framingham Study* (Dawber, T. R., 1980) describes the background and findings of this longitudinal study based on 24 years follow-up. In addition to the narrative presentation regarding risk factors for the various manifestations of cardiovascular disease, summary statistics in the form of tables and charts are presented.

The bivariate analysis conducted in this dissertation can be seen as an attempt to improve upon an analysis based on simple means in *The Framingham Study*. In the investigation of the effects of risk factors, one outcome measure used is the average age of onset of cardiovascular
disease. The average age of onset of coronary heart disease and sudden death is found to be earlier in smokers than nonsmokers. The age at which coronary heart disease develops is approximately two years earlier in men than in women. When cross-classified by hypertensive status at examination 1, gender and age cohort, the average age of onset for coronary heart disease is earliest for hypertensive men.

An investigation concerned with the age of hypertension detection as a risk factor for the age of onset of cardiovascular disease is a direct extension of the approach presented in *The Framingham Study*. The application of bivariate survival analysis methodology is intended to provide better estimates of the effects of hypertension and other risk factors on the age of first cardiovascular disease. A number of studies suggest that a method which takes account of the age at which a person crosses the threshold from normotensive to hypertensive, in addition to other risk factors, could be a good predictor of subsequent cardiovascular disease.

The logistic regression methodology, which was developed specifically for the analysis of the Framingham Heart Study, has provided many of the estimates of cardiovascular risk factors for which the Framingham Study is so well known (Truett, J., et al., 1967).

One study has utilized both the methods of least squares regression and logistic regression in an investigation of the effects of blood pressure change on subsequent cardiovascular disease (Hofman, A., et al., 1983). For each individual, the change in blood pressure over twelve exams was determined by least squares regression. This estimate of
slope was compared with the baseline blood pressure in a logistic regression to predict cardiovascular disease. Hypertensive status and the change in blood pressure over time were both related to subsequent cardiovascular disease.

Estimates of life expectancy for hypertensives on medication and hypertensives not on medication provide the basis for a comprehensive cost-effectiveness analysis (Weinstein, M. C., and W. B. Stason, 1976). Including adjustments for side effects and morbidity (quality adjusted life-years), anti-hypertensive medication is estimated to provide increases in life expectancy from 2.3 to 5.0 years for women and 1.4 to 8.1 years for men, depending on age.

Life expectancy was determined by adjusting the national life table mortality rates by the mortality rates based on the logistic regression estimates from the Framingham Heart Study (Gordon, T., and D. Shurtleff, 1973; Shurtleff, D., 1974). The effects of risk factors on life expectancy should be similar to the effects of risk factors on the age at first cardiovascular disease. These estimates are based on the same data, therefore, it is reasonable to anticipate similar results even though the outcome measure is different.

Logistic regression has been used in a way referred to as the 'pooling of repeated observations method' (Cupples, L. A., et al., 1988). This method treats each of the biennial examinations as independent two-year follow-up studies. This method does not account for the dependence which results from repeated observations on the same subject, rather than tracking individuals over time this method has been
described as tracking risk factors over time. This method is motivated by a desire to utilize all of the information obtained from the repeated interviewing of study subjects in a longitudinal study.

The 'pooling of repeated observations' method with logistic regression has been used to investigate the effects of risk factors on cardiovascular disease based on fifteen examinations (Stokes, J., et al., 1987). This method has been used to estimate the effects of anti-hypertensive treatment on subsequent cardiovascular disease (Shea, S., et al., 1985).

Other studies have a more narrow focus on specific manifestations of cardiovascular disease. Estimates of risk factors for sudden death (Shatzkin, A., et al., 1984), stroke (Wolf, P. A., et al., 1987), and peripheral arterial disease (Kannel, W. B., and D. L. McGee, 1985) have utilized logistic regression by the 'pooling of repeated observations' method.

A more moderate approach to 'pooling of repeated observations' uses eight-year follow-up intervals and ignores some of the intervening risk factor measurements. The relationship between baseline measurements of risk factors and subsequent cardiovascular disease at eight years follow-up has been explored through logistic regression (McGee, D., 1973). Using these eight-year predictions, the effect of blood pressure reduction on cardiovascular disease has been shown to differ for men and women, and for combinations of other risk factors; emphasizing the necessity of a multivariate approach. (Madhavan, S., and M. H. Alderman, 1981).
Logistic regression methods have been used to demonstrate that hypertension is a risk factor for both recognized and unrecognized myocardial infarction based on blood pressure measurements taken two years prior (Kannel, W. B., et al., 1985). Baseline measurements have been used to demonstrate the effect of obesity, independent of other risk factors, on the probability of cardiovascular disease twenty-six years later (Hubert, H. B., et al., 1983).

Although logistic regression methods have been widely used in the analysis of the Framingham Heart Study, survival analysis methods are being used more frequently. The nonparametric method of Kaplan-Meier (Kaplan, E. L., and P. Meier, 1958) has been used to investigate the long term effects of anti-hypertensive medication. Some of the people who reported anti-hypertensive medication use at two successive examinations later ceased medication use, but maintained normotensive blood pressures. Kaplan-Meier methods were used to examine the length of time these individuals maintained normal blood pressures following cessation of anti-hypertensive medication use (Dannenberg, A. L., et al., 1987).

The proportional hazard survival model (Cox, D. R., 1972) has been used in the analysis of risk factors in the Framingham Heart Study. Cardiovascular disease, broadly defined, is the outcome of interest in investigations of the effects of hypertension and other risk factors using the proportional hazard method (Harris, T., et al., 1985; Kannel, W. B., et al., 1987). Recently, the relationship between cigarette smoking and stroke has been reexamined in the Framingham Heart Study using the proportional hazard model (Wolf, P. A., et al., 1988).
The traditional outcome measure of survival analysis, mortality, has also been examined using the proportional hazard model in the Framingham Heart Study. A measure of body mass has been used to investigate the effect of obesity on mortality (Harris, T., et al., 1988). The effect of elevated serum cholesterol on subsequent death, has also been investigated by the proportional hazard model (Anderson, K. M., et al., 1987).

In summary, the covariates investigated in this dissertation have all been shown to be important risk factors for cardiovascular disease based on several different methodological approaches. None of the research to date has utilized information about the age at first detection of hypertension as a predictor of the age at first cardiovascular disease. There have been some attempts to estimate the effects of gender, hypertension, anti-hypertensive medication and age on the average age of onset for coronary heart disease and life expectancy. The methods which have been used to obtain these estimates have not taken advantage of survival analytic approaches.

In Chapter II, the Clayton-Oakes bivariate survival function is used to develop a method for the analysis of the Framingham Heart Study data. Explicit computational details of the likelihood function, a likelihood ratio test of independence and conditional expectation are provided. In Chapter III, the Lee-Klein bivariate survival function is used to develop a competing analytical method with explicit details of the likelihood function and conditional expectation.
Chapter IV provides detailed descriptions of the outcomes and covariates investigated in the analysis of the Framingham Heart Study. The covariate effects are investigated through the applications of some standard statistical methods: cross-classification frequency tables, Kaplan-Meier survival curves, logistic and Cox regression.

In Chapters V and VI, the analytical results of the Clayton-Oakes and Lee-Klein Models are presented in the form of tables of the parameter estimates and graphs of the expected age of cardiovascular disease by the age at first detection of hypertension and other risk factors. Included is a discussion of significant findings. Finally, in Chapter VII, the two models are compared on the basis of several criteria.
CHAPTER II
A DESCRIPTION OF THE CLAYTON-OAKES MODEL

In this chapter, the notion of the four groups of individuals which occur in The Framingham Heart Study (outlined in Chapter I) is developed into an analytical method. The maximum likelihood method can be applied to answer the following question: Given the data observed in the Framingham Heart Study, what is the most probable (likely) estimate of the correlation between the age at first detection of hypertension and the age at first cardiovascular disease (and other risk factor coefficients and parametric estimates)?

It is important to recognize that every individual contributes some information to the estimates obtained through this method. This is a desirable property, especially in consideration of the financial resources required to obtain long term repeated measurements of risk factors for those individuals who never develop hypertension or cardiovascular disease during the course of the study.
FORMULATION OF THE LIKELIHOOD FUNCTION

A likelihood function is developed which allows for the occurrence of hypertension and cardiovascular disease in an individual. The occurrence of any one event does not preclude the occurrence of the other event. The likelihood function that allows for right censoring of the times for either or both events is a multinomial form that consists of four components.

\[
L = \prod_{i} \left\{ P(X=x_i, Y=y_i) \right\}^{N_1} \times \left\{ P(X=x_i, Y>y_i) \right\}^{N_2} \times \left\{ P(X>x_i, Y=y_i) \right\}^{N_3} \times \left\{ P(X>x_i, Y>y_i) \right\}^{N_4}
\]  

(1)

The corresponding log likelihood is written as:

\[
\log L = \sum_{i \in g_1} \log \left\{ P(X=x_i, Y=y_i) \right\} + \sum_{i \in g_2} \log \left\{ P(X=x_i, Y>y_i) \right\} + \sum_{i \in g_3} \log \left\{ P(X>x_i, Y=y_i) \right\} + \sum_{i \in g_4} \log \left\{ P(X>x_i, Y>y_i) \right\}
\]  

(2)

\(x_i\) is the observed or censored time for the first type of event (age at first evidence of cardiovascular disease) in the \(i^{th}\) individual. \(y_i\) is the observed or censored time of the second type of event (age at first detection of hypertension) in the \(i^{th}\) individual. The individual is a member of group one \((i \in g_1)\) if \(x_i\) and \(y_i\) are both observed times (hypertensive with cardiovascular disease). The individual is a member of group two \((i \in g_2)\) if \(x_i\) is an observed time and \(y_i\) is a censored time (normotensive with cardiovascular disease). The individual is a member of group three \((i \in g_3)\) if \(x_i\) is a censored time and \(y_i\) is an observed time (hypertensive free of cardiovascular disease). The
individual is a member of group four \((i \in g_4)\) if \(x_i\) and \(y_i\) are both censored times (normotensive free of cardiovascular disease). \(N_1 (=317)\), \(N_2 (=410)\), \(N_3 (=875)\), and \(N_4 (=2929)\) are the number of individuals in group 1, group 2, group 3, and group 4, respectively.

The bivariate survival function, Equation 3, is the probability that an individual has not experienced either event at the end of the study. The other probabilities are obtained by differentiation of the survival function.

\[
P(X>x, Y>y) = S(x, y) \tag{3}
\]

\[
P(X=x, Y>y) = -\frac{\partial S(x, y)}{\partial x} \tag{4}
\]

\[
P(X>x, Y=y) = -\frac{\partial S(x, y)}{\partial y} \tag{5}
\]

\[
P(X=x, Y=y) = f(x, y) = \frac{\partial^2 S(x, y)}{\partial x \partial y} \tag{6}
\]

The Clayton-Oakes Model will allow a dependence to exist between the times for the two types of events. Let \(F_0(x)\) and \(G_0(y)\) be the marginal survival distributions for the first type of event and the second type of event, respectively. The parameter \(\theta\) is a measure of dependence which is related to Kendall's tau coefficient of concordance.

\[
\tau = \frac{\theta-1}{\theta+1} \tag{7}
\]

The joint survival function (Oakes, D., 1982a) can be written as follows:

\[
S(x, y) = \left\{ F_0(x)^{1-\theta} + G_0(y)^{1-\theta} - 1 \right\}^{\frac{1}{1-\theta}} \tag{8}
\]
Notice that as $\theta$ approaches 1 (Equation 9) there is independence of the times for the two types of events. This will permit a test for independence, by the likelihood ratio test.

$$\lim_{\theta \to 1} S(x,y) = F_0(x) \cdot G_0(y)$$

(9)

In general, the derivatives necessary to calculate the likelihood function may be written. The bivariate density function, $f(x,y)$ (Equation 10) is equivalent to an equation already published (Oakes, D., 1982a).

$$\frac{\partial^2 S(x,y)}{\partial x \partial y} = \frac{\theta}{(1-\theta)^2} \left\{ F_0(x)^{1-\theta} + G_0(y)^{1-\theta} - 1 \right\}$$

$$\frac{\partial}{\partial x} \left\{ F_0(x)^{1-\theta} \right\} \quad \text{and} \quad \frac{\partial}{\partial y} \left\{ G_0(y)^{1-\theta} \right\}$$

(10)

$$\frac{\partial S(x,y)}{\partial x} = \frac{-1}{1-\theta} \left\{ F_0(x)^{1-\theta} + G_0(y)^{1-\theta} - 1 \right\} \frac{\partial}{\partial x} \left\{ F_0(x)^{1-\theta} \right\}$$

(11)

$$\frac{\partial S(x,y)}{\partial y} = \frac{-1}{1-\theta} \left\{ F_0(x)^{1-\theta} + G_0(y)^{1-\theta} - 1 \right\} \frac{\partial}{\partial y} \left\{ G_0(y)^{1-\theta} \right\}$$

(12)

The Weibull distribution is used to define the marginal survival functions (Equation 13 and Equation 14) with their corresponding density functions (Equation 15 and Equation 16) for the two event times.

$$F_0(x_i) = \exp(-\lambda_1 x_i^{\eta_1}) \quad \lambda_1, \eta_1 > 0$$

(13)

$$G_0(y_i) = \exp(-\lambda_2 y_i^{\eta_2}) \quad \lambda_2, \eta_2 > 0$$

(14)

$$f_0(x_i) = \lambda_1 \eta_1 x_i^{\eta_1-1} \exp(-\lambda_1 x_i^{\eta_1})$$

(15)

$$g_0(y_i) = \lambda_2 \eta_2 y_i^{\eta_2-1} \exp(-\lambda_2 y_i^{\eta_2})$$

(16)
Covariates are introduced into the Weibull scale parameters (Equation 17 and Equation 18).

\[ \lambda_1 = \exp(\beta'_1 z_{1i}) \]  
\[ \lambda_2 = \exp(\beta'_2 z_{2i}) \]  

Because \( 1 < \theta < \infty \), a more convenient form (Equation 19) is used which allows for the possible inclusion of covariate effects on the degree of association.

\[ \theta - 1 = \exp(\beta'_3 z_{3i}) \]  

It is useful to express the estimate of Kendall's tau in terms of the estimated coefficient (Equation 20).

\[ \tau = \frac{\exp(\beta'_3 z_{3i})}{\exp(\beta'_3 z_{3i}) + 2} \]  

In the above notation, \( z_{1i} \), \( z_{2i} \), and \( z_{3i} \) are row vectors of covariates for the first type of event for the \( i^{th} \) individual, the second type of event for the \( i^{th} \) individual, and the parameter of association, respectively. \( \beta'_1 \), \( \beta'_2 \) and \( \beta'_3 \) are column vectors of coefficients so that,

\[ \beta'_1 z_{1i} = \sum_{j=1}^{J} \beta'_{1j} z_{1ij} \]  
\[ \beta'_2 z_{2i} = \sum_{k=1}^{K} \beta'_{2k} z_{2ik} \]  
\[ \beta'_3 z_{3i} = \sum_{l=1}^{L} \beta'_{3l} z_{3il} \]  

where \( J \), \( K \), and \( L \) are the numbers of covariates for the first type of event, the second type of event, and for the association parameter, respectively. To simplify the notation let...
\[ Q_i = \exp \left( x_i^{\eta_1} \exp(\beta'_1 z_{1i} + \beta'_3 z_{3i}) \right) + \exp \left( y_i^{\eta_2} \exp(\beta'_2 z_{2i} + \beta'_3 z_{3i}) \right) - 1 \]  

(24)

Now, the four components of the likelihood may be written.

\[ P(X=x_i, Y=y_i) = \left\{ \exp(\beta'_3 z_{3i}) + 1 \right\}^{-1} Q_i \]

\[ \eta_1 \exp(\beta'_1 z_{1i}) \exp \left( x_i^{\eta_1} \exp(\beta'_1 z_{1i} + \beta'_3 z_{3i}) \right) \]

\[ \eta_2 \exp(\beta'_2 z_{2i}) \exp \left( y_i^{\eta_2} \exp(\beta'_2 z_{2i} + \beta'_3 z_{3i}) \right) \]  

(25)

Taking logarithms of the four components simplifies the expressions and permits writing the log likelihood function.

\[ \log P(X=x_i, Y=y_i) = \log \left\{ \exp(\beta'_3 z_{3i}) + 1 \right\} + \frac{1+2\exp(\beta'_3 z_{3i})}{\exp(\beta'_3 z_{3i})} \log(Q_i) + \]

\[ \beta'_1 z_{1i} + \beta'_3 z_{3i} + \log(\eta_1) + \log(\eta_2) + (\eta_1 - 1) \log(x_i) + (\eta_2 - 1) \log(y_i) + \]

\[ x_i^{\eta_1} \exp(\beta'_1 z_{1i} + \beta'_3 z_{3i}) + y_i^{\eta_2} \exp(\beta'_2 z_{2i} + \beta'_3 z_{3i}) \]  

(29)

\[ \log P(X=x_i, Y=y_i) = \frac{\exp(\beta'_3 z_{3i}) + 1}{-\exp(\beta'_3 z_{3i})} \log(Q_i) + \beta'_1 z_{1i} + \log(\eta_1) + \]

\[ (\eta_1 - 1) \log(x_i) + x_i^{\eta_1} \exp(\beta'_1 z_{1i} + \beta'_3 z_{3i}) \]  

(30)
\[
\log P(X > x_i, Y = y_i) = \begin{cases} \frac{\exp(\beta'z_{3i})+1}{-\exp(\beta'z_{3j})} \log(Q_i) \end{cases} + \beta'z_{2i} + \log(\eta_2) + \log(\eta_1) \tag{31}
\]

\[
\log P(X > x_i, Y > y_i) = \begin{cases} \frac{1}{-\exp(\beta'z_{3i})} \log(Q_i) \end{cases} \tag{32}
\]

After combining terms and summing over the four groups, the log likelihood equation for the Clayton-Oakes Model with Weibull marginal distributions is:

\[
\log L = \sum_{i \in E_1} \left\{ \log \left\{ \exp(\beta'z_{3i})+1 \right\} - 2 \log(Q_i) \right\} + \\
\sum_{j \in E_1, E_2} \left\{ \beta'z_{1j} + (\eta_1-1) \log(x_j) + x_j^{\eta_1} \exp(\beta'z_{1j} + \beta'z_{3j}) + \log(\eta_1) \right\} + \\
\sum_{j \in E_1, E_3} \left\{ \beta'z_{2j} + (\eta_2-1) \log(y_j) + y_j^{\eta_2} \exp(\beta'z_{2j} + \beta'z_{3j}) + \log(\eta_2) \right\} - \\
\sum_{j \in E_3} \log(Q_j) - \sum_{j \in E_1, E_2, E_3, E_4} \exp(-\beta'z_{3j}) \log(Q_j) \tag{33}
\]

It is useful to have expressions for the first and second derivatives of the likelihood function to use iterative methods for obtaining the maximum likelihood estimates. These derivatives are written in Appendix A.
A LIKELIHOOD RATIO TEST FOR INDEPENDENCE

The purpose of this section is to develop a formal statistical test of the following question: Does knowledge about the age at first detection of hypertension provide any information about the age at which a person might develop cardiovascular disease?

To construct a likelihood ratio test for independence, first obtain the maximized log likelihood of the Clayton-Oakes that includes parameters of association ($LL(b_{\text{full}})$). Obtain the maximum log likelihood for the model of independence ($LL(b_{\text{reduced}})$). A Chi-squared statistic with degrees of freedom equal to the number of association parameters in the full model ($L$) tests for independence.

$$\chi^2_{d.f.:L} = 2\{LL(b_{\text{full}})-LL(b_{\text{reduced}})\}$$ (34)

Reject the null hypothesis of independence for large values of the statistic which exceed the critical table chi-square values with $L$ degrees of freedom.

Writing the likelihood for the model of independence is straightforward. The four components of the likelihood are as follows:

$$P(X=x_i,Y=y_i) = f(x_i,y_i) = f_0(x_i)\cdot g_0(y_i) =$$
$$\lambda_1 \eta_1 x_i^{-\eta_1-1} \lambda_2 \eta_2 y_i^{-\eta_2-1} \exp(-\lambda_1 x_i^{\eta_1})\exp(-\lambda_2 y_i^{\eta_2})$$ (35)

$$P(X=x_i,Y>y_i) = f_0(x_i)\cdot G_0(y_i) = \lambda_1 \eta_1 x_i^{-\eta_1-1} \exp(-\lambda_1 x_i^{\eta_1})\exp(-\lambda_2 y_i^{\eta_2})$$ (36)

$$P(X>x_i,Y=y_i) = F_0(x_i)\cdot g_0(y_i) = \lambda_2 \eta_2 y_i^{-\eta_2-1} \exp(-\lambda_1 x_i^{\eta_1})\exp(-\lambda_2 y_i^{\eta_2})$$ (37)

$$P(X>x_i,Y>y_i) = S(x_i,y_i) = F_0(x_i)G_0(y_i) = \exp(-\lambda_1 x_i^{\eta_1})\exp(-\lambda_2 y_i^{\eta_2})$$ (38)

Express the Weibull scale parameters as functions of covariates (as in Equation 17 and Equation 18), then take logarithms and sum over
the appropriate groups to obtain the log likelihood.

$$\log L = \sum_{i \in s_1, s_2} \{ \beta'_1 z_{1i} + (\eta_1 - 1) \log(x_i) \} + \sum_{i \in s_3, s_4} \{ \beta'_2 z_{2i} + (\eta_2 - 1) \log(y_i) \} - $$

$$\sum_{i \in s_1, s_2, s_3, s_4} \{ x_i^{\eta_1} \exp(\beta'_1 z_{1i}) + y_i^{\eta_2} \exp(\beta'_2 z_{2i}) \} + $$

$$(N1+N2) \log(\eta_1) + (N1+N3) \log(\eta_2) \quad (39)$$

In practice, the value for the log likelihood under independence is obtained by fitting univariate Weibull regression models separately for the two time-to-event variables. The sum of the log likelihoods of the two independently fitted models is the log likelihood of the model under independence. For the sake of completeness, the details of the univariate Weibull regression model are included here. First consider the Weibull survival function (Equation 40) and the Weibull density (Equation 41).

$$P(X > x_i) = S(x_i) = \exp(-\lambda x_i^\gamma) \quad (40)$$

$$P(X=x_i) = f(x_i) = \lambda \eta x_i^{\gamma-1} \exp(-\lambda x_i^\gamma) \quad (41)$$

This binomial type of likelihood is written as:

$$L = \prod_{i=1}^N \exp(-\lambda x_i^\gamma) \{ \lambda \eta x_i^{\gamma-1} \}^{\delta_i} \quad (42)$$

where $\delta$ is an indicator equal to 1 for deaths and equal to 0 for censored observation times. Introduce covariates through the scale parameter.

$$\lambda = \exp(\beta' z_i) \quad (43)$$

Now the log likelihood can be written.

$$\log L = \sum_{i=1}^N -\exp(\beta' z_i) x_i^\gamma + \sum_{i \in d} \beta' z_i + (\eta-1) \sum_{i \in d} \log(x_i) + D \log(\eta) \quad (44)$$
where \(i \in d\) indicates that the observed time is a death time and \(D\) is the number of deaths. The derivatives necessary for maximum likelihood estimation are as follows:

\[
\frac{\partial \log L}{\partial \beta_j} = -\sum_{i=1}^{N} z_{ij} x_i^\eta \exp(\beta' z_i) + \sum_{i \in d} z_{ij} \tag{45}
\]

\[
\frac{\partial \log L}{\partial \eta} = -\sum_{i=1}^{N} x_i^\eta \exp(\beta' z_i) \log(x_i) + \sum_{i \in d} \log(x_i) + \frac{D}{\eta} \tag{46}
\]

\[
\frac{\partial^2 \log L}{\partial \beta_j \partial \beta_j} = -\sum_{i=1}^{N} z_{ij}^2 x_i^\eta \exp(\beta' z_i) \tag{47}
\]

\[
\frac{\partial^2 \log L}{\partial \eta^2} = -\sum_{i=1}^{N} x_i^\eta \exp(\beta' z_i) (\log x_i)^2 - \frac{D}{\eta^2} \tag{48}
\]

\[
\frac{\partial^2 \log L}{\partial \beta_j \partial \eta} = -\sum_{i=1}^{N} z_{ij} x_i^\eta \exp(\beta' z_i) \log(x_i) \tag{49}
\]
In this section, a method is described for predicting the expected (or mean) age of cardiovascular disease, using information about a person's risk factor profile and age at first detection of hypertension, based on the estimates obtained by the methods of the previous sections.

\[ \mu_{x|y} = \int_0^\infty P(X > x | Y = y) \, dx \tag{50} \]

The conditional probability is available from quantities already specified.

\[ P(X > x | Y = y) = \frac{P(X > x, Y = y)}{P(Y = y)} \tag{51} \]

The numerator of the right hand expression above is Equation 31 and the denominator is the Weibull density function for \( Y \) (Equation 16). The resulting probability expression is:

\[ P(X > x | Y = y) = Q \left\{ \frac{\exp(\beta' y z_3) + 1}{\exp(\beta' y z_3)} \right\} \exp\left\{ y^\prime y \exp(\beta' z_2) \{ \exp(\beta' z_3) + 1 \} \right\} \tag{52} \]

Where \( Q \) is identical to Equation 24. The conditional expectation can now be written in the following form.

\[ \mu_{x|y} = \exp\left\{ y^\prime y \exp(\beta' z_2) \{ \exp(\beta' z_3) + 1 \} \right\} \int_0^\infty \frac{\exp(\beta' y z_3) + 1}{\exp(\beta' y z_3)} \, dx \tag{53} \]

Because there is no apparent closed form expression for the above integral, the solution is obtained by numerical integration. A finite upper bound for the integral is chosen (120 years of age) such that the probability of living beyond this age is very small.
CHAPTER III
A DESCRIPTION OF THE LEE-KLEIN MODEL

In this chapter, an alternative method is considered for the estimation of the correlation between the age at first evidence of hypertension and the age at first cardiovascular disease. The Lee-Klein Model differs from the Clayton-Oakes Model in the mechanisms by which the dependence is generated between the age at first detection of hypertension and the age at first cardiovascular disease. However, the correlation structure of these two models is identical. The intention is to compare these two approaches for the purpose of determining which model is superior.

The first section of this chapter is parallel to the first section of Chapter II for the development of a likelihood function. Next, a likelihood ratio test for independence is considered. The third section of this chapter describes a method based on the Lee-Klein Model for obtaining the expected (or mean) age of cardiovascular disease in a spirit similar to the last section of Chapter II.
FORMULATION OF THE LIKELIHOOD FUNCTION

To write the bivariate survival function for the Lee-Klein Model it is necessary to specify the cumulative hazard functions. Cumulative hazard functions of the Weibull form (Equation 54 and Equation 55) are used in the Lee-Klein Model.

\[ H_0(x_i) = \lambda_1 x_i^{\eta_1} \quad \lambda_1, \eta_1 > 0 \] (54)

\[ H_0(y_i) = \lambda_2 y_i^{\eta_2} \quad \lambda_2, \eta_2 > 0 \] (55)

The parameter \( \alpha \) is a measure of dependence which is related to Kendall's tau coefficient of concordance (Equation 56).

\[ \tau = \frac{1}{1+2\alpha} \] (56)

Notice that \( 0 < \alpha < \infty \). Also notice the relationship between \( \alpha \) and \( \tau \). As \( \alpha \to \infty \), \( \tau \to 0 \) and, \( \alpha \to 0 \), \( \tau \to 1 \). Lee and Klein have shown that the dependence structure of this model is identical to the dependence structure of the Clayton-Oakes Model (Equation 57 and Equation 58).

\[ \theta = 1 + \frac{1}{\alpha} \] (57)

\[ \alpha = \frac{1}{\theta - 1} \] (58)

The joint survival function can be written:

\[ S(x, y) = \{ 1 + H_0(x) + H_0(y) \}^{-\alpha} \] (59)

In general, the derivatives necessary to calculate the likelihood function may be written.

\[ -\frac{\partial S(x, y)}{\partial x} = \alpha \{ 1 + H_0(x) + H_0(y) \}^{-\alpha-1} \left[ \frac{\partial H_0(x)}{\partial x} \right] \] (60)
Express the Weibull scale parameters as functions of covariates (as in Equation 17 and Equation 18). In order to allow for the possible inclusion of covariate effects on the degree of association $\alpha$ is re-expressed as in Equation 63.

$$\alpha = \exp(\beta' z_3_i)$$

It is useful to express the estimate of Kendall's tau in terms of the estimated coefficient (Equation 64).

$$\tau = \frac{1}{1+2\exp(\beta' z_3_i)}$$

$z_{1i}$, $z_{2i}$, and $z_{3i}$ are row vectors of covariates defined as in the Clayton-Oakes Model. The groups $g_1$, $g_2$, $g_3$, and $g_4$ define the censoring groups as before and $N_1$, $N_2$, $N_3$, and $N_4$ are the numbers of observations in the respective groups.

To simplify the notation let

$$R_i = \{ 1 + x_i^{n_1}\exp(\beta'_1 z_{1i}) + y_i^{n_2}\exp(\beta'_2 z_{2i}) \}$$

Notice that $R_i$, unlike $Q_i$ in the Clayton-Oakes Model, does not contain any components of the association parameter $\alpha$. This will simplify the expressions for the first and second derivatives considerably.
Now, the four components of the likelihood may be written.

\[ P(X=x_i, Y=y_i) = \exp(\beta'_{\tau} z_{\tau_i}) \{ \exp(\beta'_{\tau} z_{\tau_i} + 1) \} R_i^{-\exp(\beta'_{\tau} z_{\tau_i})+2} \]

\[ \eta_1 x_i^{-\eta_1 - 1} \exp(\beta'_{\tau} z_{\tau_i}) \eta_2 y_i^{-\eta_2 - 1} \exp(\beta'_{\tau} z_{\tau_i}) \] \hspace{1cm} (66)

\[ P(X=x_i, Y>y_i) = \exp(\beta'_{\tau} z_{\tau_i}) R_i^{-\exp(\beta'_{\tau} z_{\tau_i})+1} \eta_1 x_i^{-\eta_1} \exp(\beta'_{\tau} z_{\tau_i}) \] \hspace{1cm} (67)

\[ P(X>x_i, Y=y_i) = \exp(\beta'_{\tau} z_{\tau_i}) R_i^{-\exp(\beta'_{\tau} z_{\tau_i})+1} \eta_2 y_i^{-\eta_2} \exp(\beta'_{\tau} z_{\tau_i}) \] \hspace{1cm} (68)

\[ P(X>x_i, Y>y_i) = R_i^{-\exp(\beta'_{\tau} z_{\tau_i})} \] \hspace{1cm} (69)

Taking logarithms of the four components simplifies the expressions and permits writing the log likelihood function.

\[ \log P(X=x_i, Y=y_i) = \beta'_{\tau} z_{\tau_i} + \log \{ \exp(\beta'_{\tau} z_{\tau_i} + 1) \} - \{ \exp(\beta'_{\tau} z_{\tau_i} + 2) \} \log(R_i) + \]

\[ \beta'_{\tau} z_{\tau_i} + \beta'_{\tau} z_{\tau_i} + \log(\eta_1) + \log(\eta_2) + \]

\[ (\eta_1 - 1) \log(x_i) + (\eta_2 - 1) \log(y_i) \] \hspace{1cm} (70)

\[ \log P(X=x_i, Y>y_i) = \beta'_{\tau} z_{\tau_i} - \{ \exp(\beta'_{\tau} z_{\tau_i} + 1) \} \log(R_i) + \]

\[ \beta'_{\tau} z_{\tau_i} + \log(\eta_1) + (\eta_1 - 1) \log(x_i) \] \hspace{1cm} (71)

\[ \log P(X>x_i, Y=y_i) = \beta'_{\tau} z_{\tau_i} - \{ \exp(\beta'_{\tau} z_{\tau_i} + 1) \} \log(R_i) + \]

\[ \beta'_{\tau} z_{\tau_i} + \log(\eta_2) + (\eta_2 - 1) \log(y_i) \] \hspace{1cm} (72)

\[ \log P(X>x_i, Y>y_i) = -\exp(\beta'_{\tau} z_{\tau_i}) \log(R_i) \] \hspace{1cm} (73)

After combining terms and summing over the four groups, the log likelihood equation for the Lee-Klein Model is:
\[
\log L = \sum_{i \in S_1} \{ \log \{ \exp(\beta'_1 z_{3i})+1 \} - 2 \log(R_i) \} + \\
\sum_{i \in S_1, S_2} \{ \beta'_1 z_{1i} + (\eta_i - 1) \log(x_i) + \log(\eta_i) \} + \\
\sum_{i \in S_1, S_3} \{ \beta'_2 z_{2i} + (\eta_i - 1) \log(y_i) + \log(\eta_i) \} - \\
\sum_{i \in S_2, S_3} \log(R_i) + \sum_{i \in S_1, S_2, S_3} \beta'_3 z_{3i} - \sum_{i \in S_1, S_2, S_3, S_4} \exp(\beta'_3 z_{3i}) \log(R_i)
\] (74)

It is useful to have expressions for the first and second derivatives of the likelihood function to use iterative methods for obtaining the maximum likelihood estimates. These derivatives are listed in Appendix B.

**A LIKELIHOOD RATIO TEST FOR INDEPENDENCE**

The likelihood ratio test of independence for the Lee-Klein Model is conducted in the same manner as the likelihood ratio test of independence for the Clayton-Oakes Model. Where \((LL(b_{full}))\) is the maximized log likelihood from the Lee-Klein Model to be used in Equation 34.
CONDITONAL EXPECTATION

A closed form expression of the conditional expectation based on model II can be obtained from results on the multivariate Burr's distribution (Takahasi, 1965). First note that Equation 66 is the density function of a bivariate Burr's distribution. Takahasi shows that a conditional distribution of a bivariate Burr's distribution is also a Burr's distribution with parameters related to the bivariate distribution.

\[ f(x | y) = f \left( x, \eta_1, \lambda = \frac{\lambda_1}{\lambda_1 + \lambda_2 x^{\eta_2}}, \alpha = \alpha + 1 \right) \]  \hspace{1cm} (75)

Based on the expression for the expectation of a univariate Burr's distribution, the conditional expectation can be written.

\[ \mu = \left[ \frac{1 + \lambda_2 y^{\eta_2}}{\lambda_1} \right]^{\frac{1}{\eta_2} \Gamma \left( \frac{1}{\eta_2} + 1 \right) \Gamma \left( \frac{\alpha + 1}{\eta_2} - \frac{1}{\eta_2} \right)}{\Gamma \left( \alpha + 1 \right)} \]  \hspace{1cm} (76)

Where \( \Gamma \) designates the Gamma function and \( \lambda_1, \lambda_2 \) and \( \alpha \) may be functions of covariates.
CHAPTER IV
DEFINITIONS AND PRELIMINARY ANALYSES

EXCLUSION CRITERIA

Table 1 shows the number of Framingham Study participants who were excluded from this analysis and the reasons for exclusion. The exclusion criteria were applied in a stepwise fashion. A total of 678 study participants were excluded from this analysis. This left a total of 4531 individuals included in the analysis.

A DEFINITION OF CARDIOVASCULAR DISEASE

The criteria used to identify those individuals with evidence of cardiovascular disease are outlined in Table 2. Angina is the most frequent and myocardial infarction is the second most frequent cardiovascular disease event. The severity of the first cardiovascular disease event ranges from the relatively mild angina to the extreme of sudden death.

This broad definition of cardiovascular disease has some implications for the estimation of covariate effects in the subsequent analyses. The approach used here assumes that the covariate effects are the same for coronary heart disease and other forms of cardiovascular disease.
Table 1

*Numbers of Framingham Participants Excluded from Analysis and Reasons for Exclusion.*

<table>
<thead>
<tr>
<th>Number Excluded</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>139</td>
<td>Pre-existing Cardiovascular Disease at examination 1.</td>
</tr>
<tr>
<td>29</td>
<td>Missing information on smoking at both examinations 1 and 4.</td>
</tr>
<tr>
<td>464</td>
<td>Hypertensive by sphygmomanometer measurement at examination 1.</td>
</tr>
<tr>
<td>21</td>
<td>Self-reported Anti-hypertensive medication use at examination 4.</td>
</tr>
<tr>
<td>16</td>
<td>Missing information on serum cholesterol at all examinations.</td>
</tr>
<tr>
<td>6</td>
<td>Missing height information at examination 1.</td>
</tr>
<tr>
<td>3</td>
<td>Missing weight information at examination 1.</td>
</tr>
</tbody>
</table>
The Framingham Heart Study is exceptional in the extensive follow-up regarding cardiovascular disease events. There has been a strong effort to obtain complete information about the occurrence of cardiovascular events among participants in the study by hospital surveillance and follow-up with family physicians and relatives. This investigation assumes that all individuals without cardiovascular disease events and alive at examination 10 were free of disease at examination 10. For these persons, age at last follow-up was calculated by determining age at examination 10.

After determining the exam at which the first cardiovascular disease event occurred, it is necessary to approximate the age of the subject at the time of the event. The age in months of each subject is known on December, 1947 and at the first examination. For examinations 2 through 10 the age at examination is known in whole years. Although on average, the interval between successive exams was 2 years, age at examination provides information specific to the individual. Assuming the subject is halfway through the interval between two birthdays on examination day, an additional half year is added to the whole number value of age at examination.

A more refined estimate of the age at which the first cardiovascular disease event occurred may be obtained by assuming that the event occurred at the midpoint of the interval preceding the examination which records the event. Consider the following example: A subject is reported to have experienced the first cardiovascular disease event at examination 5. To the recorded age at examination 4, the
value of one half a year is added. To this result is added half the number of years between examination 4 and examination 5. This method was used to determine the age at which an event occurred from information about the examination at which the event was recorded.
Table 2

*Type and Frequency of Cardiovascular Events.*

<table>
<thead>
<tr>
<th>Type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Angina Pectoris.</td>
<td>213</td>
</tr>
<tr>
<td>2. Myocardial Infarction.</td>
<td>167</td>
</tr>
<tr>
<td>3. Cerebrovascular Accident.</td>
<td>74</td>
</tr>
<tr>
<td>4. Intermittent Claudication.</td>
<td>74</td>
</tr>
<tr>
<td>5. Congestive Heart Failure.</td>
<td>60</td>
</tr>
<tr>
<td>6. Sudden Death: Coronary Heart Disease.</td>
<td>50</td>
</tr>
<tr>
<td>7. Non-sudden Death: Coronary Heart Disease.</td>
<td>38</td>
</tr>
<tr>
<td>8. Coronary Insufficiency.</td>
<td>25</td>
</tr>
<tr>
<td>9. Death: Cerebrovascular Accident.</td>
<td>17</td>
</tr>
<tr>
<td>10. Death: Other Cardiovascular Disease.</td>
<td>9</td>
</tr>
</tbody>
</table>

Total                                   727
A DEFINITION OF HYPERTENSION

Two criteria were used in determining hypertensive status. The first criterion is based on actual blood pressure measurements taken during the biennial physical examinations which were part of the Framingham Heart Study. The second criterion is based on self-reported use of anti-hypertensive medication as recorded during the interview portion of the biennial examinations. The earliest examination at which either of these criteria were satisfied was used to determine the age at first detection of hypertension.

At each biennial examination, two separate blood pressure measurements were taken by mercury sphygmomanometer. Definite hypertension was indicated by systolic pressure of at least 160 mm Hg and/or diastolic pressure of at least 95 mm Hg on both measurements. The age at first detection of hypertension was calculated by identifying the first of two successive biennial examinations with elevated blood pressure. Requiring two successive biennial examinations reduces the risk of misclassifying normotensives as hypertensive. During examinations 4 through 10, subjects were who were not identified as hypertensive by the above criterion, were asked if they were currently taking anti-hypertensive drugs. The examination at which a subject gave an affirmative answer to this question was used in determining the age at first detection of hypertension. By employing information about the use of anti-hypertensive medication, the chance of misclassifying hypertensives with controlled blood pressure (as normotensive) is decreased.
The definition of hypertension used in this dissertation is similar to a recent study of hypertension incidence (Dannenberg, A. L., et al., 1988). The definition employed here is more restrictive in requiring two successive biennial examinations with elevated blood pressure measurements. An additional attempt has been made in this study to identify a cohort of individuals who are disease-free at the start of the study.

By emphasizing "specificity" in this definition of hypertension it is possible that this classification scheme compromises "sensitivity". Consequently, an unknown number of individuals may be misclassified normotensives (truly hypertensive but fail to meet the strict criteria for classification as hypertensives). If this misclassification occurred more frequently among people who developed cardiovascular disease, it is possible that the effect of hypertension as a risk factor for cardiovascular disease is underestimated.

Another possible effect of this "strict" definition of hypertension might be to estimate the age of onset of hypertension at an older age than the actual age of hypertension onset. All hypertensives would be equally subject to such a bias which might have implications for the conditional expectation curves to be presented later. The consequence of identifying the age at onset of hypertension at a later time that the true age at onset of hypertension will be considered in Chapter VII.

For the remaining individuals in the Framingham Heart Study for whom hypertension was never demonstrated, it is necessary to determine the examination at which blood pressure measurements were last taken.
Unlike the situation with cardiovascular disease, the blood pressure status of all individuals is not known with certainty at examination 10. If death or a cardiovascular disease event (without any indication of hypertension) was recorded at a time later than the last examination at which blood pressure measurements were last taken, the later examination was used in calculating the age at last follow-up.

After determining the examination at which hypertension was detected or last follow-up, the age of the individual is then calculated in the same manner as the age of the first cardiovascular disease event.

Table 3 describes nine types of individuals in the Framingham Heart Study based on the temporal relationship of cardiovascular disease and the age at first detection of hypertension. In regard to cardiovascular disease, recognize that because of the intensive surveillance and follow-up, persons without evidence of cardiovascular disease are identified as disease free at examination 10.

Individuals who were initially enrolled in the study may not return for additional blood pressure measurements and their hypertension status is not completely known at the end of the study. There is a large number of people determined to be normotensive at some time before the end of the study.

The temporal relationship between the age at first detection of hypertension and age at first evidence of cardiovascular disease accounts for the discrepancy between the number of hypertensives in the preliminary analyses and the bivariate methods in Chapters V, VI, and VII. In the preliminary analyses of this chapter (Tables 3 through 11),
it is only appropriate to define hypertension status in a prospective manner. An individual is considered to be hypertensive only if their hypertension is detected before (or simultaneously with) any evidence of cardiovascular disease. The term 'prior hypertension' has been used to refer to this group. There are a total of 1087 (the sum of the groups 1, 2, 3, and 4 in Table 3) individuals with hypertension preceding cardiovascular disease.

Bivariate survival methods estimate the relationship between age at first cardiovascular disease and the age at first hypertension detection. The bivariate methods consider an additional group of individuals (group 5 in Table 3, N=105) in whom hypertension is detected at some time after the first evidence of cardiovascular disease. In the bivariate analyses, there are 1192 (the sum of groups 1, 2, 3, 4 and 5 in Table 3) people designated as hypertensives.
### Table 3

*Frequency of Various Temporal Relationships of First Cardiovascular Disease Event and First Detection of Hypertension.*

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension Precedes Cardiovascular Disease.</td>
<td>145</td>
</tr>
<tr>
<td>Hypertension Detected and First Cardiovascular Event at the same Examination.</td>
<td>67</td>
</tr>
<tr>
<td>Hypertension Precedes Disease Free at Last Follow-up.</td>
<td>866</td>
</tr>
<tr>
<td>Hypertension Detected and Free of Disease at Examination 10.</td>
<td>9</td>
</tr>
<tr>
<td>First Cardiovascular Event Precedes First Detection of Hypertension.</td>
<td>105</td>
</tr>
<tr>
<td>First Cardiovascular Event Precedes Normotensive at Last Follow-up.</td>
<td>257</td>
</tr>
<tr>
<td>Normotensive and First Cardiovascular Event at the Same Examination.</td>
<td>153</td>
</tr>
<tr>
<td>Normotensive Based on an Examination before Examination 10.</td>
<td>2680</td>
</tr>
<tr>
<td>Normotensive and Free of Cardiovascular Disease at Examination 10.</td>
<td>249</td>
</tr>
</tbody>
</table>
ANTI-HYPERTENSIVE MEDICATION USE

Self-reported use of anti-hypertensive medication during examination 4 through 10 was used to identify those hypertensive persons receiving pharmacologic treatment. Those individuals who had experienced a cardiovascular disease event were included in the medication group only if the use of medication preceded the cardiovascular disease event. Among the 1087 hypertensives, 742 (68.26%) were identified as having used anti-hypertensive medication. Table 4 shows the frequency of cardiovascular disease events among medicated and non-medicated hypertensives, and normotensives. The rate of cardiovascular disease is 14.95% among normotensives, increasing to 16.04% among hypertensives on medication and to 26.96% among hypertensives not on medication.

In the examination of smoking, gender, serum cholesterol and body mass index group the frequency of cardiovascular disease is examined for each of the risk factors among both normotensives and hypertensives (Tables 5, 6, 7 and 8). The frequency tables are followed by a logistic regression procedure for the purpose of testing the differences of proportions observed in these tables.

Kaplan-Meier survival curves (Figures 1, 2, 3 and 4) which illustrate the probability of remaining free of cardiovascular disease over age are also presented for these risk factors. Because the trends illustrated by The Kaplan-Meier survival curves do not differ by hypertensive status, the curves presented are based on both normotensives and hypertensives combined.
Table 4

*Frequency of Cardiovascular Disease Among Normotensives, Hypertensives on Medication and Hypertensives Not on Medication.*

Percentage Developing CVD and Free of CVD.

<table>
<thead>
<tr>
<th></th>
<th>WITH CVD</th>
<th>FREE OF CVD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMOTENSIVES</td>
<td>515</td>
<td>2929</td>
<td>3444</td>
</tr>
<tr>
<td></td>
<td>14.95%</td>
<td>85.05%</td>
<td>76.01%</td>
</tr>
<tr>
<td>HYPERTENSIVES ON MEDICATION</td>
<td>119</td>
<td>623</td>
<td>742</td>
</tr>
<tr>
<td></td>
<td>16.04%</td>
<td>83.96%</td>
<td>16.38%</td>
</tr>
<tr>
<td>HYPERTENSIVES NOT ON MEDICATION</td>
<td>93</td>
<td>252</td>
<td>345</td>
</tr>
<tr>
<td></td>
<td>26.96%</td>
<td>73.04%</td>
<td>7.61%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>727</td>
<td>3804</td>
<td>4531</td>
</tr>
<tr>
<td></td>
<td>16.05%</td>
<td>83.95%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
SMOKING

Individuals were classified as smokers or nonsmokers based on self-reported cigarette smoking on examination 1 and, because of a change in the interviewing question, cigarette or pipe smoking on examination 4. The smoking information was not coded at examinations 2 or 3. The frequency of cardiovascular disease events categorized by smoking and hypertension is presented in Table 5. Among normotensives, 12.83% of the nonsmokers and 16.47% of the smokers experienced cardiovascular disease. Among hypertensives, 18.78% of the nonsmokers and 20.27% of the smokers experienced cardiovascular disease.

A very clear effect of smoking on the probability of developing cardiovascular disease is illustrated by Kaplan-Meier curves (Figure 1). Smokers experience a distinct disadvantage or greater probability of cardiovascular disease at all ages older than 45 years of age.
Table 5

*Frequency of Cardiovascular Disease by Smoking Status Among Normotensives and Hypertensives.*

Percentage Developing CVD and Free of CVD.

<table>
<thead>
<tr>
<th></th>
<th>WITH CVD</th>
<th>FREE OF CVD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMOTENSIVE</td>
<td>184</td>
<td>1250</td>
<td>1434</td>
</tr>
<tr>
<td>NONSMOKERS</td>
<td>12.83%</td>
<td>87.17%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>331</td>
<td>1679</td>
<td>2010</td>
</tr>
<tr>
<td>SMOKERS</td>
<td>16.47%</td>
<td>83.53%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>515</td>
<td>2929</td>
<td>3444</td>
</tr>
<tr>
<td></td>
<td>14.95%</td>
<td>85.05%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>WITH CVD</th>
<th>FREE OF CVD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYPERTENSIVE</td>
<td>105</td>
<td>454</td>
<td>559</td>
</tr>
<tr>
<td>NONSMOKERS</td>
<td>18.78%</td>
<td>81.22%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>107</td>
<td>421</td>
<td>528</td>
</tr>
<tr>
<td>SMOKERS</td>
<td>20.27%</td>
<td>79.73%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>212</td>
<td>875</td>
<td>1087</td>
</tr>
<tr>
<td></td>
<td>19.50%</td>
<td>80.50%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
Figure 1: Probability of Survival Free of Cardiovascular Disease by Smoking Status.
GENDER

Because cardiovascular disease rates differ for men and women, a covariate was included to control for this difference. Table 6 presents the frequency of cardiovascular disease events categorized by gender and hypertension. Among the normotensives, 8.9% of the women and 21.57% of the men experienced cardiovascular disease. Among the hypertensives, 16.27% of the women and 25.53% of the men experienced cardiovascular disease.

Kaplan-Meier curves (Figure 2) demonstrate the increased frequency of cardiovascular disease among men. The gender differential is evident at an early age and increases steadily.
Table 6

**Frequency of Cardiovascular Disease by Gender Among Normotensives and Hypertensives.**

Percentage Developing CVD and Free of CVD.

<table>
<thead>
<tr>
<th></th>
<th>WITH CVD</th>
<th>FREE OF CVD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMOTENSIVE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEMALES</td>
<td>160</td>
<td>1638</td>
<td>1798</td>
</tr>
<tr>
<td></td>
<td>8.90%</td>
<td>91.10%</td>
<td>52.21%</td>
</tr>
<tr>
<td>NORMOTENSIVE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALES</td>
<td>355</td>
<td>1291</td>
<td>1646</td>
</tr>
<tr>
<td></td>
<td>21.57%</td>
<td>78.43%</td>
<td>47.79%</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>515</td>
<td>2929</td>
<td>3444</td>
</tr>
<tr>
<td></td>
<td>14.95%</td>
<td>85.05%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>WITH CVD</th>
<th>FREE OF CVD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYPERTENSIVE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEMALES</td>
<td>115</td>
<td>592</td>
<td>707</td>
</tr>
<tr>
<td></td>
<td>16.27%</td>
<td>83.73%</td>
<td>65.04%</td>
</tr>
<tr>
<td>HYPERTENSIVE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALES</td>
<td>97</td>
<td>283</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>25.53%</td>
<td>74.47%</td>
<td>34.96%</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>212</td>
<td>875</td>
<td>1087</td>
</tr>
<tr>
<td></td>
<td>19.50%</td>
<td>80.50%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
Figure 2: Probability of Survival Free of Cardiovascular Disease by Gender.
SERUM CHOLESTEROL

Serum cholesterol determinations in mg/100 ml were performed for a majority of the study participants at each examination. Individuals with serum cholesterol determinations above 275 mg/100 ml on any two examinations were defined to be in the high risk category. For those individuals who experienced a cardiovascular disease event, the two cholesterol determinations of at least 275 mg/100 ml must have occurred prior to the cardiovascular disease event. The value of 275 mg/100 ml was chosen to obtain approximately the upper tercile of all study participants.

Table 7 presents the frequency of cardiovascular disease events categorized by cholesterol group and hypertension status. Among normotensives, 13.96% of the low cholesterol group and 17.61% of the high cholesterol group experienced cardiovascular disease. Among hypertensives, 18.16% of the low cholesterol group and 21.78% of the high cholesterol group experienced cardiovascular disease.

Examination of Kaplan-Meier curves to assess the effect of elevated serum cholesterol (Figure 3) suggests that the probability of cardiovascular disease is not affected by serum cholesterol. The two curves are almost indistinguishable at all ages. The higher frequency of cardiovascular disease among the high cholesterol groups shown in Table 7 becomes unimportant after controlling for the age at which cardiovascular disease occurred.
Table 7

*Frequency of Cardiovascular Disease by Serum Cholesterol Status Among Normotensives and Hypertensives.*

Percentage Developing CVD and Free of CVD.

<table>
<thead>
<tr>
<th></th>
<th>WITH CVD</th>
<th>FREE OF CVD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMOTENSIVES</td>
<td>350</td>
<td>2157</td>
<td>2507</td>
</tr>
<tr>
<td>WITH LOW</td>
<td>13.96%</td>
<td>86.04%</td>
<td></td>
</tr>
<tr>
<td>CHOLESTEROL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NORMOTENSIVES</td>
<td>165</td>
<td>772</td>
<td>937</td>
</tr>
<tr>
<td>WITH HIGH</td>
<td>17.61%</td>
<td>82.39%</td>
<td></td>
</tr>
<tr>
<td>CHOLESTEROL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>515</td>
<td>2929</td>
<td>3444</td>
</tr>
<tr>
<td></td>
<td>14.95%</td>
<td>85.05%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
Figure 3: Probability of Survival Free of Cardiovascular Disease by Serum Cholesterol Status.
BODY MASS INDEX

Quetelet's Index was calculated as weight in kilograms divided by the squared value of height in centimeters. A standardized score (mean of zero, standard deviation of 1) was then calculated for men and women separately based on Quetelet's Index. Finally, an individual was assigned to the high body mass group if the standardized score was in the upper quartile.

Table 8 shows the frequency of cardiovascular disease events in the high and not high body mass groups for normotensives and hypertensives. Among normotensives, 13.50% of the not high body mass group and 20.95% of the high body mass group experienced cardiovascular disease. Among hypertensives, 17.92% of the not high body mass group and 22.90% of the high body mass group experienced cardiovascular disease.

Changes in the probability of cardiovascular disease with increasing age for the high and not high body mass groups are illustrated by Kaplan-Meier survival curves (Figure 4). A higher probability of cardiovascular disease is associated with the high body mass group.
Table 8

*Frequency of Cardiovascular Disease by Body Mass Index Group Among Normotensives and Hypertensives.*

Percentage Developing CVD and Free of CVD.

<table>
<thead>
<tr>
<th>WITH CVD</th>
<th>FREE OF CVD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMOTENSIVES</td>
<td>374</td>
<td>2397</td>
</tr>
<tr>
<td>WITH LOW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BODY MASS</td>
<td>13.50%</td>
<td>86.50%</td>
</tr>
<tr>
<td>NORMOTENSIVES</td>
<td>141</td>
<td>532</td>
</tr>
<tr>
<td>WITH HIGH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BODY MASS</td>
<td>20.95%</td>
<td>79.05%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>515</td>
<td>2929</td>
</tr>
<tr>
<td></td>
<td>14.95%</td>
<td>85.05%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WITH CVD</th>
<th>FREE OF CVD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYPERTENSIVES</td>
<td>133</td>
<td>609</td>
</tr>
<tr>
<td>WITH LOW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BODY MASS</td>
<td>17.92%</td>
<td>82.08%</td>
</tr>
<tr>
<td>HYPERTENSIVES</td>
<td>79</td>
<td>266</td>
</tr>
<tr>
<td>WITH HIGH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BODY MASS</td>
<td>22.90%</td>
<td>77.10%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>212</td>
<td>875</td>
</tr>
<tr>
<td></td>
<td>19.50%</td>
<td>80.50%</td>
</tr>
</tbody>
</table>
**Figure 4:** Probability of Survival Free of Cardiovascular Disease by Body Mass Index Group.
LOGISTIC REGRESSION ANALYSES

In order to test for differences of the percentages observed in Tables 4 through 8, logistic regression analyses were conducted. All of these two variable regressions included a dichotomous indicator variable which takes on the value of 1 for individuals with hypertension prior to cardiovascular disease, and 0 otherwise. The coefficients listed in Table 9 show the effect of the risk factor based on an analysis that controls for hypertension status prior to cardiovascular disease.

The results demonstrate statistically significant effects for all the risk factors defined in the previous sections of this Chapter. The risk factors listed in Table 9 and used in the subsequent analyses were all introduced as dichotomous indicator variables. The variable name listed in the tables corresponds to the group assigned the value of 1 (for example, smokers are assigned the value of 1). With a dichotomous indicator of hypertension status, the medication variable in Table 9 is an estimate of the effects of anti-hypertensive medication relative to the group of hypertensives not on medication.

These logistic regression analyses use no information about the age at which cardiovascular disease was first manifest in the Framingham Study participants. Information about the age of hypertension detection is not utilized in this approach.
Table 9

*Logistic Regression: Two Variable Analyses Controlling for Hypertension.*

Estimated Effects of Risk Factors on the Probability of Cardiovascular Disease Controlling for Prior Hypertension Status (N = 4531).

Two Variable Models Including Prior Hypertension Status Coded as a Dichotomous Indicator Variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.6586</td>
<td>0.1573</td>
<td>-4.19**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.2339</td>
<td>0.0830</td>
<td>2.82**</td>
</tr>
<tr>
<td>Male</td>
<td>0.8959</td>
<td>0.0844</td>
<td>10.61**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.2608</td>
<td>0.0862</td>
<td>3.03**</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.4576</td>
<td>0.0909</td>
<td>5.03**</td>
</tr>
</tbody>
</table>

(** = p-value < 0.0001)
COX REGRESSION ANALYSES

Results

Analyses based on the Cox proportional hazards model are presented in Tables 10 through 12. These analyses are presented to provide a basis for comparison with the bivariate models presented later. It is useful to examine the results of a Cox regression analysis because it is the standard survival analysis technique which would be used in the absence of a bivariate methodology. A time-dependent covariate may be used to incorporate the age at first detection of hypertension in Cox regression.

The analyses are conducted in two steps. First, in order to investigate the effects of covariates separately while still controlling for hypertension, the results of two variable models are presented. Second, a multivariate analysis is presented to investigate the covariate effects while controlling for all other risk factors.

As described in the section entitled A Definition of Hypertension, Cox regression requires that hypertension status is defined in a prospective manner. A person is considered hypertensive if they are first detected to be hypertensive prior to developing cardiovascular disease. Individuals who develop hypertension subsequent to cardiovascular disease (105 such individuals in the Framingham Heart Study, see Table 3) are considered to be normotensive for these analyses. Because of the variability in blood pressure measurement, it is likely that a large proportion of these 105 people are truly hypertensive but misclassified as normotensive. The consequences of misclassification will
be considered in the discussion section which follows. Other individuals, who never experienced cardiovascular disease by examination 10, were included in the hypertensive group if they were determined to be hypertensive at any examination.

Three methods were used to control for hypertensive status in the Cox regression analyses. First, a dichotomous indicator variable was used to identify hypertensives. Second, a time-dependent covariate was used to introduce the age at first detection of hypertension. Finally, Cox regression analyses were conducted on a subset of the Framingham data consisting entirely of hypertensives. In this last approach, attempts to use a time-dependent covariate to represent the age at first detection of hypertension were unsuccessful. For reasons related to the high degree of correlation between the time-dependent covariate and the dependent variable, the maximum likelihood estimates failed to converge.

The motivation for the analyses based only on the hypertensives is to closely examine the coefficient estimates for anti-hypertensive medication. Since medication is an attribute only of hypertensives, any analysis with both hypertensives and normotensives must adequately control for hypertensive status to obtain valid estimates of medication effects.

The Medication variable is statistically significant in all of the Cox regression analyses (Tables 10 through 12). The estimated coefficients are negative in all of the Cox regression analyses indicating a survival advantage for hypertensives on anti-hypertensive medication as compared to hypertensives not receiving medication.
Compare the coefficient estimates for Medication in the analyses which introduce hypertension as a dichotomous indicator variable with the analyses based on hypertensives only. Compare the Medication coefficients in Table 10, top to Table 12, top (two variable models) and compare Table 11, top to Table 12, bottom (multivariate models). In both the two variable models and the multivariate models, the coefficient estimates differ only slightly. This suggests adequate control of hypertension status in the analysis based on both hypertensives and normotensives (N=4531) and that the estimates of the medication effects are valid using this approach.

When the age at first detection of hypertension is introduced as a time-dependent covariate, the effect of anti-hypertensive medication is greater (with a larger Z-statistic) than when hypertension status is introduced as a dichotomous indicator variable. This pattern is consistent for both the two variable analyses and the multivariate analysis. This result suggests that the effect of medication may be underestimated as a consequence of failing to incorporate precise information about hypertension.

The coefficients estimated for Smoking and Male are statistically significant in all of the analyses and demonstrate the expected effects of these risk factors. Smokers and men are at higher risk of cardiovascular disease.

The coefficient estimates for High Cholesterol are not significant in the two variable analyses nor in the analyses limited to hypertensives only. After controlling for the effects of the other risk factors in the
multivariate analysis with a dichotomous indicator variable for hypertension, *High Cholesterol* is shown to be a significant risk factor for cardiovascular disease. In the multivariate analysis with the age at first detection of hypertension introduced as a time-dependent covariate, *High Cholesterol* approaches significance.

The coefficient estimates for *High Body Mass* are statistically significant in the two variable analysis with a dichotomous indicator of hypertension and in the multivariate analyses. All of the coefficient estimates are positive for *High Body Mass* indicating a higher risk of cardiovascular disease among this group when controlling for the effects of other risk factors.

**Discussion**

The Cox proportional hazards model uses only information about covariates which precede the onset of cardiovascular disease. As a consequence, individuals who develop hypertension subsequent to cardiovascular disease will not be accounted for in estimating the relationship between hypertension and cardiovascular disease. If these 105 individuals are actually misclassified as normotensives, the result would be to underestimate the coefficients for hypertension in Table 11.

There is another type of study participant who is not adequately represented in an analysis based on Cox regression. Consider an individual who is last observed to be normotensive, at examination 3. This person does not return for any subsequent examinations and hypertensive status is unknown beyond examination 3. However, as a result of surveillance efforts at examination 10, this person is
determined to be alive and free of cardiovascular disease. A Cox regression analysis assumes that this person is normotensive throughout the study. A bivariate analysis recognizes that the age of this person at examination 3 represents the maximum observed age at which the person is still normotensive (a censored observation).

The bivariate survival analysis methods offer the additional advantage of providing an estimate of the degree of dependence between the age at first detection of hypertension and the age at first evidence of cardiovascular disease. The interpretation of this dependence measure (as Kendall's tau or as a hazard ratio) is distinctly different from the results available from the Cox proportional hazard model with a time dependent covariate for the age at first detection of hypertension.
Table 10

*Cox Regression: Two Variable Analyses Controlling for Hypertension.*

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease Controlling for Hypertension (N = 4531).

Two Variable Models Including Prior Hypertension Status Coded as a Dichotomous Indicator Variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.4349</td>
<td>0.1385</td>
<td>-3.06**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.5768</td>
<td>0.0767</td>
<td>7.47**</td>
</tr>
<tr>
<td>Male</td>
<td>0.8151</td>
<td>0.0774</td>
<td>10.51**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>-0.0004</td>
<td>0.0784</td>
<td>-0.01</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.1959</td>
<td>0.0814</td>
<td>5.03**</td>
</tr>
</tbody>
</table>

Age at First Detection of Hypertension Coded as a Time Dependent Covariate.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.8634</td>
<td>0.1144</td>
<td>-7.55**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.5896</td>
<td>0.0766</td>
<td>7.69**</td>
</tr>
<tr>
<td>Male</td>
<td>0.8638</td>
<td>0.0770</td>
<td>11.27**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>-0.0505</td>
<td>0.0783</td>
<td>-0.64</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.1501</td>
<td>0.0811</td>
<td>1.85</td>
</tr>
</tbody>
</table>

(** = p-value < 0.0001)
### Cox Regression: Multivariate Analysis Controlling for Hypertension.

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease Controlling for Hypertension (N = 4531).

#### Multivariate Analysis Including Prior Hypertension Status Coded as a Dichotomous Indicator Variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.3222</td>
<td>0.1389</td>
<td>-2.32**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.4352</td>
<td>0.0799</td>
<td>5.45**</td>
</tr>
<tr>
<td>Male</td>
<td>0.7330</td>
<td>0.0815</td>
<td>8.99**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.1845</td>
<td>0.0811</td>
<td>2.30**</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2815</td>
<td>0.0821</td>
<td>3.43**</td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.3346</td>
<td>0.1136</td>
<td>2.95**</td>
</tr>
</tbody>
</table>

#### Multivariate Analysis Including Prior Age at First Detection of Hypertension as a Time Dependent Covariate.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.7282</td>
<td>0.1157</td>
<td>-6.30**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.4215</td>
<td>0.0798</td>
<td>5.29**</td>
</tr>
<tr>
<td>Male</td>
<td>0.7335</td>
<td>0.0813</td>
<td>9.03**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.1456</td>
<td>0.0799</td>
<td>1.82</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2495</td>
<td>0.0820</td>
<td>3.04**</td>
</tr>
<tr>
<td>Hypertension</td>
<td>1.3111</td>
<td>0.0965</td>
<td>13.58**</td>
</tr>
</tbody>
</table>

(** = p-value < 0.0001)
Table 12

*Cox Regression, Hypertensives: Two Variable and Multivariate Analysis for Individuals with Prior Hypertension (N=1087).*

Estimated Effects of Risk Factors on the Age of Cardiovascular Disease.

**Univariate Analysis.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.4154</td>
<td>0.1387</td>
<td>-2.99**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.4847</td>
<td>0.1383</td>
<td>3.50**</td>
</tr>
<tr>
<td>Male</td>
<td>0.6592</td>
<td>0.1385</td>
<td>4.76**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>-0.0759</td>
<td>0.1396</td>
<td>-0.54</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.1560</td>
<td>0.1422</td>
<td>1.10</td>
</tr>
</tbody>
</table>

**Multivariate Analysis**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.3444</td>
<td>0.1397</td>
<td>-2.47**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.3592</td>
<td>0.1439</td>
<td>2.50**</td>
</tr>
<tr>
<td>Male</td>
<td>0.5489</td>
<td>0.1478</td>
<td>3.71**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.0338</td>
<td>0.1423</td>
<td>0.24</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2423</td>
<td>0.1434</td>
<td>1.69</td>
</tr>
</tbody>
</table>

(** = p-value < 0.0001)
CHAPTER V

CLAYTON-OAKES MODEL ANALYSES

RESULTS AND DISCUSSION

The bivariate models investigated in this chapter and Chapter VI utilize covariates which are introduced into the Weibull scale parameter related to the age at onset of cardiovascular disease. The generalized description of the models in Chapters II and III allows for the introduction of covariates in either of the Weibull scale parameters or the dependence parameter. Covariates for this application are introduced in a manner consistent with developing the notion of the conditional expectation of the age at onset of cardiovascular disease.

The results of models using single covariates are presented in Table 13. The large and highly significant coefficient for the Medication variable demonstrates the beneficial effect of anti-hypertension medication on the age at first cardiovascular disease, in the presence of the age at first detection of hypertension. The positive and highly significant effects of Smoking and Male indicates the deleterious effects of these risk factors on the age of first cardiovascular disease. High Cholesterol and High Body Mass are not statistically significant in this univariate analysis, although the Z-statistic for High Body Mass is approaching significance.
Combining all risk factors into a multivariate analysis (Table 14) allows a determination of the effects of one of the risk factors controlling for all the other effects. The effects and trends in this multivariate analysis are similar to the results of the univariate investigation with the exception of High Cholesterol and High Body Mass. A small positive and statistically significant coefficient is estimated for the effect of elevated serum cholesterol. The coefficient estimate for High Body Mass is highly statistically significant in this analysis.

Two significant interaction effects were found and the results of the analysis including these interaction effects are presented in Table 15. The Smoking by Male interaction describes the differential effects of smoking for men and women. The consequences of smoking for men are more severe than the consequences of smoking for women. The Medication by Male interaction describes the differential effects of anti-hypertensive medication for men and women. The benefit derived from anti-hypertensive medication is greater for men than for women.

The Chi-square statistic for the test of independence between the age at first detection of hypertension and the age at first cardiovascular disease is very large. The value of the log likelihood under the model of independence is -10087.98. Consequently, the hypothesis of independence is rejected even after accounting for the effects of five risk factors and two interaction terms.

After controlling for the five risk factors and two interaction terms, the bivariate correlation between the age at first detection of
hypertension and the age of first cardiovascular disease is large. The estimate of Kendall's tau is 0.4424 with a 95% confidence limits of (0.3988, 0.4869). A correlation estimate of this magnitude indicates that considerable information is to be gained by using the age at first detection of hypertension to predict the age at first cardiovascular disease.

The dependence parameter is also a hazard ratio estimate. Based on the multivariate analysis with interactions, the ratio of the hazard for cardiovascular disease among hypertensives divided by the hazard for cardiovascular disease among normotensives is 2.59 with a 95% confidence limits of (2.33, 2.90).

Age, which is usually the strongest predictor of cardiovascular disease in any analysis, enters this analysis through the age at first detection of hypertension and is tied to one of the outcomes of interest, the age at onset of cardiovascular disease. In order to verify that the hazard ratio estimate is not just a measure of the risk associated with age, an auxilliary analysis was conducted identical to that in Table 15 except that age at the start of the study was included as an additonal covariate. The estimate of the dependence parameter in this analysis was not significantly different from the estimate in Table 15. The age coefficient in this auxilliary analysis was statistically significant but, negative (contrary to the usual notion of increasing risk of cardiovascular disease with age) suggesting multicollinearity effects.

Conditional expectation is investigated to assess and illustrate the effects of risk factors on the expected age of first cardiovascular
disease. Figures 5, 6, 7 and 8 are obtained by using the parameters estimated from the multivariate analysis with interactions (Table 15). Nineteen values of the age at first detection of hypertension between age 35 and 80, spaced at 2 and one-half year intervals, were used to obtain nineteen values for each combination of risk factors. A spline routine is then used to draw the conditional expectation curves.

The curvilinear relationship between the age at first detection of hypertension and the age at first cardiovascular disease is identical in each of the curves. The effects of the risk factors is to shift the curves along the vertical axis. At younger ages of the first detection of hypertension, the curve is almost flat. A ten year change in the age at first detection of hypertension from age 45 to age 35 results in a delay of one-half year in the expected age at cardiovascular disease. At older ages of the first detection of hypertension, the slope of the curve is steeper. A ten year change in the age at first detection of hypertension from age 75 to age 65 results in a delay of about six years in the expected age at cardiovascular disease.

Figure 5 shows that the benefit of anti-hypertensive medication for men consists of delaying the onset of cardiovascular disease by approximately ten and one-half years. For women, the benefit of anti-hypertensive medication is measured in terms of delaying the onset of cardiovascular disease by approximately five and one-half years.

Figure 6 illustrates the effect of smoking for men and women. Smoking women tend to experience cardiovascular disease about one-half year earlier than nonsmoking women. For men, smoking tends to lead
to cardiovascular disease approximately four years earlier than for nonsmoking men.

Elevated serum cholesterol leads to an earlier onset of cardiovascular disease of about one year as illustrated in Figure 7. The effect of a high body mass is to promote the onset of cardiovascular disease by about two years as illustrated in Figure 8.

In Tables 16 and 17, the results of the application of the Clayton-Oakes Model to the subset of hypertensive individuals is presented. As in the preliminary analyses using Cox regression, the purpose of these restricted analyses is to compare the estimates for the Medication variable with the results obtained when normotensives are included in the analysis.

In the analyses limited to hypertensives, anti-hypertensive medication has a slightly stronger effect with greater protection from cardiovascular disease although, the associated Z-statistics are slightly smaller. The similarity of the Medication coefficients in these two analytic approaches indicates that hypertension status is adequately controlled in the analysis including both normotensives and hypertensives.
Table 13

*Clayton-Oakes Model: Univariate Analysis.*

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease in a Bivariate Analysis which includes the Age at First Detection of Hypertension.

\[ N = 4531 \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.9691</td>
<td>0.0989</td>
<td>-9.80**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.5663</td>
<td>0.0733</td>
<td>7.72**</td>
</tr>
<tr>
<td>Male</td>
<td>0.8731</td>
<td>0.0722</td>
<td>12.09**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>-0.0601</td>
<td>0.0745</td>
<td>-0.81</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.1070</td>
<td>0.0771</td>
<td>1.39</td>
</tr>
</tbody>
</table>

(** = p-value < 0.0001)
Table 14

*Clayton-Oakes Model: Multivariate Analysis.*

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease in a Bivariate Analysis which includes the Age at First Detection of Hypertension.

\[
\text{log likelihood} = -9987.33 \quad N = 4531
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension Shape</td>
<td>6.1166</td>
<td>0.1367</td>
<td>----</td>
</tr>
<tr>
<td>CVD Shape parameter</td>
<td>7.5551</td>
<td>0.1999</td>
<td>----</td>
</tr>
<tr>
<td>Hypertension Intercept</td>
<td>-26.4386</td>
<td>0.5707</td>
<td>----</td>
</tr>
<tr>
<td>CVD Intercept</td>
<td>-33.5499</td>
<td>0.8507</td>
<td>----</td>
</tr>
<tr>
<td>Medication</td>
<td>-0.8354</td>
<td>0.1004</td>
<td>-8.32**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.3155</td>
<td>0.0713</td>
<td>4.43**</td>
</tr>
<tr>
<td>Male</td>
<td>0.7069</td>
<td>0.0731</td>
<td>9.67**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.1383</td>
<td>0.0706</td>
<td>1.96*</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2131</td>
<td>0.0724</td>
<td>2.94**</td>
</tr>
<tr>
<td>Dependence parameter</td>
<td>0.4630</td>
<td>0.0907</td>
<td>----</td>
</tr>
</tbody>
</table>

\(\text{(** = p-value < 0.0001, * = p-value < 0.05)}\)
Table 15

*Clayton-Oakes Model: Multivariate Analysis with Interactions.*

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease in a Bivariate Analysis which includes the Age at First Detection of Hypertension.

\[
\text{log likelihood} = -9977.75 \quad \text{N} = 4531
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension Shape parameter</td>
<td>6.1149</td>
<td>0.1367</td>
<td>----</td>
</tr>
<tr>
<td>CVD Shape parameter</td>
<td>7.5525</td>
<td>0.2000</td>
<td>----</td>
</tr>
<tr>
<td>Hypertension Intercept</td>
<td>-26.4316</td>
<td>0.5706</td>
<td>----</td>
</tr>
<tr>
<td>CVD Intercept</td>
<td>-33.4805</td>
<td>0.8529</td>
<td>----</td>
</tr>
<tr>
<td>Medication</td>
<td>-0.5926</td>
<td>0.1300</td>
<td>-4.56**</td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.0299</td>
<td>0.1179</td>
<td>-0.25</td>
</tr>
<tr>
<td>Male</td>
<td>0.5450</td>
<td>0.1130</td>
<td>4.82**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.1238</td>
<td>0.0709</td>
<td>1.75</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2178</td>
<td>0.0726</td>
<td>3.00**</td>
</tr>
<tr>
<td>Smok*Male</td>
<td>0.4750</td>
<td>0.1494</td>
<td>3.18**</td>
</tr>
<tr>
<td>Med.*Male</td>
<td>-0.5591</td>
<td>0.1993</td>
<td>-2.80**</td>
</tr>
<tr>
<td>Dependence parameter</td>
<td>0.4617</td>
<td>0.0914</td>
<td>----</td>
</tr>
<tr>
<td>(\tau=0.4424)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chi-square Test of Independence for Age of First Detection of Hypertension and Age of First Cardiovascular Disease.

\[
\chi^2 = 220.46**
\]

\((** = p-value < 0.0001)\)
Figure 5: Clayton-Oakes Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, Gender and Anti-hypertensive Medication Status.
Figure 6: Clayton-Oakes Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, Gender and Smoking Status.
Figure 7: Clayton-Oakes Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, and Serum Cholesterol Status.
Figure 8: Clayton-Oakes Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, and Body Mass Index Group.
Table 16

Clayton-Oakes Model, Hypertensives: Univariate and Multivariate Analyses for Individuals Ever Hypertensive (N=1192).

Estimated Effects of Risk Factors on the Age of Cardiovascular Disease.

Univariate Analysis Controlling for Age at First Detection of Hypertension.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-1.0095</td>
<td>0.1130</td>
<td>-8.93**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.5331</td>
<td>0.1112</td>
<td>4.79**</td>
</tr>
<tr>
<td>Male</td>
<td>0.7375</td>
<td>0.1101</td>
<td>6.69**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>-0.0780</td>
<td>0.1110</td>
<td>-0.70</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2136</td>
<td>0.1117</td>
<td>1.91</td>
</tr>
</tbody>
</table>

Multivariate Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.9239</td>
<td>0.1135</td>
<td>-8.14**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.3874</td>
<td>0.1138</td>
<td>3.40**</td>
</tr>
<tr>
<td>Male</td>
<td>0.5310</td>
<td>0.1157</td>
<td>4.59**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.0309</td>
<td>0.1122</td>
<td>0.27</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2940</td>
<td>0.1120</td>
<td>2.63**</td>
</tr>
</tbody>
</table>

(** = p-value < 0.0001)
Table 17

Clayton-Oakes Model, Hypertensives: Multivariate Analysis with Interactions for Individuals Ever Hypertensive (N=1192).

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease.

Multivariate Analysis with Interactions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.6960</td>
<td>0.1544</td>
<td>-4.51**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.1406</td>
<td>0.1624</td>
<td>0.87</td>
</tr>
<tr>
<td>Male</td>
<td>0.4014</td>
<td>0.1992</td>
<td>2.01**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>-0.0023</td>
<td>0.1118</td>
<td>-0.02</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.3044</td>
<td>0.1116</td>
<td>2.73**</td>
</tr>
<tr>
<td>Smok*Male</td>
<td>0.5499</td>
<td>0.2368</td>
<td>2.32**</td>
</tr>
<tr>
<td>Med.*Male</td>
<td>-0.5260</td>
<td>0.2354</td>
<td>-2.23**</td>
</tr>
</tbody>
</table>

(** = p-value < 0.0001)
RESULTS AND DISCUSSION

The results of analyses using the Lee-Klein Model are presented in Tables 18, 19 and 20. The significance tests for the coefficients of risk factors are similar to the results seen in the Clayton-Oakes Models with the exception of 'High Cholesterol. In the Clayton-Oakes multivariate analysis, High Cholesterol is significant, whereas in the Lee-Klein Model the Z-statistic is smaller.

These two models differ in regard to the dependence estimates. The bivariate correlation between the age at first detection of hypertension and the age at cardiovascular disease in terms of Kendall's tau is 0.5282 with 95% confidence limits of (0.4893, 0.5668). This confidence interval lies above the confidence interval for Kendall's tau estimated in the Clayton-Oakes Model. As in the Clayton-Oakes Model, this estimate of dependence from the Lee-Klein Model is based on the multivariate analysis with interaction terms which takes into account the effects of risk factors.

The estimate of the ratio of the hazard of cardiovascular disease among hypertensives divided by the hazard of cardiovascular disease among normotensives is 3.24 with 95% confidence limits of (2.92, 3.62).
This hazard ratio confidence interval also lies above the confidence interval of the hazard ratio estimated in the Clayton-Oakes Model.

The conditional expectation curves presented here are based on the coefficients estimated in the multivariate analysis with interactions (Table 20) and were generated using the same methods as Figures 5, 6, 7, and 8. The general shape of the curves based on the Lee-Klein Model is similar to the shape of curves based on the Clayton-Oakes Model.

Some general trends are evident in comparisons of the curves based on the Clayton-Oakes Models and curves based on the Lee-Klein Models. At earlier ages of the first detection of hypertension, the expected age at first cardiovascular disease is earlier for the Lee-Klein Model. At older ages of the first detection of hypertension, the expected age of cardiovascular disease is greater in the Lee-Klein Model. When the age of first detection of hypertension is in the range of about age 60 to age 65, the two models produce similar estimate of the expected age of cardiovascular disease.

At older ages of the first detection of hypertension, the curves based on the Lee-Klein Model tend to be steeper than curves based on the Clayton-Oakes Model. A ten year change in the age at first detection of hypertension from age 75 to age 65 results in a delay of about seven years in the expected age at cardiovascular disease (compare to six years in the Clayton-Oakes Model).

Figure 9 shows that the benefit of anti-hypertensive medication for men consists of delaying the onset of cardiovascular disease by
approximately twelve and one-half years. For women, the benefit of anti-hypertensive medication is measured in terms of delaying the onset of cardiovascular disease by approximately seven years.

Figure 10 illustrates the effects of smoking for men and women. Smoking women tend to experience cardiovascular disease almost one year earlier than nonsmoking women. For men, smoking tends to lead to cardiovascular disease five years earlier than for nonsmoking men. The effects of elevated serum cholesterol and high body mass (Figures 11 and 12) are nearly the same as the results observed in the Clayton-Oakes Model.

In Tables 21 and 22, the results of the application of the Lee-Klein Model to the subset of hypertensive individuals is presented. Comparing the analyses limited to hypertensives with the analyses based on all the data, anti-hypertensive medication has a slightly weaker effect and the associated Z-statistics are slightly smaller. These trends are different from the relationships observed in the Clayton-Oakes Model.
Table 18

_Lee-Klein Model: Univariate Analysis_.

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease in a Bivariate Analysis which includes the Age at First Detection of Hypertension.

\[ N = 4531 \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-1.5480</td>
<td>0.1295</td>
<td>-11.96**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.7301</td>
<td>0.0867</td>
<td>8.42**</td>
</tr>
<tr>
<td>Male</td>
<td>1.1171</td>
<td>0.0891</td>
<td>12.54**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>-0.0770</td>
<td>0.0877</td>
<td>-0.88</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.1253</td>
<td>0.0914</td>
<td>1.37</td>
</tr>
</tbody>
</table>

\[ (** = p-value < 0.0001) \]
Table 19

*Lee-Klein Model: Multivariate Analysis.*

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease in a Bivariate Analysis which includes the Age at First Detection of Hypertension.

\[ \text{log likelihood} = -9926.82 \quad \text{N} = 4531 \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension Shape</td>
<td>8.0251</td>
<td>0.2169</td>
<td>----</td>
</tr>
<tr>
<td>CVD Shape parameter</td>
<td>9.9656</td>
<td>0.3152</td>
<td>----</td>
</tr>
<tr>
<td>Hypertension Intercept</td>
<td>-33.0263</td>
<td>0.8327</td>
<td>----</td>
</tr>
<tr>
<td>CVD Intercept</td>
<td>-42.4720</td>
<td>1.2590</td>
<td>----</td>
</tr>
<tr>
<td>Medication</td>
<td>-1.3231</td>
<td>0.1307</td>
<td>-10.12**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.4847</td>
<td>0.0969</td>
<td>5.00**</td>
</tr>
<tr>
<td>Male</td>
<td>0.9204</td>
<td>0.0973</td>
<td>9.46**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.1674</td>
<td>0.0961</td>
<td>1.74</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2481</td>
<td>0.0992</td>
<td>2.50**</td>
</tr>
<tr>
<td>Dependence parameter</td>
<td>-0.7902</td>
<td>0.0800</td>
<td>----</td>
</tr>
</tbody>
</table>

\( (** = \text{p-value} < 0.0001) \)
Table 20

Lee-Klein Model: Multivariate Analysis with Interactions.

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease in a Bivariate Analysis which includes the Age at First Detection of Hypertension.

\[ \log \text{likelihood} = -9912.68 \quad N = 4531 \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension Shape parameter</td>
<td>8.0522</td>
<td>0.2179</td>
<td>----</td>
</tr>
<tr>
<td>CVD Shape parameter</td>
<td>10.0675</td>
<td>0.3196</td>
<td>----</td>
</tr>
<tr>
<td>Hypertension Intercept</td>
<td>-33.1140</td>
<td>0.8365</td>
<td>----</td>
</tr>
<tr>
<td>CVD Intercept</td>
<td>-42.8076</td>
<td>1.2786</td>
<td>----</td>
</tr>
<tr>
<td>Medication</td>
<td>-0.9836</td>
<td>0.1585</td>
<td>-6.21**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.1032</td>
<td>0.1440</td>
<td>0.72</td>
</tr>
<tr>
<td>Male</td>
<td>0.6831</td>
<td>0.1552</td>
<td>4.40**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.1363</td>
<td>0.0967</td>
<td>1.41</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2507</td>
<td>0.0995</td>
<td>2.52**</td>
</tr>
<tr>
<td>Smok*Male</td>
<td>0.7336</td>
<td>0.1963</td>
<td>3.74**</td>
</tr>
<tr>
<td>Med.*Male</td>
<td>-0.8556</td>
<td>0.2397</td>
<td>-3.57**</td>
</tr>
<tr>
<td>Dependence parameter</td>
<td>-0.8061</td>
<td>0.0795</td>
<td>----</td>
</tr>
</tbody>
</table>

Chi-square Test of Independence for Age of First Detection of Hypertension and Age of First Cardiovascular Disease.

\[ \chi^2 = 350.60^{**} \]

\( (** = p\text{-value} < 0.0001) \)
Figure 9: Lee-Klein Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, Gender and Anti-hypertensive Medication Status.
Figure 10: Lee-Klein Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, Gender and Smoking Status.
Figure 11: Lee-Klev Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, and Serum Cholesterol Status.
Figure 12: Lee-Klein Model: Expected Age of Developing Cardiovascular Disease by Age at First Detection of Hypertension, and Body Mass Index Group.
Table 21

Lee-Klein Model, Hypertensives: Univariate and Multivariate Analyses for Individuals Ever Hypertensive (N=1192).

Estimated Effects of Risk Factors on the Age of Cardiovascular Disease.

Univariate Analysis Controlling for Age at First Detection of Hypertension.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-1.2085</td>
<td>0.1261</td>
<td>-9.58**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.5949</td>
<td>0.1211</td>
<td>4.91**</td>
</tr>
<tr>
<td>Male</td>
<td>0.8035</td>
<td>0.1206</td>
<td>6.66**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>-0.0741</td>
<td>0.1217</td>
<td>-0.61</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2359</td>
<td>0.1234</td>
<td>1.91</td>
</tr>
</tbody>
</table>

Multivariate Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-1.0864</td>
<td>0.1275</td>
<td>-8.52**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.4359</td>
<td>0.1294</td>
<td>3.37**</td>
</tr>
<tr>
<td>Male</td>
<td>0.5286</td>
<td>0.1327</td>
<td>3.98**</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.0477</td>
<td>0.1270</td>
<td>0.38</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.2954</td>
<td>0.1266</td>
<td>2.33**</td>
</tr>
</tbody>
</table>

\( (** = p-value < 0.0001)\)
Table 22

Lee-Klein Model, Hypertensives: Multivariate Analysis with Interactions for Individuals Ever Hypertensive (N=1192).

Estimated Effects of Risk Factors on the Age of First Cardiovascular Disease.

Multivariate Analysis with Interactions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>-0.8374</td>
<td>0.1714</td>
<td>-4.88**</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.1777</td>
<td>0.1795</td>
<td>0.99</td>
</tr>
<tr>
<td>Male</td>
<td>0.4257</td>
<td>0.2283</td>
<td>1.86</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>0.0155</td>
<td>0.1273</td>
<td>0.12</td>
</tr>
<tr>
<td>High Body Mass</td>
<td>0.3095</td>
<td>0.1268</td>
<td>2.44**</td>
</tr>
<tr>
<td>Smok'Male</td>
<td>0.5867</td>
<td>0.2673</td>
<td>2.19**</td>
</tr>
<tr>
<td>Med.'Male</td>
<td>-0.5754</td>
<td>0.2611</td>
<td>-2.02**</td>
</tr>
</tbody>
</table>

(** = p-value < 0.0001)
CHAPTER VII
SUMMARY AND CONCLUSIONS

Having estimated coefficients and examined the results from fitting two bivariate survival analysis models to the Framingham Heart Study data, it is of interest to investigate which of the two models is more credible. Since the models have the same number of parameters, they may be compared on the basis of the values of their log likelihoods. The value of the log likelihood in the Lee-Klein Model (Table 20) is -9912.68 which is considerably larger than the value of the log likelihood in the comparable Clayton-Oakes Model (Table 15) of -9977.75. This provides some evidence that the Lee-Klein Model is superior in this application.

There are a few differences in the results derived from the two models, but generally the conclusions to be drawn from both models are in agreement. The Chi-square tests for independence of the age at first detection of hypertension and the age at onset of cardiovascular disease strongly reject the hypothesis of independence in both models. Each of the models provide an estimate of Kendall's tau which may be considered to be a partial correlation coefficient between the age at first detection of hypertension and the age at onset of cardiovascular disease adjusted for the effects of anti-hypertensive medication, smoking, gender,
total serum cholesterol and body mass. Although the scale coefficients estimated in these models are not directly interpretable as risk ratio estimates, the effects of the covariates may be expressed in terms of the expected age at onset of cardiovascular disease.

These two models differ in the estimated effects of covariates on the expected age at onset of cardiovascular disease. The Lee-Klein Model estimates a greater effect of anti-hypertensive medication and smoking on the expected age at onset of cardiovascular disease. For anti-hypertensive medication, the Lee-Klein Model estimates a delayed onset of cardiovascular disease two years longer than the estimates, based on the Clayton-Oakes Model. For smoking, the Lee-Klein Model estimates (by one year) an earlier onset of cardiovascular disease than the estimate based on the Clayton-Oakes Model.

As described in Chapter VI, these two models differ in the estimates of the dependence parameter. The Lee-Klein Model estimates a larger value of Kendall's tau and a larger hazard ratio estimate than the Clayton-Oakes Model. Therefore, the Lee-Klein Model implies that hypertension results in greater risk for cardiovascular disease when compared to the Clayton-Oakes Model.

The Figures which show the relationship between the age at first detection of hypertension and the expected age at onset of cardiovascular disease are different in the two models. Curves based on the Lee-Klein Model suggest that the interval between the age at first detection of hypertension and the expected age of cardiovascular disease is longer at older ages when compared to the Clayton-Oakes Model.
A possible source of bias requires consideration which pertains to the figures showing the relationship between the age at first detection of hypertension and the expected age of cardiovascular disease. It is possible that the biological onset of hypertension begins at an age considerably younger than the age at first detection of hypertension. At the individual level, the number of years that the biological onset of hypertension precedes the first detection of hypertension would be subject to considerable variation which cannot be accounted for in these analyses.

The methods described in this dissertation may be applied to other problems of interest to epidemiologists. In future work, these models will be used to account for a "household" effect among spouse-pairs in The Framingham Heart Study. Rather than a correlation between two separate events in the same individual, these methods will be applied to the age at first detection of hypertension for husband-wife pairs. Another investigation in this same spirit concerns estimating the correlation between the age at retirement from the labor force for husband-wife pairs using the Retirement History Survey data.

Other bivariate survival functions have been described which could be used in the construction of likelihoods applicable to these types of problems. Finally, a conditional likelihood might be adapted for censoring and covariates which might be more suitable for prediction purposes.
APPENDIX A: LOG LIKELIHOOD FUNCTION

DERIVATIVES, CLAYTON-OAKES MODEL

\[
\frac{\partial LL}{\partial \beta_{1j}} = -2 \sum_{i \in \mathcal{E}_1} \frac{\partial \log(Q_i)}{\partial \beta_{1j}} + \sum_{i \in \mathcal{E}_1, \mathcal{E}_2} z_{1,i} \left(1 + x_i^\eta \exp(\beta'_1 z_{1i} + \beta'_3 z_{3i})\right) - 
\sum_{i \in \mathcal{E}_1, \mathcal{E}_3} \frac{\partial \log(Q_i)}{\partial \beta_{1j}} - \sum_{i \in \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_4} \exp(-\beta'_3 z_{3i}) \frac{\partial \log(Q_i)}{\partial \beta_{1j}}
\]

(77)

where,

\[
\frac{\partial \log(Q_i)}{\partial \beta_{1j}} = \frac{z_{1,i} x_i^\eta \exp\{x_i^\eta \exp(\beta'_1 z_{1i} + \beta'_3 z_{3i}) + \beta'_2 z_{2i} + \beta'_3 z_{3i}\}}{Q_i}
\]

(78)

\[
\frac{\partial LL}{\partial \beta_{2i}} = -2 \sum_{i \in \mathcal{E}_1} \frac{\partial \log(Q_i)}{\partial \beta_{2i}} + \sum_{i \in \mathcal{E}_1, \mathcal{E}_3} z_{2,i} \left(1 + x_i^\eta \exp(\beta'_2 z_{2i} + \beta'_3 z_{3i})\right) - 
\sum_{i \in \mathcal{E}_1, \mathcal{E}_3} \frac{\partial \log(Q_i)}{\partial \beta_{2i}} - \sum_{i \in \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_4} \exp(-\beta'_3 z_{3i}) \frac{\partial \log(Q_i)}{\partial \beta_{2i}}
\]

(79)

where,

\[
\frac{\partial \log(Q_i)}{\partial \beta_{2i}} = \frac{z_{2,i} x_i^\eta \exp\{x_i^\eta \exp(\beta'_2 z_{2i} + \beta'_3 z_{3i}) + \beta'_2 z_{2i} + \beta'_3 z_{3i}\}}{Q_i}
\]

(80)

\[
\frac{\partial LL}{\partial \beta_{3i}} = \sum_{i \in \mathcal{E}_1} \frac{z_{3,i} \exp(\beta'_3 z_{3i})}{\exp(\beta'_3 z_{3i}) + 1} - 2 \frac{\partial \log(Q_i)}{\partial \beta_{3i}} + \sum_{i \in \mathcal{E}_1, \mathcal{E}_2} z_{3,i} x_i^\eta \exp(\beta'_1 z_{1i} + \beta'_3 z_{3i}) + 
\sum_{i \in \mathcal{E}_1, \mathcal{E}_3} \frac{\partial \log(Q_i)}{\partial \beta_{3i}} + 
\sum_{i \in \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_4} \exp(-\beta'_3 z_{3i}) \left(\frac{\partial \log(Q_i)}{\partial \beta_{3i}} - z_{3,i} \log(Q_i)\right)
\]

(81)
where,

$$\frac{\partial \log(Q_i)}{\partial \beta_{3i}} = \left\{ \begin{array}{l} x_i^{\eta_i} \exp \left\{ x_i^{\eta_i} \exp (\beta_{1} z_{1i} + \beta_{3} z_{3i}) + \beta_{4} z_{4i} + \beta_{5} z_{5i} \right\} \\ y_i^{\eta_i} \exp \left\{ y_i^{\eta_i} \exp (\beta_{2} z_{2i} + \beta_{3} z_{3i}) + \beta_{4} z_{4i} + \beta_{5} z_{5i} \right\} \end{array} \right\}$$

(82)

$$\frac{\partial LL}{\partial \eta_1} = -2 \sum_{i \in \mathcal{G}_1} \frac{\partial \log(Q_i)}{\partial \eta_1} + \sum_{i \in \mathcal{G}_3} (\log x_i) \left\{ 1 + x_i^{\eta_i} \exp (\beta_{1} z_{1i} + \beta_{3} z_{3i}) \right\} - \sum_{i \in \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4} \frac{\partial \log(Q_i)}{\partial \eta_1} \frac{\partial \log(Q_i)}{\partial \eta_1} + \frac{N_1 + N_2}{\eta_1}$$

(83)

where,

$$\frac{\partial \log(Q_i)}{\partial \eta_1} = \frac{x_i^{\eta_i} (\log x_i) \exp \left\{ x_i^{\eta_i} \exp (\beta_{1} z_{1i} + \beta_{3} z_{3i}) + \beta_{4} z_{4i} + \beta_{5} z_{5i} \right\}}{Q_i}$$

(84)

$$\frac{\partial LL}{\partial \eta_2} = -2 \sum_{i \in \mathcal{G}_1} \frac{\partial \log(Q_i)}{\partial \eta_2} + \sum_{i \in \mathcal{G}_3} (\log y_i) \left\{ 1 + y_i^{\eta_i} \exp (\beta_{2} z_{2i} + \beta_{3} z_{3i}) \right\} - \sum_{i \in \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4} \frac{\partial \log(Q_i)}{\partial \eta_2} \frac{\partial \log(Q_i)}{\partial \eta_2} + \frac{N_1 + N_3}{\eta_2}$$

(85)

where,

$$\frac{\partial \log(Q_i)}{\partial \eta_2} = \frac{y_i^{\eta_i} (\log y_i) \exp \left\{ y_i^{\eta_i} \exp (\beta_{2} z_{2i} + \beta_{3} z_{3i}) + \beta_{4} z_{4i} + \beta_{5} z_{5i} \right\}}{Q_i}$$

(86)

In this list of second derivatives, when \( j = j' \) diagonal elements of the matrix are indicated, otherwise, when \( j \neq j' \) the off-diagonal elements of the matrix are indicated.
\[
\frac{\delta^2 LL}{\delta \beta_{1j} \delta \beta_{1j}} = -2 \sum_{i \in \xi_1} \frac{\delta^2 \log(Q_i)}{\delta \beta_{1j} \delta \beta_{1j}} + \sum_{i \in \xi_1, \xi_2} z_{1i} z_{1j} x_i^\eta \exp(\beta'_1 z_{1i} + \beta'_3 z_{3i}) - \\
\sum_{i \in \xi_2, \xi_3} \frac{\delta^2 \log(Q_i)}{\delta \beta_{1j} \delta \beta_{1j}} - \sum_{i \in \xi_4, \xi_3} \exp(-\beta'_3 z_{3i}) \frac{\delta^2 \log(Q_i)}{\delta \beta_{1j} \delta \beta_{1j}}
\]  

(87)

where,

\[
\frac{\delta^2 \log(Q_i)}{\delta \beta_{1j} \delta \beta_{1j}} = \frac{Q_i \left[ \frac{\delta^2 Q_i}{\delta \beta_{1j} \delta \beta_{1j}} \right] \cdot \left[ \frac{\delta Q_i}{\delta \beta_{1j}} \cdot \frac{\delta Q_i}{\delta \beta_{1j}} \right]}{Q_i^2}
\]  

(88)

and,

\[
\frac{\delta^2 Q_i}{\delta \beta_{1j} \delta \beta_{1j}} \left\{ z_{1i} \{ x_i^\eta \exp(\beta'_1 z_{1i} + \beta'_3 z_{3i}) + 1 \} \right\}
\]  

(89)

\[
\frac{\delta^2 LL}{\delta \beta_{2j} \delta \beta_{2j}} = -2 \sum_{i \in \xi_1} \frac{\delta^2 \log(Q_i)}{\delta \beta_{2j} \delta \beta_{2j}} + \sum_{i \in \xi_1, \xi_3} z_{2i} z_{2j} y_i^\eta \exp(\beta'_2 z_{2i} + \beta'_3 z_{3i}) - \\
\sum_{i \in \xi_2, \xi_3} \frac{\delta^2 \log(Q_i)}{\delta \beta_{2j} \delta \beta_{2j}} - \sum_{i \in \xi_1, \xi_3} \exp(-\beta'_3 z_{3i}) \frac{\delta^2 \log(Q_i)}{\delta \beta_{2j} \delta \beta_{2j}}
\]  

(90)

where,

\[
\frac{\delta^2 \log(Q_i)}{\delta \beta_{2j} \delta \beta_{2j}} = \frac{Q_i \left[ \frac{\delta^2 Q_i}{\delta \beta_{2j} \delta \beta_{2j}} \right] \cdot \left[ \frac{\delta Q_i}{\delta \beta_{2j}} \cdot \frac{\delta Q_i}{\delta \beta_{2j}} \right]}{Q_i^2}
\]  

(91)

and,

\[
\frac{\delta^2 Q_i}{\delta \beta_{2j} \delta \beta_{2j}} \left\{ z_{2i} \{ y_i^\eta \exp(\beta'_2 z_{2i} + \beta'_3 z_{3i}) + 1 \} \right\}
\]  

(92)
\[
\frac{\delta^2 LL}{\partial \beta_3 \partial \beta_{3'}} = \sum_{i \in \mathcal{E}_1} \frac{z_{3i} \exp(\beta' z_{3i})}{\exp(2 \cdot \beta' z_{3i}) + 2 \exp(\beta' z_{3i}) + 1} - 2 \frac{\delta^2 \log(Q_i)}{\partial \beta_3 \partial \beta_{3'}} + \]

\[
\sum_{i \in \mathcal{E}_1 \mathcal{E}_2} z_{3i} \exp(\beta' z_{3i}) x_i \exp(\beta' z_{3i} + \beta' z_{3i}) + \]

\[
\sum_{i \in \mathcal{E}_1 \mathcal{E}_2} z_{3i} \exp(\beta' z_{3i}) - \sum_{i \in \mathcal{E}_3 \mathcal{E}_3} \frac{\delta^2 \log(Q_i)}{\partial \beta_3 \partial \beta_{3'}} - \]

\[
\sum_{i \in \mathcal{E}_1 \mathcal{E}_2 \mathcal{E}_3} \exp(-\beta' z_{3i}) \left\{ \frac{\partial^2 \log(Q_i)}{\partial \beta_3 \partial \beta_{3'}} - z_{3i} \frac{\partial Q_i}{\partial \beta_{3'}} \right\} \quad (93)\]

where,

\[
\frac{\delta^2 \log(Q_i)}{\partial \beta_3 \partial \beta_{3'}} = \frac{Q_i \left\{ \frac{\partial^2 Q_i}{\partial \beta_3 \partial \beta_{3'}} \right\} - \left\{ \frac{\partial Q_i}{\partial \beta_3} \right\} \left\{ \frac{\partial Q_i}{\partial \beta_{3'}} \right\}}{Q_i^2} \quad (94)\]

and,

\[
\frac{\delta^2 Q_i}{\partial \beta_3 \partial \beta_{3'}} = z_{3i} \left\{ y_i \exp \left\{ y_i \exp(\beta' z_{2i} + \beta' z_{3i}) + \beta' z_{2i} + \beta' z_{3i} \right\} \right\} \times \]

\[
\left\{ z_{3i} \left\{ y_i \exp(\beta' z_{2i} + \beta' z_{3i}) + 1 \right\} \right\} + \]

\[
z_{3i} \left\{ x_i \exp \left\{ x_i \exp(\beta' z_{1i} + \beta' z_{3i}) + \beta' z_{1i} + \beta' z_{3i} \right\} \right\} \times \]

\[
z_{3i} \left\{ x_i \exp(\beta' z_{1i} + \beta' z_{3i}) + 1 \right\} \quad (95)\]

\[
\frac{\delta^2 LL}{\partial \eta_i^2} = -2 \sum_{i \in \mathcal{E}_1} \frac{\delta^2 \log(Q_i)}{\partial \eta_i^2} + \sum_{i \in \mathcal{E}_1 \mathcal{E}_2} (\log x_i)^2 x_i \exp(\beta' z_{1i} + \beta' z_{3i}) - \]

\[
\sum_{i \in \mathcal{E}_2 \mathcal{E}_3} \frac{\delta^2 \log(Q_i)}{\partial \eta_i^2} - \sum_{i \in \mathcal{E}_1 \mathcal{E}_2 \mathcal{E}_3 \mathcal{E}_4} \exp(-\beta' z_{3i}) \frac{\delta^2 \log(Q_i)}{\partial \eta_i^2} - \frac{N_1 + N_2}{\eta_i^2} \quad (96)\]
where,

$$\frac{\delta^2 \log (Q_i)}{\delta \eta_1^2} = \frac{Q_i}{\delta \eta_1^2} \left[ \frac{\delta^2 Q_i}{\delta \eta_1^2} \right] - \left[ \frac{\delta Q_i}{\delta \eta_1} \right]^2$$

(97)

and,

$$\frac{\delta^2 Q_i}{\delta \eta_1^2} = x_i \eta_1 (\log x_i)^2 \exp \left\{ \eta_1 \exp (\beta' z_{i1} + \beta' z_{i3}) + 2 \beta' z_{i1} + 2 \beta' z_{i3} \right\} +$$

$$\log (x_i) \left[ \frac{\delta Q_i}{\delta \eta_1} \right]$$

(98)

$$\frac{\delta^2 \mathcal{L}}{\delta \eta_2^2} = -2 \sum_{i \in \mathcal{F}_1} \frac{\delta^2 \log (Q_i)}{\delta \eta_2^2} + \sum_{i \in \mathcal{F}_2} (\log y_i)^2 \eta_2 \exp (\beta' z_{2i} + \beta' z_{3i}) -$$

$$\sum_{i \in \mathcal{F}_2} \frac{\delta^2 \log (Q_i)}{\delta \eta_2^2} - \sum_{i \in \mathcal{F}_2} \exp (-\beta' z_{3i}) \frac{\delta^2 \log (Q_i)}{\delta \eta_2^2} - \frac{N1 + N3}{\eta_2^2}$$

(99)

where,

$$\frac{\delta^2 \log (Q_i)}{\delta \eta_2^2} = \frac{Q_i}{\delta \eta_2^2} \left[ \frac{\delta^2 Q_i}{\delta \eta_2^2} \right] - \left[ \frac{\delta Q_i}{\delta \eta_2} \right]^2$$

(100)

and,

$$\frac{\delta^2 Q_i}{\delta \eta_2^2} = y_i \eta_2 (\log y_i)^2 \exp \left\{ y_i \eta_2 \exp (\beta' z_{2i} + \beta' z_{3i}) + 2 \beta' z_{2i} + 2 \beta' z_{3i} \right\} +$$

$$\log (y_i) \left[ \frac{\delta Q_i}{\delta \eta_2} \right]$$

(101)
\[
\frac{\partial^2 LL}{\partial \beta_1 \partial \beta_2} = -2 \sum_{i \in E_1} \frac{\partial^2 \log(Q_i)}{\partial \beta_1 \partial \beta_2} - \sum_{i \in E_2} \frac{\partial^2 \log(Q_i)}{\partial \beta_1 \partial \beta_2}
\]

where,

\[
\frac{\partial^2 \log(Q_i)}{\partial \beta_1 \partial \beta_2} = \left[ \frac{\partial \log(Q_i)}{\partial \beta_1} \right] \left[ \frac{\partial \log(Q_i)}{\partial \beta_2} \right]
\]

\[
\frac{\partial^2 LL}{\partial \beta_1 \partial \beta_3} = -2 \sum_{i \in E_1} \frac{\partial^2 \log(Q_i)}{\partial \beta_1 \partial \beta_3} + \sum_{i \in E_2} z_{1ji} z_{3ji} x_i^\eta \exp(\beta_1 z_{li} + \beta_3 z_{3ji}) - \\
\sum_{i \in E_3} \exp(-\beta_3 z_{3ji}) \left\{ \frac{\partial^2 \log(Q_i)}{\partial \beta_1 \partial \beta_3} \right\} - z_{3ji} \frac{\partial \log(Q_i)}{\partial \beta_1} \frac{\partial \log(Q_i)}{\partial \beta_3}
\]

where,

\[
\frac{\partial^2 \log(Q_i)}{\partial \beta_1 \partial \beta_3} = \frac{Q_i \left[ \frac{\partial^2 Q_i}{\partial \beta_1 \partial \beta_3} \right] - \left[ \frac{\partial Q_i}{\partial \beta_1} \right] \left[ \frac{\partial Q_i}{\partial \beta_3} \right]}{Q_i^2}
\]

and,

\[
\frac{\partial^2 Q_i}{\partial \beta_1 \partial \beta_3} = \left\{ \frac{\partial Q_i}{\partial \beta_1} \right\} \left\{ \{ z_{3ji} \{ x_i^\eta \exp(\beta_1 z_{li} + \beta_3 z_{3ji}) + 1 \} \right\}
\]

\[
\frac{\partial^2 LL}{\partial \beta_1 \partial \eta_1} = -2 \sum_{i \in E_1} \frac{\partial^2 \log(Q_i)}{\partial \beta_1 \partial \eta_1} + \sum_{i \in E_2} z_{1ji} (\log x_i) x_i^\eta \exp(\beta_1 z_{li} + \beta_3 z_{3ji}) - \\
\sum_{i \in E_3} \exp(-\beta_3 z_{3ji}) \frac{\partial^2 \log(Q_i)}{\partial \beta_1 \partial \eta_1} - \sum_{i \in E_4} \exp(\beta_3 x_{3ji}) \frac{\partial^2 \log(Q_i)}{\partial \beta_1 \partial \eta_1}
\]
where,

\[
\frac{\delta^2 \log(Q_i)}{\delta \beta_{i1} \delta \eta_1} = Q_i \left[ \frac{\delta^2 Q_i}{\partial \beta_{1j} \partial \eta_i} \right] - \frac{\delta Q_i}{\partial \beta_{1j}} \cdot \frac{\delta Q_i}{\partial \eta_i} \tag{108}
\]

and,

\[
\frac{\delta^2 Q_i}{\delta \beta_{1j} \delta \eta_1} = \left[ \frac{\delta Q_i}{\delta \eta_1} \right] \left\{ z_{i1} \{ x_i^{\eta_i} \exp(\beta_1 z_{1i} + \beta_3 z_{3i}) + 1 \} \right\} \tag{109}
\]

\[
\frac{\delta^2 LL}{\delta \beta_{1j} \delta \eta_2} = -2 \sum_{i \in \xi_1} \frac{\delta^2 \log(Q_i)}{\delta \beta_{1j} \delta \eta_2} - \sum_{i \in \xi_2, \xi_3, \xi_4} \frac{\delta^2 \log(Q_i)}{\delta \beta_{1j} \delta \eta_2} - \sum_{i \in \xi_3} \exp(-\beta_3 z_{3i}) \frac{\delta^2 \log(Q_i)}{\delta \beta_{1j} \delta \eta_2} \tag{110}
\]

where,

\[
\frac{\delta^2 \log(Q_i)}{\delta \beta_{1j} \delta \eta_2} = \left[ \frac{\delta \log(Q_i)}{\delta \beta_{1j}} \right] \left[ \frac{\delta \log(Q_i)}{\delta \eta_2} \right] \tag{111}
\]

\[
\frac{\delta^2 LL}{\delta \beta_{2i} \delta \beta_{3i}} = -2 \sum_{i \in \xi_1} \frac{\delta^2 \log(Q_i)}{\delta \beta_{2i} \delta \beta_{3i}} + \sum_{i \in \xi_1, \xi_3} z_{2i} z_{3i} y_i^{\eta_i} \exp(\beta_2 z_{2i} + \beta_3 z_{3i}) - \sum_{i \in \xi_2, \xi_3} \frac{\delta^2 \log(Q_i)}{\delta \beta_{2i} \delta \beta_{3i}} \tag{112}
\]

where,

\[
\frac{\delta^2 \log(Q_i)}{\delta \beta_{2i} \delta \beta_{3i}} = Q_i \left[ \frac{\delta^2 Q_i}{\partial \beta_{2i} \partial \beta_{3i}} \right] - \frac{\delta Q_i}{\partial \beta_{2i}} \cdot \frac{\delta Q_i}{\partial \beta_{3i}} \tag{113}
\]
and,

$$\frac{\delta^2 Q_i}{\delta \beta_{2k} \delta \beta_{3l}} = \left[ \frac{\delta Q_i}{\delta \beta_{2k}} \right] \cdot \left\{ z_{3li} \{ y_i^{\eta_2} \exp(\beta'_{2}z_{2i} + \beta'_{3}z_{3i}) + 1 \} \right\}$$  \hspace{1cm} (114)

$$\frac{\delta^2 L_{LL}}{\delta \beta_{2k} \delta \eta_1} = -2 \sum_{i \in S_1} \frac{\delta^2 \log(Q_i)}{\delta \beta_{2k} \delta \eta_1} - \sum_{i \in S_2, S_3} \frac{\delta^2 \log(Q_i)}{\delta \beta_{2k} \delta \eta_1} - \sum_{i \in S_1, S_2, S_3, S_4} \exp(-\beta'_{3}z_{3i}) \frac{\delta^2 \log(Q_i)}{\delta \beta_{2k} \delta \eta_1}$$  \hspace{1cm} (115)

where,

$$\frac{\delta^2 \log(Q_i)}{\delta \beta_{2k} \delta \eta_1} = - \left[ \frac{\delta \log(Q_i)}{\delta \beta_{2k}} \right] \cdot \left[ \frac{\delta \log(Q_i)}{\delta \eta_1} \right]$$  \hspace{1cm} (116)

$$\frac{\delta^2 L_{LL}}{\delta \beta_{2k} \delta \eta_2} = -2 \sum_{i \in S_1} \frac{\delta^2 \log(Q_i)}{\delta \beta_{2k} \delta \eta_2} + \sum_{i \in S_2, S_3} z_{2ki} (\log x_i) y_i^{\eta_2} \exp(\beta'_{2}z_{2i} + \beta'_{3}z_{3i}) - \sum_{i \in S_1, S_2, S_3, S_4} \exp(-\beta'_{3}z_{3i}) \frac{\delta^2 \log(Q_i)}{\delta \beta_{2k} \delta \eta_2}$$  \hspace{1cm} (117)

where,

$$\frac{\delta^2 \log(Q_i)}{\delta \beta_{2k} \delta \eta_2} = Q_i \left[ \frac{\delta^2 Q_i}{\delta \beta_{2k} \delta \eta_2} \right] - \left[ \frac{\partial Q_i}{\delta \beta_{2k}} \right] \cdot \left[ \frac{\partial Q_i}{\delta \eta_2} \right]$$  \hspace{1cm} (118)

and,

$$\frac{\delta^2 Q_i}{\delta \beta_{2k} \delta \eta_2} = \left\{ \frac{\delta Q_i}{\delta \eta_2} \right\} \cdot \left\{ z_{2ki} \{ y_i^{\eta_2} \exp(\beta'_{2}z_{2i} + \beta'_{3}z_{3i}) + 1 \} \right\}$$  \hspace{1cm} (119)
\[
\frac{\partial^2 LL}{\partial \beta_3 \partial \eta_1} = -2 \sum_{i \in \mathcal{E}_1} \frac{\partial^2 \log(Q_i)}{\partial \beta_3 \partial \eta_1} + \sum_{i \in \mathcal{E}_1, \mathcal{E}_2} \frac{z_{3i}}{\beta_3, \eta_1} \exp(\beta \cdot z_{3i} + \beta \cdot z_{3i}^2) - \\
\sum_{i \in \mathcal{E}_2, \mathcal{E}_3} \frac{\partial^2 \log(Q_i)}{\partial \beta_3 \partial \eta_1} - \\
\sum_{i \in \mathcal{E}_3} \exp(-\beta \cdot z_{3i}^2) \left\{ \frac{\partial^2 \log(Q_i)}{\partial \beta_3 \partial \eta_1} - \beta \cdot z_{3i} \frac{\partial \log(Q_i)}{\partial \eta_1} \right\} 
\]
(120)

where,
\[
\frac{\partial^2 \log(Q_i)}{\partial \beta_3 \partial \eta_1} = Q_i \frac{\partial^2 Q_i}{\partial \beta_3 \partial \eta_1} - \frac{\partial Q_i}{\partial \beta_3} \cdot \frac{\partial Q_i}{\partial \eta_1} 
\]
(121)

and,
\[
\frac{\partial^2 Q_i}{\partial \beta_3 \partial \eta_1} = \{ \frac{\partial Q_i}{\partial \eta_1} \} \{ z_{3i} \left\{ x_i \eta_i \exp(\beta \cdot z_{3i} + \beta \cdot z_{3i}^2) + 1 \right\} \} 
\]
(122)

\[
\frac{\partial^2 LL}{\partial \beta_3 \partial \eta_2} = -2 \sum_{i \in \mathcal{E}_1} \frac{\partial^2 \log(Q_i)}{\partial \beta_3 \partial \eta_2} + \sum_{i \in \mathcal{E}_1, \mathcal{E}_2} \frac{z_{3i}}{\beta_3, \eta_2} \exp(\beta \cdot z_{2i} + \beta \cdot z_{2i}^2) - \\
\sum_{i \in \mathcal{E}_2, \mathcal{E}_3} \frac{\partial^2 \log(Q_i)}{\partial \beta_3 \partial \eta_2} - \\
\sum_{i \in \mathcal{E}_3} \exp(-\beta \cdot z_{3i}^2) \left\{ \frac{\partial^2 \log(Q_i)}{\partial \beta_3 \partial \eta_2} - \beta \cdot z_{3i} \frac{\partial \log(Q_i)}{\partial \eta_2} \right\} 
\]
(123)

where,
\[
\frac{\partial^2 \log(Q_i)}{\partial \beta_3 \partial \eta_2} = Q_i \frac{\partial^2 Q_i}{\partial \beta_3 \partial \eta_2} - \frac{\partial Q_i}{\partial \beta_3} \cdot \frac{\partial Q_i}{\partial \eta_2} 
\]
(124)
and,

\[
\frac{\partial^2 \mathcal{L}}{\partial \eta_1 \partial \eta_2} = -2 \sum_{i \in \mathcal{E}_1} \frac{\partial^2 \log(Q_i)}{\partial \eta_1 \partial \eta_2} - \sum_{i \in \mathcal{E}_2} \frac{\partial^2 \log(Q_i)}{\partial \eta_1 \partial \eta_2} - \\
\sum_{i \in \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4} \exp(-\beta' z'_{3i}) \frac{\partial^2 \log(Q_i)}{\partial \eta_1 \partial \eta_2}
\]  

(126)

where,

\[
\frac{\partial^2 \log(Q_i)}{\partial \eta_1 \partial \eta_2} = - \left\{ \frac{\partial \log(Q_i)}{\partial \eta_1} \right\} \left\{ \frac{\partial \log(Q_i)}{\partial \eta_2} \right\}
\]  

(127)
APPENDIX B: LOG LIKELIHOOD FUNCTION

DERIVATIVES, LEE-KLEIN MODEL

\[
\frac{\partial LL}{\partial \beta_{1,i}} = -2 \sum_{i \in e_1} \frac{\partial \log(R_i)}{\partial \beta_{1,i}} + \sum_{i \in e_1, e_2} z_{1,i} - \sum_{i \in e_1, e_2} \frac{\partial \log(R_i)}{\partial \beta_{1,i}} - \\
\sum_{i \in e_1, e_2, e_3, e_4} \exp(\beta' z_{3,i}) \frac{\partial \log(R_i)}{\partial \beta_{1,i}} \tag{128}
\]

where,

\[
\frac{\partial \log(R_i)}{\partial \beta_{1,i}} = \frac{z_{1,i} y_i^{\eta_i} \exp(\beta' z_{1,i})}{R_i} \tag{129}
\]

\[
\frac{\partial LL}{\partial \beta_{2,i}} = -2 \sum_{i \in e_1} \frac{\partial \log(R_i)}{\partial \beta_{2,i}} + \sum_{i \in e_1, e_3} z_{2,i} - \sum_{i \in e_1, e_3} \frac{\partial \log(R_i)}{\partial \beta_{2,i}} - \\
\sum_{i \in e_1, e_2, e_3, e_4} \exp(\beta' z_{3,i}) \frac{\partial \log(R_i)}{\partial \beta_{2,i}} \tag{130}
\]

where,

\[
\frac{\partial \log(R_i)}{\partial \beta_{2,i}} = \frac{z_{2,i} y_i^{\eta_i} \exp(\beta' z_{2,i})}{R_i} \tag{131}
\]

\[
\frac{\partial LL}{\partial \beta_{3,i}} = \sum_{i \in e_1} \frac{z_{3,i} \exp(\beta' z_{3,i})}{\exp(\beta' z_{3,i}) + 1} + \sum_{i \in e_1, e_2, e_3} z_{3,i} - \\
\sum_{i \in e_1, e_2, e_3, e_4} z_{3,i} \exp(\beta' z_{3,i}) \log(R_i) \tag{132}
\]
\[
\frac{\delta LL}{\delta \eta_1} = -2 \sum_{i \in \xi_1} \frac{\delta \log(R_i)}{\delta \eta_1} + \sum_{i \in \xi_1, \xi_2} \left( \log(x_i) - \sum_{i \in \xi_2, \xi_3, \xi_4} \frac{\delta \log(R_i)}{\delta \eta_1} \right) - \\
\sum_{i \in \xi_1, \xi_2, \xi_3, \xi_4} \exp(\beta' z_i) \frac{\delta \log(R_i)}{\delta \eta_1} + \frac{N_1 + N_2}{\eta_1}
\] (133)

where,
\[
\frac{\delta \log(R_i)}{\delta \eta_1} = \frac{x_i^{\eta_1} \exp(\beta' z_i) \log(x_i)}{R_i}
\] (134)

\[
\frac{\delta LL}{\delta \eta_2} = -2 \sum_{i \in \xi_1} \frac{\delta \log(R_i)}{\delta \eta_2} + \sum_{i \in \xi_1, \xi_3} \left( \log(y_i) - \sum_{i \in \xi_2, \xi_3} \frac{\delta \log(R_i)}{\delta \eta_2} \right) - \\
\sum_{i \in \xi_1, \xi_2, \xi_3, \xi_4} \exp(\beta' z_i) \frac{\delta \log(R_i)}{\delta \eta_2} + \frac{N_1 + N_3}{\eta_2}
\] (135)

where,
\[
\frac{\delta \log(R_i)}{\delta \eta_2} = \frac{y_i^{\eta_2} \exp(\beta' z_i) \log(y_i)}{R_i}
\] (136)

In this list of second derivatives, when \( j = j' \) diagonal elements of the matrix are indicated, otherwise, when \( j \neq j' \) the off-diagonal elements of the matrix are indicated.

\[
\frac{\delta^2 LL}{\delta \beta_1 \delta \beta_{1j}} = -2 \sum_{i \in \xi_1} \frac{\delta^2 \log(R_i)}{\delta \beta_1 \delta \beta_{1j}} - \sum_{i \in \xi_2, \xi_3} \frac{\delta^2 \log(R_i)}{\delta \beta_1 \delta \beta_{1j}} - \\
\sum_{i \in \xi_1, \xi_2, \xi_3, \xi_4} \exp(\beta' z_i) \frac{\delta^2 \log(R_i)}{\delta \beta_1 \delta \beta_{1j}}
\] (137)
where,

\[
\frac{\delta^2 \log(R_i)}{\partial \beta_{1,j} \partial \beta_{1,j}'} = R_i \left[ \left( \frac{\partial^2 R_i}{\partial \beta_{1,j} \partial \beta_{1,j}'} \right) - \left( \frac{\partial R_i}{\partial \beta_{1,j}} \right) \left( \frac{\partial R_i}{\partial \beta_{1,j}'} \right) \right] \frac{1}{R_i^2} \tag{138}
\]

and,

\[
\frac{\delta^2 R_i}{\partial \beta_{1,j} \partial \beta_{1,j}'} = Z_{1,ji} \left[ \frac{\partial R_i}{\partial \beta_{1,j}} \right] \tag{139}
\]

\[
\frac{\delta^2 LL}{\partial \beta_{2k} \partial \beta_{2k}'} = -2 \sum_{i \in \mathcal{E}_1} \frac{\delta^2 \log(R_i)}{\partial \beta_{2k} \partial \beta_{2k}'} - \sum_{i \in \mathcal{E}_2} \exp(\beta_3 z_{3i}) \frac{\delta^2 \log(R_i)}{\partial \beta_{2k} \partial \beta_{2k}'} - \sum_{i \in \mathcal{E}_3} \exp(\beta_3 z_{3i}) \frac{\delta^2 \log(R_i)}{\partial \beta_{2k} \partial \beta_{2k}'} - \sum_{i \in \mathcal{E}_4} \exp(\beta_3 z_{3i}) \frac{\delta^2 \log(R_i)}{\partial \beta_{2k} \partial \beta_{2k}'} \tag{140}
\]

where,

\[
\frac{\delta^2 \log(R_i)}{\partial \beta_{2k} \partial \beta_{2k}'} = R_i \left[ \left( \frac{\partial^2 R_i}{\partial \beta_{2k} \partial \beta_{2k}'} \right) - \left( \frac{\partial R_i}{\partial \beta_{2k}} \right) \left( \frac{\partial R_i}{\partial \beta_{2k}'} \right) \right] \frac{1}{R_i^2} \tag{141}
\]

and,

\[
\frac{\delta^2 R_i}{\partial \beta_{2k} \partial \beta_{2k}'} = Z_{2k} \left[ \frac{\partial R_i}{\partial \beta_{2k}} \right] \tag{142}
\]

\[
\frac{\delta^2 LL}{\partial \beta_3 \partial \beta_3'} = \sum_{i \in \mathcal{E}_1} \frac{Z_{3i}}{\exp(2 \beta_3 z_{3i}) + 2 \exp(\beta_3^2 z_{3i}^2)} - \sum_{i \in \mathcal{E}_2} \frac{Z_{3i}^2}{z_{3i}^2} \exp(\beta_3 z_{3i}) \log(R_i) \tag{143}
\]
\[
\frac{\delta^2 LL}{\partial \eta_1^2} = -2 \sum_{i \in \mathcal{E}_3} \frac{\delta^2 \log(R_i)}{\partial \eta_1^2} - \sum_{i \in \mathcal{E}_3 \times \mathcal{E}_3} \frac{\delta \log^2(R_i)}{\partial \eta_1^2} - \\
\sum_{i \in \mathcal{E}_3 \times \mathcal{E}_3} \exp(\beta_i Z_{3i}) \frac{\delta^2 \log(R_i)}{\partial \eta_1^2} \frac{\delta \log(R_i)}{\partial \eta_1} - \frac{N_1 + N_2}{\eta_1^2}
\]

where,

\[
\frac{\delta^2 \log(R_i)}{\partial \eta_1^2} = R_i \left| \frac{\delta^2 R_i}{\partial \eta_1^2} \right| - \left( \frac{\delta R_i}{\partial \eta_1} \right)^2
\]

(145)

and,

\[
\frac{\delta^2 R_i}{\partial \eta_1^2} = \eta_i \exp(\beta_i z_{3i}) \log x_i^2
\]

(146)

\[
\frac{\delta^2 \log \left( \frac{R_i}{\eta_1} \right)}{\partial \eta_1^2} = \left( \frac{\delta \log(R_i)}{\partial \eta_1} \right) \left( \frac{\delta \log \left( \frac{R_i}{\eta_1} \right)}{\partial \eta_1} \right)
\]

(147)

\[
\frac{\delta^2 LL}{\partial \eta_2^2} = -2 \sum_{i \in \mathcal{E}_3} \frac{\delta^2 \log(R_i)}{\partial \eta_2^2} - \sum_{i \in \mathcal{E}_3 \times \mathcal{E}_3} \frac{\delta^2 \log(R_i)}{\partial \eta_2^2} - \\
\sum_{i \in \mathcal{E}_3 \times \mathcal{E}_3} \exp(\beta_i Z_{3i}) \frac{\delta^2 \log(R_i)}{\partial \eta_2^2} \frac{\delta \log(R_i)}{\partial \eta_2} - \frac{N_1 + N_3}{\eta_2^2}
\]

where,

\[
\frac{\delta^2 \log(R_i)}{\partial \eta_2^2} = \frac{R_i \left| \frac{\delta^2 R_i}{\partial \eta_2^2} \right| - \left( \frac{\delta R_i}{\partial \eta_2} \right)^2}{\frac{R_i^2}{\eta_2^2}}
\]

(149)
and,

\[ \frac{\delta^2 R_i}{\delta \eta_i^2} = y_i^\eta \exp(\beta' z_i) (\log y_i)^2 \]  

(150)

\[ \frac{\delta^2 \log (R_i)}{\delta \eta_i^2} = \left\{ \frac{\delta \log (R_i)}{\delta \eta_i} \right\} \cdot \left\{ \log (y_i) - \frac{\delta \log (R_i)}{\delta \eta_i} \right\} \]  

(151)

\[ \frac{\delta^2 LL}{\delta \beta_{1,j} \delta \beta_{2,k}} = -2 \sum_{i \in \epsilon_1} \frac{\delta^2 \log (R_i)}{\delta \beta_{1,j} \delta \beta_{2,k}} - \sum_{i \in \epsilon_2, \epsilon_3} \frac{\delta^2 \log (R_i)}{\delta \beta_{1,j} \delta \beta_{2,k}} - \sum_{i \in \epsilon_4} \exp(\beta' z_i) \frac{\delta^2 \log (R_i)}{\delta \beta_{1,j} \delta \beta_{2,k}} \]  

(152)

where,

\[ \frac{\delta^2 \log (R_i)}{\delta \beta_{1,j} \delta \beta_{2,k}} = - \left\{ \frac{\delta \log (R_i)}{\delta \beta_{1,j}} \right\} \cdot \left\{ \frac{\delta \log (R_i)}{\delta \beta_{2,k}} \right\} \]  

(153)

\[ \frac{\delta^2 LL}{\delta \beta_{1,j} \delta \beta_{3,l}} = - \sum_{i \in \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4} z_{3,l} \exp(\beta' z_i) \left\{ \frac{\delta \log (R_i)}{\delta \beta_{1,j}} \right\} \]  

(154)

\[ \frac{\delta^2 LL}{\delta \beta_{1,j} \delta \eta_i} = -2 \sum_{i \in \epsilon_1} \frac{\delta^2 \log (R_i)}{\delta \beta_{1,j} \delta \eta_i} - \sum_{i \in \epsilon_2, \epsilon_3} \frac{\delta^2 \log (R_i)}{\delta \beta_{1,j} \delta \eta_i} - \sum_{i \in \epsilon_4} \exp(\beta' z_i) \frac{\delta^2 \log (R_i)}{\delta \beta_{1,j} \delta \eta_i} \]  

(155)

where,

\[ \frac{\delta^2 \log (R_i)}{\delta \beta_{1,j} \delta \eta_i} = \frac{z_{1,j} R_i}{\delta \eta_i} \left( \frac{\delta R_i}{\delta \eta_i} \right) - \left( \frac{\delta R_i}{\delta \beta_{1,j}} \right) \cdot \frac{\delta R_i}{\delta \eta_i} \]  

(156)

or,

\[ \frac{\delta^2 \log (R_i)}{\delta \beta_{1,j} \delta \eta_i} = \left\{ \frac{\delta \log (R_i)}{\delta \eta_i} \right\} \cdot \left( z_{1,j} - \frac{\delta \log (R_i)}{\delta \beta_{1,j}} \right) \]  

(157)
\[
\frac{\partial^2 LL}{\partial \beta_1 \partial \eta_2} = -2 \sum_{i \in e_1} \frac{\partial^2 \log(R_i)}{\partial \beta_1 \partial \eta_2} - \sum_{i \in e_2, e_3} \frac{\partial^2 \log(R_i)}{\partial \beta_1 \partial \eta_2} - \sum_{i \in e_1, e_2, e_3, e_4} \exp(\beta_3 z_{3i}) \frac{\partial^2 \log(R_i)}{\partial \beta_1 \partial \eta_2} \tag{158}
\]

where,

\[
\frac{\partial^2 \log(R_i)}{\partial \beta_1 \partial \eta_2} = - \left( \frac{\partial \log(R_i)}{\partial \beta_1} \right) \left( \frac{\partial \log(R_i)}{\partial \eta_2} \right) \tag{159}
\]

\[
\frac{\partial^2 LL}{\partial \beta_2 \partial \beta_3 l} = - \sum_{i \in e_1, e_2, e_3, e_4} z_{3i} \exp(\beta_3 z_{3i}) \left( \frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \beta_3 l} \right) \tag{160}
\]

\[
\frac{\partial^2 LL}{\partial \beta_2 \partial \eta_1} = -2 \sum_{i \in e_1} \frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \eta_1} - \sum_{i \in e_2, e_3} \frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \eta_1} - \sum_{i \in e_1, e_2, e_3, e_4} \exp(\beta_3 z_{3i}) \frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \eta_1} \tag{161}
\]

where,

\[
\frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \eta_1} = - \left( \frac{\partial \log(R_i)}{\partial \beta_2} \right) \left( \frac{\partial \log(R_i)}{\partial \eta_1} \right) \tag{162}
\]

\[
\frac{\partial^2 LL}{\partial \beta_2 \partial \eta_2} = -2 \sum_{i \in e_1} \frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \eta_2} - \sum_{i \in e_2, e_3} \frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \eta_2} - \sum_{i \in e_1, e_2, e_3, e_4} \exp(\beta_3 z_{3i}) \frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \eta_2} \tag{163}
\]

where,

\[
\frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \eta_2} = \frac{z_{2ki} R_i}{R_i^2} \left( \frac{\partial R_i}{\partial \eta_2} \right) - \left( \frac{\partial R_i}{\partial \beta_2} \right) \left( \frac{\partial R_i}{\partial \eta_2} \right) \tag{164}
\]
or,

$$\frac{\partial^2 \log(R_i)}{\partial \beta_2 \partial \eta_2} = \left[ \frac{\partial \log(R_i)}{\partial \eta_2} \right] \left\{ \frac{2k_i}{\partial \beta_2} - \frac{\partial \log(R_i)}{\partial \beta_2} \right\}$$

(165)

$$\frac{\partial^2 \text{LL}}{\partial \beta_3 \partial \eta_1} = - \sum_{i \in f_1, f_2, f_3, f_4} z_{3i} \exp(\beta'_3 z_{3i}) \left\{ \frac{\partial \log(R_i)}{\partial \eta_1} \right\}$$

(166)

$$\frac{\partial^2 \text{LL}}{\partial \beta_3 \partial \eta_2} = - \sum_{i \in f_1, f_2, f_3, f_4} z_{3i} \exp(\beta'_3 z_{3i}) \left\{ \frac{\partial \log(R_i)}{\partial \eta_2} \right\}$$

(167)

$$\frac{\partial^2 \text{LL}}{\partial \eta_1 \partial \eta_2} = -2 \sum_{i \in f_1} \frac{\partial^2 \log(R_i)}{\partial \eta_1 \partial \eta_2} - \sum_{i \in f_2 \neq f_1} \frac{\partial^2 \log(R_i)}{\partial \eta_1 \partial \eta_2}$$

$$\sum_{i \in f_3, f_4, f_2 \neq f_1} \exp(\beta'_3 z_{3i}) \frac{\partial^2 \log(R_i)}{\partial \eta_1 \partial \eta_2}$$

(168)

where,

$$\frac{\partial^2 \log(R_i)}{\partial \eta_1 \partial \eta_2} = - \left[ \frac{\partial \log(R_i)}{\partial \eta_1} \right] \left[ \frac{\partial \log(R_i)}{\partial \eta_2} \right]$$

(169)
APPENDIX C: FORTRAN PROGRAMS

// JOB , 'J. T. WASSELL'.
// REGION=2048K, TIME=5
// EXEC VSF2CG, TIME=5
// FORT. SYSIN DD '

C
C
C------ PURPOSE --------------------------------------------------
C
C THIS PROGRAM READS AND PRINTS SOME OF THE DATA.
C INITIAL STARTING VALUES FOR THE PARAMETERS ARE SET. OTHER SUBROUTINES ARE CALLED TO PERFORM ITERATIVE CALCULATIONS.
C
CONSULT THE DESCRIPTIONS IN SUBROUTINES PARML - THIS SUBROUTINE CALLS FOR THE VALUE OF THE LOG LIKELIHOOD AND DERIVATIVES. IT Chooses Parameter ESTIMATES WHICH MAXIMIZE THE LIKELIHOOD BY NEWTON-RAPSON OR STEEPEST DESCENT.
C UW BULL - THIS SUBROUTINE RETURNS THE LOG LIKELIHOOD AND DERIVATIVES REQUIRED FOR A UNIVARIATE WEIBULL REGRESSION.
C OKDER - THIS SUBROUTINE RETURNS THE LOG LIKELIHOOD AND DERIVATIVES REQUIRED BY THE BIVARIATE CLAYTON-OAKES MODEL.
C LKDER - THIS SUBROUTINE RETURNS THE LOG LIKELIHOOD AND DERIVATIVES REQUIRED BY THE BIVARIATE LEE-KLEIN MODEL.
C
C
COMMON Z1,Z2,Z3,X,Y,IX,IY
INTEGER IX(4531),IY(4531),
* N,DIM(3),VAR,ITER,GOODI,OUTP,PO
REAL*8 X(4531),Y(4531),
* Z1(4531,6),Z2(4531,9),Z3(4531,3),NP1(20),
* LLIK1,SE(20),ZZZ(20),VNC(20),
* HYPER,SMOK,MALE,CHOL,PDU,AGE,QD,DIAB
EXTERNAL PARML,SOLN,OUT,UWBULL,OKDER,LKDER
C
C N REPRESENTS THE NUMBER OF OBSERVATIONS TO BE
ANALYZED. IT MUST AGREE WITH THE DIMENSIONS OF X, Y, IX, IY, AND THE FIRST DIMENSIONS OF Z1, Z2, AND Z3.

N=4531
OUTP=6
DIM(1)=1
DIM(2)=8
DIM(3)=1
DO 10 I=1,4531
   Z1(I,1)=1.0D0
   Z2(I,1)=1.0D0
   Z3(I,1)=1.0D0
10 CONTINUE
DO 20 I=1,20
   NP1(I)=0.0D0
   VNC(I)=0.0D0
20 CONTINUE
LLIKH1=0.0D0
ITER=0
CONV=1
GOOD=0
DO 40 I=1,N
   READ(5,500) X(I),IX(I),HYPER,Y(I),IY(I),PDU,SMOK,
               MALE,CHOL,AGE,QD,DIA
   Z2(I,2)=PDU
   Z2(I,3)=SMOK
   Z2(I,4)=MALE
   Z2(I,5)=CHOL
   Z2(I,6)=QD
   Z2(I,7)=SMOK*MALE
   Z2(I,8)=PDU*MALE
40 CONTINUE
PRINT *, 'SET VAR TO REQUEST OTHER MODELS'
PRINT *, '----- OAKES MODEL -----'
PRINT *, 'Z2(2)=PDU'
PRINT *, 'Z2(3)=SMOK, Z2(4)=MALE, Z2(5)=CHOL'
PRINT *, 'Z2(6)=QD'
PRINT *, 'Z2(7)=SMOK*MALE'
PRINT *, 'Z2(8)=DRUG*MALE'

THE NEXT THREE STATEMENTS WRITE OUT THE INFORMATION FOR 50 OF THE CASES. CHANGE THE NEXT STATEMENT TO PRINT MORE OR FEWER CASES.

DO 45 I=1,50
   WRITE(6,600) X(I),IX(I),Y(I),IY(I),Z2(I,2),
                Z2(I,3),Z2(I,4),Z2(I,5),Z2(I,6),Z2(I,7),Z2(I,8)
45 CONTINUE
VAR DEFINES THE VARIABLES AND THE MODEL TO BE USED.

VAR=1 PERFORMS THE WEIBULL REGRESSION ON THE X VAR.

VAR=2 PERFORMS THE WEIBULL REGRESSION ON THE Y VAR.

VAR=3 PERFORMS THE BIVARIATE OAKES MODEL ANALYSIS.

VAR=4 PERFORMS THE BIVARIATE L & K MODEL ANALYSIS.

VAR=3

PO=1
NP1(1)=6.114923D0
NP1(2)=7.552508D0
NP1(3)=-26.431688D0
NP1(4)=-33.480571D0
NP1(5)=-0.59260D0
NP1(6)=-0.02996454D0
NP1(7)=0.5450982D0
NP1(8)=0.12387732D0
NP1(9)=0.217887D0
NP1(10)=0.475012D0
NP1(11)=-0.559152D0
NP1(12)=0.461772D0
CALL PARML(NP1,VNC,LLIKH1,DIM,VAR,N,ITER,CONV,GOODI,PO)

DO 67 J=1,2-DIM(1)-DIM(2)+DIM(3)
   SE(J)=VNC(J)**0.50D0
   IF (SE(J) .EQ. 0.0D0) GO TO 67
   ZZZ(J)=NP1(J)/SE(J)
   PRINT *, 'COEFF=',NP1(J), 'STD ERR=',SE(J), 'Z=',ZZZ(J)
   PRINT *, '-----------------------------'
67 CONTINUE

FORMAT STATEMENT 500 IS USED TO READ THE DATASET.

FORMAT STATEMENT 600 IS USED TO WRITE OUT THE INFORMATION ASSOCIATED WITH 50 OF THE CASES.

500 FORMAT(6X,F7.4,1X,I1,I1,I1,F7.4,1X,I1,2X,F1.0,6X,
   F1.0,11X,F1.0,1X,F1.0,2X,F7.4,1X,F1.0,13X,F1.0)

600 FORMAT(1X,F9.4,1X,I1,2X,F9.4,1X,I1,6(3X,F3.0),
   3X,F7.4)

9999 STOP

END
SUBROUTINE PARML(NP, VNC, LLIKH, DIM, VAR, N, ITER, CONV, GOODI, PO)

THIS IS THE PARAMETER LOOP FOR MARQUARDT'S
COMPROMISE, USED TO MAXIMIZE THE LOG LIKELIHOOD.
IT CALLS SUBROUTINES FOR THE LOG LIKELIHOOD AND
THE FIRST AND SECOND DERIVATIVES.
THIS SUBROUTINE CHECKS ESTIMATES FOR CONVERGENCE
AND COUNTS ITERATIONS.
THE SUBROUTINE
OUT IS CALLED TO PRINT SUMMARY RESULTS.
THE SUBROUTINE SOLN IS CALLED FOR NEW PARAMETER
ESTIMATES AFTER A SUCCESSFUL ITERATION.
THE IMSL SUBROUTINE DWRRRL PRINTS OUT A MATRIX.
THE IMSL SUBROUTINE DLINRG RETURNS THE INVERSE
OF THE MATRIX OF SECOND DERIVATIVES.
ARGUMENTS:
NP - STARTING VALUES FOR PARAMETER ESTIMATES
(INPUT). FINAL PARAMETER ESTIMATES (OUTPUT).
VNC - VARIANCE ESTIMATES OF THE PARAMETERS IN
NP (OUTPUT).
LLIKH - THE FINAL VALUE OF THE MAXIMIZED LOG
LIKELIHOOD (OUTPUT).
DIM - DIM(1) IS SET EQUAL TO 1 + THE NUMBER OF
COVARIATES FOR THE FIRST TIME: X.
DIM(2) IS SET EQUAL TO 1 + THE NUMBER OF
COVARIATES FOR THE SECOND TIME: Y.
DIM(3) IS SET EQUAL TO 1 + THE NUMBER OF
COVARIATES FOR THE DEPENDENCE PARAMETER.
NOTE THAT THE TOTAL NUMBER OF PARAMETERS IS
EQUAL TO 2 + DIM(1) + DIM(2) + DIM(3).
VAR - DIRECTS WHICH SUBROUTINE IS TO BE CALLED
FOR THE LOG LIKELIHOOD AND 1ST AND 2ND
DERIVATIVES, ETC. (INPUT).
N - THE NUMBER OF OBSERVATIONS, THE DIMENSION OF
X, IX, Y, IY (INPUT).
ITER - COUNTS THE NUMBER OF TIMES A NEW SET
OF PARAMETER ESTIMATES WERE TRIED
(Both SUCCESSFULLY AND UNSUCCESSFULLY)
CONV - ALWAYS SET TO 0 ON INPUT.
1 - IF STUCK IN STEEPEST DESCENT.
2 - STARTING VALUE ESTIMATES INAPPROPRIATE.
3 - CONVERGENCE FAILED.
GOODI - THE NUMBER OF SUCCESSFUL ITERATIONS
(OUTPUT).
PO - IF = 0 NOTHING IS PRINTED AT INTERMEDIATE
ITERATIONS. (USE FOR PLOTTING THE LIKELIHOOD
EXTERNAL SOLN.OUT.OKDER

INTEGER DIM(3), MD, N, VAR, DER, CONV, GOODI, PO,
* IER, ITER, ITORER, FOUL, FAIR, LAST, MAX, PLOT,
* PARM, EXIT1(20), EXIT2(20), SUCESS(200)
REAL*8 LL1(200), DL(20), D2L(20, 20), H(20, 20),
* DB(20), BP, TOL, DTOL, LIKTOL, LOGL, DELTA,
* OP(200, 20), ODL(200, 20), OD2L(200, 20, 20),
* VNC(20), VCOV(20, 20), WKAREA(460), LLIKH,
* NP(20)
CHARACTER*8 FMT, RL(1), CL(1)

INITIALIZE THE CONTROL VARIABLES: ITER COUNTS
THE NUMBER OF ITERATIONS.
THE VARIABLE MAX SETS THE UPPER LIMIT (<=200)
FOR THE NUMBER OF ITERATIONS IN ANY ONE CALL TO
THIS SUBROUTINE. CONVERGENCE IS TESTED BY
THE VARIABLES TOL, DTOL AND EXIT.
CONVERGENCE IS DECLARED WHEN THE ESTIMATES
AND THE FIRST DERIVATIVES ARE CHANGING BY LESS
THAN THE VALUE OF TOL BETWEEN ITERATIONS.
THE ABSOLUTE VALUE OF THE FIRST DERIVATIVES MUST
ALSO BE LESS THAN DTOL BEFORE CONVERGENCE IS
DECLARED. ALL PARAMETER ESTIMATES MUST PASS THE
TESTS BEFORE CONVERGENCE IS DECLARED.

IF (VAR .NE. 1 .AND. VAR .NE. 2 .AND.
* VAR .NE. 3 .AND. VAR .NE. 4 .AND. VAR .NE. 5
* .AND. VAR .NE. 6) GO TO 9000
ITER=1
ITORER = 0
FOUL=0
FAIR=0
LAST=1
DER=1
CONV=1
GOODI=0
BP=0.0D0
LOGL=0.0D0
TOL=1.0D-02
DTOL=1.0D-02
LIKTOL=(-1.0D-03)
MAX=20
DO 10 I=1,20
DL(I)=0.0D0
DB(I)=0.0DO 
VNC(I)=0.0DO 
EXIT1(I)=0 
EXIT2(I)=0 
DO 10 J=1,20 
    VC0V(I,J)=0.0DO 
    H(I,J)=0.0DO 
    D2L(I,J)=0.0DO 
10 CONTINUE 
DO 20 I=1,200 
    LL1(I)=0.0DO 
    SUCESS(I)=0 
    OP(I,J)=0.0DO 
    CDL(I,J)=0.0DO 
    DO 20 J=1,20 
    OD2L(I,J,K)=0.0DO 
20 CONTINUE 
IF (VAR.EQ.1 .OR. VAR.EQ.2) MD=DIM(VAR)+1 
IF (VAR.EQ.3 .OR. VAR.EQ.4) 
    MD=DIM(1)+DIM(2)+DIM(3)+2 
IF (VAR.EQ.5) MD=DIM(1)+DIM(2)+1 
IF (VAR.EQ.6) MD=DIM(1)+DIM(2) 
----------------------------------------------------------------------------------------------------------------------
PLOTS OF THE LIKELIHOOD FUNCTION AND THE 1ST DERIV.
ARE REQUESTED WHEN PLOT=1. OTHERWISE LET PLOT=0.
PLOTTING INSTRUCTIONS: 1. CHOOSE THE PARM NUMBER.
2. SELECT THE INCREMENT. 3. SET START VALUE.
---------------------------------------------------------------------------------------------------------------
PLOT=0 
IF (PLOT .EQ. 1) THEN 
    PARM=5 
    DELTA = 0.100DO 
ENDIF 

IF (PO .EQ. 1) WRITE (6,6010) 
1000 CONTINUE 
---------------------------------------------------------------------------------------------------------------
** ENTER THE PARAMETER LOOP. 
CALL A SUBROUTINE TO GET THE LOGLIKELIHOOD AND FIRST AND SECOND DERIVATIVES. 
---------------------------------------------------------------------------------------------------------------
IF (VAR .EQ. 1 .OR. VAR .EQ. 2) 
    CALL UWBULL(D2L,DL,NP,LOGL,DIM,VARN,PLOT, 
    ITORER,DER)
IF (VAR .EQ. 3)
* CALL OKDER(NP,LOGL,DL,D2L,DIM,N,PLOT,ITORER,DER)

IF (VAR .EQ. 4)
* CALL LKDER(NP,LOGL,DL,D2L,DIM,N,PLOT,ITORER,DER)

LL1(ITER)=LOGL
DO 1100 I=1,MD
   OP(ITER,I)=NP(I)
   ODL(ITER,I)=DL(I)
   DO 1100 J=1,MD
      OD2L(ITER,I,J)=D2L(I,J)
1100 CONTINUE

IF (PLOT .EQ. 1) THEN
   NP(PARM)=NP(PARM)+DELTA
   GO TO 1400
ENDIF

IF (PO .EQ. 1 ) THEN
   WRITE (6,6100) ITER,LL1(ITER)
   WRITE (6,6110) (NP(J), J=1, MD)
   IF (DER .EQ. 0) GO TO 1110
   WRITE (6,6120) (DL(J), J=1,MD)
   CALL PAGE(-1.129)
   FMT='(W15.6)'
   RL(1)="NUMBER"
   CALL DWRRRL( "2ND DER. =", MD, MD, D2L,20,0,FMT,RL,CL)
   WRITE (6,6000)
1110 ENDIF

C-------------------------------
C CHECK IF THE LOGLIKELIHOOD INCREASED (UNLESS IT'S
C THE SECOND ATTEMPT), TO SET THE BLENDING PARAMETER.
C-------------------------------

IF (ITORER .EQ. 9) GO TO 3000
IF (ITORER .GE. 2 .AND. ITER .EQ. 1 ) GO TO 3000
IF (ITER .EQ. 1) GO TO 1250
IF (ITORER .GE. 1) GO TO 1150
IF (LL1(ITER) - LL1(LAST) .LT. LIKTOL) GO TO 1140

   DER = DER - 1
   DER = IABS(DER)
   IF (DER .NE. 1 .AND. DER .NE. 0) DER = 1
   IF (DER .EQ. 1) GO TO 1000
   LAST = ITER
   FAIR = FAIR + 1
   SUCESS(ITER) = FAIR
   IF (PO .EQ. 1) WRITE (6,6130) SUCESS(ITER)
FOUL = 0
BP = 0.0D0
GO TO 1300

1140 CONTINUE
ITORER = 1
1150 CONTINUE
SUCESS(ITER) = 0
FOUL = FOUL + 1
IF (FOUL .EQ. 1) BP = 0.01D0
IF (FOUL .EQ. 2) BP = 0.1D0
IF (FOUL .EQ. 3) BP = 1.0D0
IF (FOUL .EQ. 4) BP = 5.0D0
IF (FOUL .EQ. 5) BP = 10.0D0
IF (FOUL .GE. 6) BP = FOUL**2
IF (PO .EQ. 1) THEN
IF (ITORER .EQ. 1) WRITE (6,6140) ITER
IF (ITORER .EQ. 2) WRITE (6,6150) ITER
IF (ITORER .EQ. 3) WRITE (6,6155) ITER
WRITE (6,6160) BP
ENDIF
IF (FOUL .GE. 20) GO TO 3500
ITORER = 0
DO 1200 I = 1,MD
   NP(I) = OP(LAST.I)
   DL(I) = ODL(LAST.I)
   DO 1200 J = 1,MD
      D2L(I,J) = OD2L(LAST,I,J)
1200 CONTINUE
GO TO 1300
1250 DER = 0
1300 CALL SOLN(H, D2L, DL, DB, NP, BP, MD)
C---------------------------------------------------------------------
C CHECK FOR CONVERGENCE.
C THEN STORE THE ESTIMATES IN THE ARRAY FOR PRINTING.
C EXIT THE LOOP IF CONVERGENCE IS ACHIEVED.
C---------------------------------------------------------------------
S = 0
DO 1350 I = 1, MD
   EXIT1(I) = 0
   EXIT2(I) = 0
   IF (DABS(DB(I)).LT.DABS(NP(I)*TOL)) EXIT1(I) = 1
   IF (DABS(DL(I)).LE.DTOL) EXIT2(I) = 1
   S = S + EXIT1(I) + EXIT2(I)
1350 CONTINUE
IF (S .EQ. MD*2) CONV = 0
C EXIT THE LOOP IF THE MAXIMUM NUMBER OF
C ITERATIONS HAVE BEEN TRIED WITHOUT SUCCESS.

1400 CONTINUE
IF (ITER.EQ.MAX .OR. CONV .EQ. 0) GO TO 4000
ITER=ITER+1
IF (ITER.LE.MAX .AND. CONV .EQ. 1) GO TO 1000

3000 CONTINUE
CONV=2
GO TO 4000

3500 CONTINUE
CONV=3

4000 CONTINUE

GOODI=SUCCESS(LAST)
LLIKH=LL1(LAST)

IF (PO .EQ. 1 .OR. PLOT .EQ. 1)
  CALL OUT(OP,ODL,H,LL1,SUCESS,ITER,LAST,MAX,
            PARM,MD,CONV,PLOT)

IF (CONV .NE. 0) GO TO 9000
IER=0
IDGT=0
CALL DLINRG(MD,H,20,VCOV,20)
DO 4100 I=1,20
     VNC(I)=VCOV(I,1)
4100 CONTINUE

6000 FORMAT('0',62(2H =))
6010 FORMAT('1',//62(2H =)////)
6100 FORMAT('0','ITERATION NUMBER= ',I4,9X,
    'LOGLIKELIHOOD= ',F15.8)
6110 FORMAT('0','PARMS= ',8(F13.6,1X)/7X,8(F13.6,1X)/
    7X,8(F13.6,1X))
6120 FORMAT('0','1ST DER= ',8(F13.6,1X)/9X,8(F13.6,1X)/
    9X,8(F13.6,1X))
6130 FORMAT('0','SUCCESS COUNTER: ',5X,I4,5X,
    'CUMULATIVE SUCCESSES')
6140 FORMAT('0','10(2H* ),'USING STEEPEST DESCENT IN ',
    'ITERATION NUMBER ',I4,
    ' THE LIKELIHOOD DID NOT INCREASE.')
6150 FORMAT('0','10(2H* ),'USING STEEPEST DESCENT IN ',
    'ITERATION NUMBER ',I4,
    ' PARAMETER VALUES ARE OUT OF BOUNDS.')
6155 FORMAT('0','10(2H* ),'USING STEEPEST DESCENT IN ',
    'ITERATION NUMBER ',I4,
    ' OVERFLOW OCCURED IN CALCULATIONS.')
THE BLENDING PARAMETER VALUE, IS: 

```
SUBROUTINE SOLN(H, D2L, DL, DB, NP, BP, MD)
C-------------------------------------------------------------------
C-------------------------------------------------------------------
REAL*8 DL(20), G(20), D2L(20,20),
* SC(20,20), ID(20,20), SI(20,20), SINV(20,20),
* WKAREA(460), BP.
* C(20,20), C1(20,20), C2(20,20), C3(20,20),
* H(20,20), DB(20), NP(20)
INTEGER MD, IER
C-------------------------------------------------------------------
C SC IS THE SCALING MATRIX WHICH PUTS ONES ON THE DIAGONALS OF H BEFORE ADJUSTMENT BY THE BLENDING PARAMETER. H IS THE HESSIAN MATRIX - NEGATIVE OF THE 2ND DERIV' S.
C-------------------------------------------------------------------
DO 1000 I=1,MD
G(I)=(-1.0D0)*DL(I)
DO 1000 J=1,MD
H(I,J)=(-1.0D0)*D2L(I,J)
SC(I,J)=0.0D0
IF (I .EQ. J) SC(I,J)=
* 1.0D0*(DABS(D2L(I,J))**0.50D0)
1000 CONTINUE
C-------------------------------------------------------------------
C PRE AND POST MULTIPLY THE HESSIAN BY THE SCALING MATRIX. THE RESULT IS CALLED C1.
C-------------------------------------------------------------------
IER=0
CALL DMRRRR(MD, MD, SC, 20, MD, MD, H, 20, MD, MD, C, 20)
IER=0
CALL DMRRRR(MD, MD, C, 20, MD, MD, SC, 20, MD, MD, C1, 20)
C-------------------------------------------------------------------
ADD BP (LAMDA) X IDENTITY MATRIX TO C1. CALL THE RESULT SI.
THE BLENDING PARAMETER, BP, IS SMALL IF THE NEWTON-RHAPSON METHOD IS USED. AND IT IS LARGE IF THE METHOD OF STEEPEST DESCENT IS TO BE USED.
```
DO 1300 I=1,MD
   DO 1300 J=1,MD
      IF (I .EQ. J) ID(I,J)=1.0D0*BP
   1300 SI(I,J)=C1(I,J)+ID(I,J)
C GET THE INVERSE OF SI, THEN UNSCALE IT BY PRE AND i
C POST MULTIPLICATION OF THE SCALING MATRIX. (SC)
C---------------------------------------------------------------------
IER=0
IDGT=0
CALL DLINRG(MD,SI,20,SINV,20)
IER=0
CALL DMRRRR(MD,MD,SC,20,MD,MD,SINV,20,MD,MD,C2,20)
IER=0
CALL DMRRRR(MD,MD,C2,20,MD,MD,SC,20,MD,MD,C3,20)
C FINALLY, POST MULTIPLY BY THE VECTOR OF 1ST DERIVATIVES
C TO OBTAIN THE INCREMENTAL VECTOR DB.
C---------------------------------------------------------------------
IER=0
CALL DMURRV(MD,MD,C3,20,MD,G,1,MD, DB)
C PRINT *.'DB='.(DB(J),J=1, MD)
DO 1600 I=1,MD
   NP(I)=NP(I)-DB(I)
1600 CONTINUE
9000 RETURN
9999 END
C---------------------------------------------------------------------
SUBROUTINE OUT(OP,ODL,H,LL1,SUCESS,ITER,LAST,MAX,
   * PARM,MD,CONV,PLOT)
---------------------------------------------------------------------
REAL*8 OP(200,20),ODL(200,20),LL1(200),
   * XV(200),YV(200,2),RANGE(4),
   * H(20,20),VCOV(20,20),WKAREA(460)
CHARACTER*8 FMT,RL(1),CL(1)
INTEGER ITER,MAX,SUCESS(200),MD,CONV,PLOT,PARM,
   * IER,LAST
---------------------------------------------------------------------
C H IS THE HESSIAN MATRIX - NEGATIVE OF THE 2ND DERIV'S.
C---------------------------------------------------------------------
IF (PLOT .EQ. 1) GO TO 300
IF (CONV .EQ. 2) GO TO 60
IF (CONV .EQ. 0) GO TO 50
WRITE (6,6001)
WRITE (6,6006) MAX,SUCESS(LAST)
WRITE (6,6002)  
WRITE (6,6001)  
WRITE (6,6008) ITER  
WRITE (6,6002)  
IER=0  
IDGT=0  
FMT='(W15.6)'  
RL(1)= 'NUMBER'  
CL(1)= 'NUMBER'  
CALL PAGE(-1,129)  

CALL DLINRGR(MD, H, 20, VCOV, 20)  
CALL DWRRRL('VCOV=', MD, MD, VCOV, 20, -1, FMT, RL, CL)  
WRITE (6,6002)  
IF (CONV .EQ. 3) GO TO 70  
GO TO 100  
60 CONTINUE  
WRITE (6,6001)  
WRITE (6,6030)  
GO TO 100  
70 CONTINUE  
WRITE (6,6001)  
WRITE (6,6040)  
100 WRITE (6,6001)  
DO 200 I=1,ITER  
   IF (I .LE. 1 .OR. SUCESS(I) .GT. SUCESS(I-1)) THEN  
      WRITE (6,6010) SUCESS(I), LL1(I), (OP(I, J), J=1, MD)  
      WRITE (6,6020) (ODL(I, J), J=1, MD)  
   ENDIF  
200 CONTINUE  
WRITE (6,6002)  
GO TO 9000  
300 CONTINUE  
RANGE(1)=0.0D0  
RANGE(2)=0.0D0  
RANGE(3)=0.0D0  
RANGE(4)=0.0D0  
DO 400 I = 1, MAX  
   XV(I)=OP(I, PARM)  
   YV(I,1)=LL1(I)  
   YV(I,2)=ODL(I, PARM)  
400 CONTINUE  
IER=0  
CALL DPLOTOP(MAX, 2, XV, YV, 200, 1, RANGE, 'LD',  
             'PARAMETER VALUES',)
SUBROUTINE UWBULL(D2L, DL, NP, LOGL, DIM, VAR, N, PLOT, ITOER, DER)

COMMON Z1, Z2, Z3, X, Y, IX, IY
INTEGER IX(4531), IY(4531), DEAD,
* I, J, K, L, N, PLOT, ITOER, DIM(3), VAR, DER
REAL*8 X(4531), Y(4531), TIME,
* Z1(4531, 3), Z2(4531, 3), Z3(4531, 3), COV(9),
* NP(20), BETA(9), E1, XE1, LX, A1, EA1,
* D1E1, D1B1, D2B1, D2B1E1(9),
* D2B1(9, 9),
* LOGL, DL(20),
* D2L(20, 20)

ENTER THE ITERATIVE LOOP OF CALCULATIONS INVOLVING
THE PARAMETERS TO BE ESTIMATED.
INITIALIZE THE SUMMING VARIABLES AND MATRICES.
CALL ERRSET(207,0,-1,1,0,208)
IND=2
CALL OVERFL(IND)

E1=NP(1)
IF (E1 .LE. 0.0D0 .OR. E1 .GT. 174.6) GO TO 3000
DO 3 I=1,DIM(VAR)
   BETA(I)=NP(I+1)
   D1B1(I)=0.0D0
   D2B1E1(I)=0.0D0
   DO 3 J = 1, DIM(VAR)
    D2B1(I,J)=0.0D0
  3 CONTINUE

LOGL=0.0D0
D1E1=0.0D0
D2E1=0.0D0

PASS THROUGH THE DATA.

DO 1000 I=1,N
IF (VAR .EQ. 1) THEN
   TIME=X(I)
   DEAD=IX(I)
   DO 5 J=1,DIM(1)
    COV(J)=Z1(I,J)
  5 CONTINUE
ENDIF
IF (VAR .EQ. 2) THEN
   TIME=Y(I)
   DEAD=IY(I)
   DO 6 J=1,DIM(2)
    COV(J)=Z2(I,J)
  6 CONTINUE
ENDIF

A1=0.0D0
DO 10 J=1,DIM(VAR)
   A1=A1+(BETA(J)*COV(J))
10 CONTINUE
IF (DABS(A1).GT. 174.6) GO TO 3000
XE1=TIME**E1
EA1=DEXP(A1)*XE1
LX=DLOG(TIME)
C CALCULATE THE COMPONENTS OF THE LOGLIKELIHOOD.

\[
\text{LOGL} = \text{LOGL} - \text{EA1} + (A1 + (E1-1) \times LX + \log(E1)) \times \text{DEAD}
\]

IF (DER .EQ. 0) GO TO 1000

C CALCULATE 1ST DERIVATIVES

\[
D1E1 = D1E1 - (EA1 \times LX) + (LX \times \text{DEAD}) + (\text{DEAD}/E1)
\]

DO 110 J = 1, DIM(VAR)
    D1B1(J) = D1B1(J) - (COV(J) \times EA1) + (COV(J) \times \text{DEAD})
110 CONTINUE

C CALCULATE 2ND DERIVATIVES W.R.T. WEIBULL SHAPE PARMs.

\[
D2E1 = D2E1 - (EA1 \times LX^2) - (DEAD/(E1^2))
\]

C CALCULATE 2ND DERIV. ASSOCIATED W. ITH COVARIATES.

DO 210 J = 1, DIM(VAR)
    D2B1E1(J) = D2B1E1(J) - (COV(J) \times EA1 \times LX)
210 CONTINUE

DO 220 JP = 1, DIM(VAR)
    DO 220 J = JP, DIM(VAR)
        D2B1(J, JP) = D2B1(J, JP) - (COV(J) \times COV(JP) \times EA1)
220 CONTINUE

1000 CONTINUE

C NOW INITIALIZE THE VECTOR OF FIRST DERIVATIVES AND
C INITIALIZE THE MATRIX OF SECOND DERIVATIVES.
C THEN READ THE VALUES COMPUTED ABOVE
C INTO THE VECTOR AND MATRIX, RESPECTIVELY.

DO 1500 I = 1, 20
    DL(I) = 0.0DO
    DO 1500 J = 1, 20
        D2L(I, J) = 0.0DO
        D2L(J, I) = 0.0DO
1500 CONTINUE

DL(1) = D1E1
D2L(1, 1) = D2E1

DO 2000 I = 1, DIM(VAR)
    DL(I+1) = D1B1(I)
    D2L(I+1, 1) = D2B1E1(I)
    D2L(1, I+1) = D2B1E1(I)
2000 CONTINUE
DO 2000 J=1,DIM(VAR)
   D2L(I+1,J+1)=D2B1(I,J)
2000 CONTINUE
DO 2001 I=1,20
   DO 2001 J=1,20
      D2L(I,J)=D2L(J,I)
2001 CONTINUE
CALL OVERFL(IND)
IF (IND .NE. 2) GO TO 4000
RETURN
3000 CONTINUE
   ITORER=2
   RETURN
4000 CONTINUE
   ITORER=3
   RETURN
END

C
C
C
C
C

SUBROUTINE OKDER(NP, LOGL, DL, D2L, DIM, N, PLOT, ITORER, DER)

THIS SUBROUTINE USES THE CLAYTON-OAKES LIKELIHOOD FUNCTION WITH WEIBULL MARGINAL DISTNS.

COMMON Z1, Z2, Z3, X, Y, IX, IY
INTEGER IX(4531), IY(4531),
   * I, J, K, L, N, PLOT, ITORER, DER,
   * DIM1, DIM2, DIM3, DIM(3),
   * P1, P2, P12, P13
REAL*8 X(4531), Y(4531),
   * Z1(4531, 6), Z2(4531, 9), Z3(4531, 3),
   * NP(20),
   * B1(6), B2(9), B3(3), E1, E2,
   * N1, N2, N3, N4,
   * XE1, YE2
REAL*8 Q, LQ, LX, LY,
   * A1, A2, A3, A13, A23,
   * EA13, EA23,
* D1E1, D1E2,
* D2E1, D2E2, D2E3,
* D1B1(6), D1B2(9), D1B3(3)
* REAL*8 D2B1E1(6), D2B2E1(9), D2B3E1(3),
* D2B1E2(6), D2B2E2(9), D2B3E2(3),
* D2B1(6, 6), D2B2B1(6, 9), D2B1B3(6, 3),
* D2B2(9, 9), D2B2B3(9, 3),
* D2B3(3, 3),
* LC(10), D1C(20, 10),
* D2C(20, 20, 10),
* LOGL, DL(20),
* D2L(20, 20)
* REAL*8 T1, T2, T3, T4, T5, T6, T7, T8, T9,
* G1, G2, G3, G4, G12, G13, G23
* CALL ERRSET(207, 0, -1, 1, 0, 208)
* IND=2
* CALL OVERFL(IND)

C-----------------------------INITIALIZE A FEW SUMMING VARIABLES.
C-----------------------------

N1=0.0D0
N2=0.0D0
N3=0.0D0
N4=0.0D0
DIM1=DIM(1)
DIM2=DIM(2)
DIM3=DIM(3)

C-----------------------------ENTER THE ITERATIVE LOOP OF CALCULATIONS INVOLVING
C THE PARAMETERS TO BE ESTIMATED.
C INITIALIZE THE SUMMING VARIABLES AND MATRICES.
C-----------------------------

DO 1 I=1, 110
   LC(I)=0.0D0
   DO 1 J=1, 20
      D1C(J, I)=0.0D0
      DO 1 K=1, 20
         D2C(K, J, I)=0.0D0
1    CONTINUE
E1=NP(1)
E2=NP(2)
IF (E1 .LE. 0.0D0 .OR. E1 .GT. 174.6) GO TO 3000
IF (E2 .LE. 0.0D0 .OR. E2 .GT. 174.6) GO TO 3000
DO 3 I=1, DIM1
   B1(I)=NP(I+2)
3    CONTINUE
DO 4 J=1, DIM2
   B2(J)=NP(J+2+DIM1)
4    CONTINUE
DO 5 L=1, DIM3
   B3(L)=NP(L+2+DIM1+DIM2)
CONTINUE

PASS THROUGH THE DATA. CALCULATE THE BASIC QUANTITY Q. AND THE DERIV. QUANTITIES TO BE SUMMED.

DO 1000 I=1,N

G1=0.0D0
G2=0.0D0
G3=0.0D0
G4=0.0D0
G12=0.0D0
G13=0.0D0
G23=0.0D0

IF (IX(I).EQ.1 .AND. IY(I).EQ.1) G1=1.0D0
IF (IX(I).EQ.1 .AND. IY(I).EQ.0) G2=1.0D0
IF (IX(I).EQ.0 .AND. IY(I).EQ.1) G3=1.0D0
IF (IX(I).EQ.0 .AND. IY(I).EQ.0) G4=1.0D0

IF (G1+G2 .EQ. 1.0D0) G12=1.0D0
IF (G1+G3 .EQ. 1.0D0) G13=1.0D0
IF (G2+G3 .EQ. 1.0D0) G23=1.0D0

N1=N1*G1
N2=N2+G2
N3=N3+G3
N4=N4-G4

A1=0.0D0
A2=0.0D0
A3=0.0D0

DO 10 J=1,DIM1
   A1=A1+(B1(J)*Z1(I,J))
10 CONTINUE

DO 20 K=1,DIM2
   A2=A2+(B2(K)*Z2(I,K))
20 CONTINUE

DO 30 L=1,DIM3
   A3=A3+(B3(L)*Z3(I,L))
30 CONTINUE

IF (DABS(A3) .GT. 10.0D0) GO TO 3000
A13=A1+A3
IF (DABS(A13) .GT. 174.6) GO TO 3000
A23=A2+A3
IF (DABS(A23) .GT. 174.6) GO TO 3000
E3=DEXP(-1.0D0*A3)
XE1=X(I)**E1
YE2=Y(I)**E2
LX=DLOG(X(I))
135

\[ \text{LY} = \text{DLOG}(Y(I)) \]

\[ \text{EA13} = \text{DEXP}(A13) \times XE1 \]

\[ \text{IF (DABS(EA13).GT. 174.6) GO TO 3000} \]

\[ \text{EA23} = \text{DEXP}(A23) \times YE2 \]

\[ \text{IF (DABS(EA23).GT. 174.6) GO TO 3000} \]

\[ Q = \text{DEXP}(EA13) + \text{DEXP}(EA23) - 1.0D0 \]

\[ LQ = \text{DLOG}(Q) \]

--

C CALCULATE THE COMPONENTS OF THE LOGLIKELIHOOD.

C

\[ \text{LC(1)} = \text{LC(1)} - (LQ \times 2.0D0 \times G1) \]

\[ \text{LC(2)} = \text{LC(2)} - (LQ \times G23) \]

\[ \text{LC(3)} = \text{LC(3)} - (EA3 \times LQ) \]

\[ \text{LC(4)} = \text{LC(4)} + (\text{DLOG(DEXP(A3) + 1.0D0)} \times G1) \]

\[ \text{LC(5)} = \text{LC(5)} + (A1 \times G12) \]

\[ \text{LC(6)} = \text{LC(6)} + (A2 \times G13) \]

\[ \text{LC(7)} = \text{LC(7)} + ((E1 - 1.0DO) \times LX \times G12) \]

\[ \text{LC(8)} = \text{LC(8)} + ((E2 - 1.0DO) \times LY \times G13) \]

\[ \text{LC(9)} = \text{LC(9)} + ((EA13 + \text{DLOG(E1)}) \times G12) \]

\[ \text{LC(10)} = \text{LC(10)} + ((EA23 + \text{DLOG(E2)}) \times G13) \]

\[ \text{IF (DER . EQ. 0) GO TO 1000} \]

--

C CALCULATE 1ST DERIVATIVES W.R.T. WEIBULL SHAPE PARMS.

C

\[ \text{D1E1} = (XE1 \times LX \times \text{DEXP(EA13 + A13)}) \]

\[ \text{D1C(1,1)} = \text{D1C(1,1)} - ((\text{D1E1}/Q) \times 2.0D0 \times G1) \]

\[ \text{D1C(1,2)} = \text{D1C(1,2)} - ((\text{D1E1}/Q) \times G23) \]

\[ \text{D1C(1,3)} = \text{D1C(1,3)} - (EA3 \times (\text{D1E1}/Q)) \]

\[ \text{D1C(1,4)} = \text{D1C(1,4)} + (LX \times G12) \]

\[ \text{D1C(1,5)} = \text{D1C(1,5)} + (EA13 \times LX \times G12) \]

\[ \text{D1C(1,6)} = (N1 + N2) / E1 \]

\[ \text{D1E2} = (YE2 \times LY \times \text{DEXP(EA23 + A23)}) \]

\[ \text{D1C(2,1)} = \text{D1C(2,1)} - ((\text{D1E2}/Q) \times 2.0D0 \times G1) \]

\[ \text{D1C(2,2)} = \text{D1C(2,2)} - ((\text{D1E2}/Q) \times G23) \]

\[ \text{D1C(2,3)} = \text{D1C(2,3)} - (EA3 \times (\text{D1E2}/Q)) \]

\[ \text{D1C(2,4)} = \text{D1C(2,4)} + (LY \times G13) \]

\[ \text{D1C(2,5)} = \text{D1C(2,5)} + (EA23 \times LY \times G13) \]

\[ \text{D1C(2,6)} = (N1 + N3) / E2 \]

--

C GET 1ST DERIV. ASSOCIATED WITH HUSBANDS COVARIATES

C

DO 110 J = 1, DIM1

P1 = J + 2

\[ \text{D1B1(J)} = Z1(I,J) \times XE1 \times \text{DEXP(EA13 + A13)} \]

\[ \text{D1C(P1,1)} = \text{D1C(P1,1)} - ((\text{D1B1(J)}/Q) \times 2.0D0 \times G1) \]

\[ \text{D1C(P1,2)} = \text{D1C(P1,2)} - ((\text{D1B1(J)}/Q) \times G23) \]

\[ \text{D1C(P1,3)} = \text{D1C(P1,3)} - ((\text{D1B1(J)}/Q) \times EA3) \]

\[ \text{D1C(P1,4)} = \text{D1C(P1,4)} + (Z1(I,J) \times G12) \]

\[ \text{D1C(P1,5)} = \text{D1C(P1,5)} + (Z1(I,J) \times EA13 \times G12) \]
DO 120 K=1, DIM2
Pl=K+(DIM1+2)
D1B2(K)=22(I,K)*YE2*DEXP(EA23+A23)
D1C(P1,1)=D1C(P1,1)-
  ((D1B2(K)/Q)*2.0D0*G1)
D1C(P1,2)=D1C(P1,2)-((D1B2(K)/Q)*G23)
D1C(P1,3)=D1C(P1,3)-((D1B2(K)/Q)*EA3)
D1C(P1,4)=D1C(P1,4)+(Z2(I,K)*G13)
D1C(P1,5)=D1C(P1,5)+(Z2(I,K)*EA23*G13)
120 CONTINUE

DO 130 L=1, DIM3
Pl=L+(DIM1+DIM2+2)
D1B3(L)=Z3(I,L)*(YE2*DEXP(EA23+A23)+
  XE1*DEXP(EA13+A13))
D1C(P1,1)=D1C(P1,1)-
  ((D1B3(L)/Q)*2.0D0*G1)
D1C(P1,2)=D1C(P1,2)-((D1B3(L)/Q)*G23)
D1C(P1,3)=D1C(P1,3)-((D1B3(L)/Q)*EA3)
D1C(P1,4)=D1C(P1,4)+(Z3(I,L)*G13)
D1C(P1,5)=D1C(P1,5)+((Z3(I,L)*DEXP(A3))
  /(DEXP(A3)+1.0D0)*G1)
D1C(P1,6)=D1C(P1,6)+(Z3(I,L)*EA13*G12)
D1C(P1,7)=D1C(P1,7)+(Z3(I,L)*EA23*G13)
130 CONTINUE

IF (PLOT .EQ. 1) GO TO 1000

DO 140 I=1, DIM2
D1E1=(X(I)**(2*E1))*(LX**2)*DEXP(EA13+(2*A13))+(LX*D1E1)
D2E1=(D1E1/Q) - ((D1E1/Q)**2)
D2C(1,1,1)=D2C(1,1,1)-(D2E1*2.0D0*G1)
D2C(1,1,2)=D2C(1,1,2)-(D2E1*G23)
D2C(1,1,3)=D2C(1,1,3)-(EA3*D2E1)
D2C(1,1,4)=D2C(1,1,4)+(EA13*(LX**2)*G12)
D2C(1,1,5)=-(N1+N2)/(E1**2))

D2E2=(Y(I)**(2*E2))*(LY**2)*DEXP(EA23+(2*A23))+(LY*D1E2)
D2E2=(D2E2/Q) - ((D1E2/Q)**2)
D2C(2,2,1)=D2C(2,2,1)-(D2E2*2.0D0*G1)
D2C(2,2,2)=D2C(2,2,2)-(D2E2*G23)
D2C(2,2,3)=D2C(2,2,3)-(EA3*D2E2)
D2C(2,2,4) = D2C(2,2,4) + (EA23*(LY**2)*G13)
D2C(2,2,5) = (N1+N3)/(E2**2)

D2E12 = (-1.0D0)*(D1E1/Q)*(D1E2/Q)
D2C(2,1,1) = D2C(2,1,1) - (D2E12*G23)
D2C(2,1,2) = D2C(2,1,2) - (D2E12*EA3)

C--------------------------------------------------------------------------------------------------------------------------------
C GET 2ND DERIV. ASSOCIATED WITH HUSBANDS COVARIATES.
C--------------------------------------------------------------------------------------------------------------------------------

DO 210 J=1.DIM1
   PL=J+2
   T1=D1E1*(Z1(I,J)*(EA13+1.0D0))
   D2B1E1(J) = (T1/Q) - ((D1B1(J)/Q)*(D1E1/Q))
   D2C(PL,1,1) = D2C(PL,1,1) - (D2B1E1(J)*2.0D0)*G1
   D2C(PL,1,2) = D2C(PL,1,2) - (D2B1E1(J)*G23)
   D2C(PL,1,3) = D2C(PL,1,3) - (D2B1E1(J)*EA3)
   D2C(PL,1,4) = D2C(PL,1,4) + (Z1(I,J)*LX*EA13*G12)

D2B1E2(J) = (-1.0D0*D1E2*D1B1(J))/(Q**2)
D2C(PL,1,1) = D2C(PL,1,1) - (D2B1E2(J)*2.0D0)*G1
D2C(PL,2,1) = D2C(PL,2,1) - (D2B1E2(J)*G23)
D2C(PL,2,2) = D2C(PL,2,2) - (D2B1E2(J)*EA3)
D2C(PL,2,3) = D2C(PL,2,3) - (Z1(I,J)*LX*EA13*G12)

210 CONTINUE

DO 220 J=1.DIM1
   PL=J+2
   DO 220 J=J,PL
   T2=D1B1(J)*(Z1(I,J)*(EA13+1.0D0))
   D2B1(J,JP) = (T2/Q) - (D1B1(J)/Q)
   D2C(PL,JP,1) = D2C(PL,JP,1) - (D2B1(J,JP)*2.0D0)*G1
   D2C(PL,JP,2) = D2C(PL,JP,2) - (D2B1(J,JP)*G23)
   D2C(PL,JP,3) = D2C(PL,JP,3) - (D2B1(J,JP)*EA3)
   D2C(PL,JP,4) = D2C(PL,JP,4) + (Z1(I,J)*LX*EA13*G12)

220 CONTINUE

C--------------------------------------------------------------------------------------------------------------------------------
C GET 2ND DERIV. ASSOCIATED WITH WIFE'S COVARIATES.
C--------------------------------------------------------------------------------------------------------------------------------

DO 230 K=1.DIM2
   PL=K+(DIM1+2)
   T3=D1E2*(Z2(I,K)*(EA23+1.0D0))
   D2B2E2(K) = (T3/Q) - (D1B2(K)/Q)*(D1E2/Q)
D2C(P1, 2, 1) = D2C(P1, 2, 1) -
   (D2B2E2(K)*2.0D0*G1)
D2C(P1, 2, 2) = D2C(P1, 2, 2) - (D2B2E2(K)*G23)
D2C(P1, 2, 3) = D2C(P1, 2, 3) - (D2B2E2(K)*EA3)
D2C(P1, 2, 4) = D2C(P1, 2, 4) +
   (Z2(I, K)*LY*EA23*G13)
D2B2E1(K) = (-1.0D0)*(D1E1/Q)*(D1B2(K)/Q)
D2C(P1, 1, 1) = D2C(P1, 1, 1) -
   (D2B2E2(K)*2.0D0*G1)
D2C(P1, 1, 2) = D2C(P1, 1, 2) - (D2B2E1(K)*G23)
D2C(P1, 1, 3) = D2C(P1, 1, 3) - (D2B2E1(K)*EA3)
*Z2(I, K)*EA23*G13)

CONTINUE
DO 240 KP=1, DIM2
P2=KP*(DIM1+2)
   DO 240 K=KP, DIM2
      P1=K+(DIM1+2)
      T4=D1B2(K)*(Z2(I, KP)*(EA23+1.0D0))
      D2B2(K, KP) = (T4/Q) - (D1B2(K)/Q)*
         (D1B2(K)/Q)
      D2C(P1, P2, 1) = D2C(P1, P2, 1) -
         (D2B2(K, KP)*2.0D0*G1)
      D2C(P1, P2, 2) = D2C(P1, P2, 2) -
         (D2B2(K, KP)*G23)
      D2C(P1, P2, 3) = D2C(P1, P2, 3) -
         (D2B2(K, KP)*EA3)
      D2C(P1, P2, 4) = D2C(P1, P2, 4) + (Z2(I, K)
         *Z2(I, KP)*EA23*G13)

CONTINUE
DO 250 J=1, DIM1
P12=J+2
   DO 250 K=1, DIM2
      P1=K+(DIM1+2)
      D2B1B2(J, K) = (-1.0D0)*(D1B1(J)/Q)*
         (D1B2(K)/Q)
      D2C(P1, P12, 1) = D2C(P1, P12, 1) -
         (D2B1B2(J, K)*2.0D0*G1)
      D2C(P1, P12, 2) = D2C(P1, P12, 2) -
         (D2B1B2(J, K)*G23)
      D2C(P1, P12, 3) = D2C(P1, P12, 3) -
         (D2B1B2(J, K)*EA3)

CONTINUE

C-----------------------------------------------
C GET 2ND DERIV. ASSOCIATED WITH ASSOCIATION COVARIATES
C-----------------------------------------------

DO 260 L=1, DIM3
   P1=L+(DIM1+DIM2+2)
      T5=D1E1*(Z3(I, L)*(EA13+1.0D0))
      D2B3E1(L) = (T5/Q) - (D1B3(L)/Q)*(D1E1/Q)
      D2C(P1, 1, 1) = D2C(P1, 1, 1) -
         (D2B3E1(L)*2.0D0*G1)
      D2C(P1, 1, 2) = D2C(P1, 1, 2) - (D2B3E1(L)*G23)
D2C(P1, 1, 3) = D2C(P1, 1, 3) - (D2B3E1(L) * EA3) * (D1E1/Q))
D2C(P1, 1, 4) = D2C(P1, 1, 4) + (Z3(I, L) * EA3)
T6 = D1E2 * (Z3(I, L) * (EA23 + 1.0D0))
D2B3E2(L) = (T6/Q) - (D1B3(L)/Q) * (D1E2/Q)
D2C(P1, 1, 5) = D2C(P1, 1, 5) + ((Z3(I, L)* EA3)
* (LX) * G12)

D2B3E2(L) = (T6/Q) - (D1B3(L)/Q) * (D1E2/Q)

CONTINUE

DO 270 LP = 1, DIM3
P2 = LP + (DIM1 + DIM2 + 2)
DO 270 L = LP, DIM3
P1 = L + (DIM1 + DIM2 + 2)
T7 = YE2 * Z3(I, L) * (Z3(I, LP) *
(1.0DO) * DEXP(EA23 + A23) +
XE1 * Z3(I, L) * (Z3(I, LP) * 
(1.0DO) * DEXP(EA13 + A13))
D2B3(L, LP) = (T7/Q) - (D1B3(L)/Q) *
(D1B3(LP)/Q)
D2C(P1, P2, 1) = D2C(P1, P2, 1) -
(D2B3(L, LP) * 2.0D0 * G1)
D2C(P1, P2, 2) = D2C(P1, P2, 2) -
(D2B3(L, LP) * G23)
D2C(P1, P2, 3) = D2C(P1, P2, 3) -
(D2B3(L, LP) * EA3)
D2C(P1, P2, 4) = D2C(P1, P2, 4) + (Z3(I, LP) *
* EA3 * (D1B3(L)/Q))
D2C(P1, P2, 5) = D2C(P1, P2, 5) + (Z3(I, L) *
* EA3 * (D1B3(LP)/Q))
D2C(P1, P2, 6) = D2C(P1, P2, 6) - (Z3(I, L) *
* Z3(I, LP) * EA3 * LQ)
D2C(P1, P2, 7) = D2C(P1, P2, 7) + (Z3(I, L) *
* Z3(I, LP) * EA13 * G12)
D2C(P1, P2, 8) = D2C(P1, P2, 8) + (Z3(I, L) *
* Z3(I, LP) * EA23 * G13)
D2C(P1, P2, 9) = D2C(P1, P2, 9) + ((Z3(I, L) *
* Z3(I, LP) * DEXP(A3))/((
DEXP(2.0DO * A3) + (2.0DO *
DEXP(A3)) + 1.0DO)) * G1)

CONTINUE

DO 280 J = 1, DIM1
P12 = J + 2
DO 280 L = 1, DIM3
P1 = L + (DIM1 + DIM2 + 2)
T8 = D1B1(J) * (Z3(I, L) * (EA13 + 1.0DO))
D2B1B3(J,L) = (T8/Q) - (D1B1(J)/Q) * (D1B3(L)/Q) 
D2C(P1,P12,1) = D2C(P1,P12,1) - (D2B1B3(J,L) * 2.0D0 * G1) 
D2C(P1,P12,2) = D2C(P1,P12,2) - (D2B1B3(J,L) * G23) 
D2C(P1,P12,3) = D2C(P1,P12,3) - (D2B1B3(J,L) * EA3) 
D2C(P1,P12,4) = D2C(P1,P12,4) + (Z3(I,L) * EA3) 
D2C(P1,P12,5) = D2C(P1,P12,5) + (Z3(I,L) * Z2(I,K) * EA23 * G13)

280 CONTINUE
DO 290 K = 1, DIM2
  P13 = K + (DIM1 + 2)
  DO 290 L = 1, DIM3
    T9 = D1B2(K) * (Z3(I,L) * (EA23 + 1.0D0))
    D2B2B3(K,L) = (T9/Q) - (D1B2(K)/Q) * (D1B3(L)/Q) 
    D2C(P1,P13,1) = D2C(P1,P13,1) - (D2B2B3(K,L) * 2.0D0 * G1) 
    D2C(P1,P13,2) = D2C(P1,P13,2) - (D2B2B3(K,L) * G23) 
    D2C(P1,P13,3) = D2C(P1,P13,3) - (D2B2B3(K,L) * EA3) 
    D2C(P1,P13,4) = D2C(P1,P13,4) + (Z3(I,L) * EA3) 
    D2C(P1,P13,5) = D2C(P1,P13,5) + (Z3(I,L) * Z2(I,K) * EA23 * G13)
 290 CONTINUE

1000 CONTINUE
C-------------------------------------------------------------------------------------------------------------------
C NOW COMPUTE THE LOGLIKELIHOOD (LL1)
C AND
C THE FIRST DERIV'S : DL(I) AND,
C THE SECOND DERIV'S : D2L(I)
C-------------------------------------------------------------------------------------------------------------------

LOGL = 0.0D0
DO 1500 I = 1, 20
  DL(I) = 0.0D0
  DO 1500 J = 1, 20
    D2L(I,J) = 0.0D0
1500 CONTINUE
DO 2000 I = 1, 10
  LOGL = LOGL + LC(I)
  DO 2000 J = 1, DIM1 + DIM2 + DIM3 + 2
DL(J) = DL(J) + D1C(J, I)
DO 2000 K = 1, DIM1 + DIM2 + DIM3 + 2
   D2L(J, K) = D2L(J, K) + D2C(J, K, 1)
   D2L(K, J) = D2L(J, K)
2000 CONTINUE
   CALL OVERFL(IND)
   IF (IND .NE. 2) GO TO 4000
   RETURN
3000 CONTINUE
   ITORER = 2
   RETURN
4000 CONTINUE
   ITORER = 3
   RETURN
   END
   C
   C
   SUBROUTINE LKDER(NP, LOGL, DL, D2L, DIM, N, 
   PLOT, ITORER, DER)
   C
   ; THIS SUBROUTINE USES THE LEE-KLEIN
   ; LIKELIHOOD FUNCTION WITH BURR MARGINAL DISTNS.
   ; THREE SETS OF COVARIATES ARE PERMITTED IN THIS
   ; MODEL: B1 AND B2 ARE INTRODUCED INTO THE
   ; LOCATION PARAMETERS, B3 IS INTRODUCED INTO THE
   ; DEPENDENCE PARAMETER. B1, B2, AND B3 ARE THE
   ; ESTIMATED COEFFICIENTS FOR THE CORRESPONDING
   ; COVARIATES Z1, Z2, AND Z3.
   COMMON Z1, Z2, Z3, X, Y, IX, IY
   INTEGER IX(4531), IY(4531), 
   * I, J, K, L, N, PLOT, ITORER, DER, 
   * DIM1, DIM2, DIM3, DIM(3), 
   * P1, P2, P12, P13
   REAL*8 X(4531), Y(4531), 
   * Z1(4531, 6), Z2(4531, 9), Z3(4531, 3), 
   * NP(20), 
   * B1(6), B2(9), B3(6), E1, E2, 
   * N1, N2, N3, N4, 
   * XE1, YE2, 
   * Q, LQ, LX, LY, 
   * A1, A2, A3, 
   * EA1, EA2, EA3, 
   * D1E1, D1E2, 
   * D2E1, D2E2, D2E12, 
   * D1B1(6), D1B2(9).
* D2B1E1(6), D2B2E1(9),
* D2B1E2(6), D2B2E2(9),
* D2B1(6,6), D2B1B2(6,9),
* D2B2(9,9)

REAL*8 G1, G2, G3, G4, G12, G13, G23, G123,
* LC(10), D1C(20, 10),
* D2C(20, 20, 10),
* LOGL, DL(20),
* D2L(20, 20)

CALL ERRSET(207, 0, -1, 1, 0, 208)
IND = 2
CALL OVERFL(IND)

C-----------------------------------------------
C INITIALIZE A FEW SUMMING VARIABLES.
C-----------------------------------------------
N1 = 0. 0D0
N2 = 0. 0D0
N3 = 0. 0D0
N4 = 0. 0D0
DIM1 = DIM(1)
DIM2 = DIM(2)
DIM3 = DIM(3)

C-----------------------------------------------
C ENTER THE ITERATIVE LOOP OF CALCULATIONS INVOLVING
C THE PARAMETERS TO BE ESTIMATED.
C INITIALIZE THE SUMMING VARIABLES AND MATRICES.
C-----------------------------------------------
DO 1 I = 1, 10
   LC(I) = 0. 0D0
   DO 1 J = 1, 20
      D1C(J, I) = 0. 0D0
      DO 1 K = 1, 20
         D2C(K, J, I) = 0. 0D0
      CONTINUE
1  CONTINUE
E1 = NP(1)
E2 = NP(2)
IF (E1 .LE. 0. 0D0 .OR. E1 .GT. 174. 6) GO TO 3000
IF (E2 .LE. 0. 0D0 .OR. E2 .GT. 174. 6) GO TO 3000
DO 3 I = 1, DIM1
   B1(I) = NP(I + 2)
3  CONTINUE
DO 4 J = 1, DIM2
   B2(J) = NP(J + 2 + DIM1)
4  CONTINUE
DO 5 L = 1, DIM3
   B3(L) = NP(L + 2 + DIM1 + DIM2)
5  CONTINUE
C
C
C
C   PASS THROUGH THE DATA. CALCULATE THE BASIC QUANTITY Q, AND THE DERIV. QUANTITIES TO BE SUMMED.

DO 1000 I=1,N
   G1=0.0DO
   G2=0.0DO
   G3=0.0DO
   G4=0.0DO
   G12=0.0DO
   G13=0.0DO
   G23=0.0DO
   G123=0.0DO
   IF (IX(I).EQ.1 AND. IY(I).EQ.1) G1=1.0DO
   IF (IX(I).EQ.1 AND. IY(I).EQ.0) G2=1.0DO
   IF (IX(I).EQ.0 AND. IY(I).EQ.1) G3=1.0DO
   IF (IX(I).EQ.0 AND. IY(I).EQ.0) G4=1.0DO
   IF (G1+G2 .EQ. 1.0DO) G12=1.0DO
   IF (G1+G3 .EQ. 1.0DO) G13=1.0DO
   IF (G2+G3 .EQ. 1.0DO) G23=1.0DO
   IF (G1+G2+G3 .EQ. 1.0DO) G123=1.0DO
   N1=N1+G1
   N2=N2-G2
   N3=N3-G3
   N4=N4-G4
   A1=0.0DO
   A2=0.0DO
   A3=0.0DO
   DO 10 J=1,DIM1
      A1=A1+(B1(J)*Z1(I,J))
   10 CONTINUE
   DO 20 K=1,DIM2
      A2=A2+(B2(K)*Z2(I,K))
   20 CONTINUE
   DO 30 L=1,DIM3
      A3=A3+(B3(L)*Z3(I,L))
   30 CONTINUE
   IF (DABS(A1).GT. 174.6) GO TO 3000
   IF (DABS(A2).GT. 174.6) GO TO 3000
   IF (DABS(A3) .GT. 10.0DO) GO TO 3000
   XE1=X(I)**E1
   YE2=Y(I)**E2
   EA1=DEXP(A1)*XE1
   EA2=DEXP(A2)*YE2
   EA3=DEXP(A3)
   LX=DLOG(X(I))
   LY=DLOG(Y(I))
   Q=1.0DO+EA1+EA2
   LQ=DLOG(Q)
C  CALCULATE THE COMPONENTS OF THE LOGLIKELIHOOD.

LC(1) = LC(1) - (LQ*2.0D0*G1)
LC(2) = LC(2) - (LQ*G23)
LC(3) = LC(3) - (EA1*LQ)
LC(4) = LC(4) + (A3*G123)
LC(5) = LC(5) + (A1*G12)
LC(6) = LC(6) + (A2*G13)
LC(7) = LC(7) + ((E1-1.0D0)*LX*G12) +
      (DLOG(E1)*G12)
LC(8) = LC(8) + ((E2-1.0D0)*LY*G13) +
      (DLOG(E2)*G13)
LC(9) = LC(9) + (DLOG(EA3+1.0D0))*G1
IF (DER .EQ. 0) GO TO 1000

C  CALCULATE 1ST DERIVATIVES W.R.T. WEIBULL SHAPE PARMS.

D1E1 = (EA1*LX)
D1C(1,1) = D1C(1,1) - (((D1E1/Q)*2.0D0*G1)
D1C(1,2) = D1C(1,2) - (((D1E1/Q)*G23)
D1C(1,3) = D1C(1,3) - (EA1*(D1E1/Q))
D1C(1,4) = D1C(1,4) + (LX*G12)
D1C(1,5) = (N1+N2)/E1

D1E2 = (EA2*LY)
D1C(2,1) = D1C(2,1) - (((D1E2/Q)*2.0D0*G1)
D1C(2,2) = D1C(2,2) - (((D1E2/Q)*G23)
D1C(2,3) = D1C(2,3) - (EA2*(D1E2/Q))
D1C(2,4) = D1C(2,4) + (LY*G13)
D1C(2,5) = (N1+N3)/E2

C  GET 1ST DERIV. ASSOCIATED WITH HUSBANDS COVARIATES.

DO 110 J=1,DIM1
   P1 = J+2
   D1B1(J) = Z1(I,J)*EA1
   D1C(P1,1) = D1C(P1,1) -
                (((D1B1(J)/Q)*2.0D0*G1)
   D1C(P1,2) = D1C(P1,2) - (((D1B1(J)/Q)*G23)
   D1C(P1,3) = D1C(P1,3) - ((D1B1(J)/Q)*EA3)
   D1C(P1,4) = D1C(P1,4) + (Z1(I,J)*G12)
110 CONTINUE

C  GET 1ST DERIV. ASSOCIATED WITH WIFE'S COVARIATES.

DO 120 K=1,DIM2
   P1 = K+(DIM1+2)
   D1B2(K) = Z2(I,K)*EA2
   D1C(P1,1) = D1C(P1,1) -
                (((D1B2(K)/Q)*2.0D0*G1)
   D1C(P1,2) = D1C(P1,2) - (((D1B2(K)/Q)*G23)
D1C(P1,3) = D1C(P1,3) - ((D1B2(K)/Q)*EA3)
D1C(P1,4) = D1C(P1,4) + (Z2(I,K)*G13)

120 CONTINUE

C GET 1ST DERIV. ASSOCIATED WITH ASSOCIATION COVARIATE 1

DO 130 L = 1, DIM3
    P1 = L + (DIM1 + DIM2 + 2)

    D1B3(L) =
    D1C(P1,1) = D1C(P1,1) - (LQ*Z3(I,L)*EA3)
    D1C(P1,2) = D1C(P1,2) + (Z3(I,L)*G123)
    D1C(P1,3) = D1C(P1,3) + ((Z3(I,L)*EA3)/
        (EA3 + 1.0D0))*G1

130 CONTINUE

IF (PLOT .EQ. 1) GO TO 1000

C GET 2ND DERIVATIVES W.R.T. WEIBULL SHAPE PARMS.

D2E1 = (D1E1/Q)*(LX-(D1E1/Q))
D2C(1,1,1) = D2C(1,1,1) - (D2E1*2.0D0*G1)
D2C(1,1,2) = D2C(1,1,2) - (D2E1*G23)
D2C(1,1,3) = D2C(1,1,3) - (EA3*D2E1)
D2C(1,1,4) = -((N1+N2)/(E1**2))

D2E2 = (D1E2/Q)*(LY-(D1E2/Q))
D2C(2,2,1) = D2C(2,2,1) - (D2E2*2.0D0*G1)
D2C(2,2,2) = D2C(2,2,2) - (D2E2*G23)
D2C(2,2,3) = D2C(2,2,3) - (EA3*D2E2)
D2C(2,2,4) = -((N1+N3)/(E2**2))

D2E12 = (-1.0D0)*(D1E1/Q)*(D1E2/Q)
D2C(2,1,1) = D2C(2,1,1) - (D2E12*2.0D0*G1)
D2C(2,1,2) = D2C(2,1,2) - (D2E12*G23)
D2C(2,1,3) = D2C(2,1,3) - (EA3*D2E12)

C GET 2ND DERIV. ASSOCIATED WITH HUSBANDS COVARIATES.

DO 210 J = 1, DIM1
    P1 = J + 2

    D2B1E1(J) = (D1E1/Q)*(Z1(I,J)-(D1B1(J)/Q))
    D2C(P1,1,1) = D2C(P1,1,1) -
        (D2B1E1(J)*2.0D0*G1)
    D2C(P1,1,2) = D2C(P1,1,2) - (D2B1E1(J)*G23)
    D2C(P1,1,3) = D2C(P1,1,3) - (D2B1E1(J)*EA3)

    D2B1E2(J) = -1.0D0*(D1E2/Q)*(D1B1(J)/Q)
    D2C(P1,2,1) = D2C(P1,2,1) -
        (D2B1E2(J)*2.0D0*G1)
    D2C(P1,2,2) = D2C(P1,2,2) - (D2B1E2(J)*G23)
    D2C(P1,2,3) = D2C(P1,2,3) - (D2B1E2(J)*EA3)
CONTINUE
DO 220 JP=1,DIM1
   P2=JP+2
   DO 220 J=JP,DIM1
      Pl=J+2
      D2B1(J,JP)=(D1B1(J)/Q)*(Z1(I,JP)-(D1B1(JP)/Q))
      D2C(P1,P2,1)=D2C(P1,P2,1)-(D2B1(J,JP)*2.0D0*G1)
      D2C(P1,P2,2)=D2C(P1,P2,2)-(D2B1(J,JP)*G23)
      D2C(P1,P2,3)=D2C(P1,P2,3)-(D2B1(J,JP)*EA3)
220 CONTINUE
C--------------------------------------------------------------------------------------------------------------------------------
C GET 2ND DERIV. ASSOCIATED WITH WIFE'S COVARIATES.
C--------------------------------------------------------------------------------------------------------------------------------
DO 230 K=1,DIM2
   P1=1+(DIM1+2)
   D2B2E2(K)=(D1E2/Q)*(Z2(I,K)-(D1B2(K)/Q))
   D2C(P1,2,1)=D2C(P1,2,1)-(D2B2E2(K)*2.0D0*G1)
   D2C(P1,2,2)=D2C(P1,2,2)-(D2B2E2(K)*G23)
   D2C(P1,2,3)=D2C(P1,2,3)-(D2B2E2(K)*EA3)
   D2B2E1(K)=-1.0D0*(D1E1/Q)*(D1B2(K)/Q)
   D2C(P1,1,1)=D2C(P1,1,1)-(D2B2E1(K)*2.0D0*G1)
   D2C(P1,1,2)=D2C(P1,1,2)-(D2B2E1(K)*G23)
   D2C(P1,1,3)=D2C(P1,1,3)-(D2B2E1(K)*EA3)
230 CONTINUE
DO 240 KP=1,DIM2
   P2=KP+(DIM1+2)
   DO 240 K=1,DIM2
      Pl=K+(DIM1+2)
      D2B2(K,KP)=(D1B2(K)/Q)*(Z2(I,KP)-(D1B2(KP)/Q))
      D2C(P1,2,1)=D2C(P1,2,1)-(D2B2(K,KP)*2.0D0*G1)
      D2C(P1,2,2)=D2C(P1,2,2)-(D2B2(K,KP)*G23)
      D2C(P1,2,3)=D2C(P1,2,3)-(D2B2(K,KP)*EA3)
240 CONTINUE
DO 250 J=1,DIM1
   P12=J+2
   DO 250 K=1,DIM2
      Pl=K+(DIM1+2)
      D2B1B2(J,K)=-1.0D0*(D1B1(J)/Q)*(D1B2(K)/Q)
      D2C(P1,P12,1)=D2C(P1,P12,1)-(D2B1B2(J,K)*2.0D0*G1)
      D2C(P1,P12,2)=D2C(P1,P12,2)-(D2B1B2(J,K)*G23)
      D2C(P1,P12,3)=D2C(P1,P12,3)-(D2B1B2(J,K)*EA3)
250 CONTINUE
* \((D_2B_1B_2(J,K) \times 2.0D0 \times G_1)\)
\(D_2C(P_1,P_12,2) = D_2C(P_1,P_12,2) - \)
* \((D_2B_1B_2(J,K) \times G_{23})\)
\(D_2C(P_1,P_12,3) = D_2C(P_1,P_12,3) - \)
* \((D_2B_1B_2(J,K) \times EA_{3})\)

250 CONTINUE

C--------------------------------------------------------------------------------------------------------------------------------
C GET 2ND DERIV. ASSOCIATED WITH ASSOCIATION COVARS.
C--------------------------------------------------------------------------------------------------------------------------------
DO 260 L=1, DIM3
P1=L+(DIM1+DIM2+2)
C D2B3E1(L)
\(D_2C(P_1,1,1) = D_2C(P_1,1,1) - \)
* \(((D_1E_1/Q) \times Z_3(I,L) \times EA_{3})\)
C D2B3E2(L)
\(D_2C(P_1,2,1) = D_2C(P_1,2,1) - \)
* \(((D_1E_2/Q) \times Z_3(I,L) \times EA_{3})\)
260 CONTINUE
DO 270 LP=1,DIM3
P2=LP+(DIM1+DIM2+2)
DO 270 L=LP, DIM3
P1=L+(DIM1+DIM2+2)
C D2B3(L,LP)
\(D_2C(P_1,P_2,1) = D_2C(P_1,P_2,1) - \)
* \(((LQ \times Z_3(I,L) \times Z_3(I,LP) \times EA_{3})/(DEXP(2.0D0 \times A_{3}) \times Z_3(I,L) \times Z_3(I,LP) \times EA_{3})/(DEXP(2.0D0 \times A_{3}) + 1.0D0))\times G_1\)
270 CONTINUE
DO 280 J=1,DIM1
P12=J+2
DO 280 L=1,DIM3
P1=L+(DIM1+DIM2+2)
C D2B1B3(J,L)
\(D_2C(P_1,P_12,1) = D_2C(P_1,P_12,1) - \)
* \(((D_1B_1(J)/Q) \times Z_3(I,L) \times EA_{3})\)
280 CONTINUE
DO 290 K=1,DIM2
P13=K+(DIM1+2)
DO 290 L=1,DIM3
P1=L+(DIM1+DIM2+2)
C D2B2B3(K,L)
\(D_2C(P_1,P_13,1) = D_2C(P_1,P_13,1) - \)
* \(((D_1B_2(K)/Q) \times Z_3(I,L) \times EA_{3})\)
290 CONTINUE
1000 CONTINUE
C--------------------------------------------------------------------------------------------------------------------------------
C NOW COMPUTE THE LOGLIKELIHOOD (LL1)
C AND
C THE FIRST DERIV'S : DL(I) AND,
THE SECOND DERIV'S : D2L(I)

```c
LOGL=0.0DO
DO 1500 I=1,20
   DL(I)=0.0DO
   DO 1500 J=1,20
      D2L(I,J)=0.0DO
1500 CONTINUE
DO 2000 I=1,10
   LOGL=LOGL+LC(I)
   DO 2000 J=1,DIM1+DIM2+DIM3+2
      DL(J)=DL(J)+D1C(J,I)
   DO 2000 K=1,DIM1+DIM2+DIM3+2
      D2L(J,K)=D2L(J,K)+D2C(J,K,I)
      D2L(K,J)=D2L(J,K)
2000 CONTINUE
CALL OVERFL(IND)
IF (IND .NE. 2) GO TO 4000
RETURN
3000 CONTINUE
ITORER=2
RETURN
4000 CONTINUE
ITORER=3
RETURN
END
```

```
GO.SYSIN DD DSN=TS2798.PAGE3,
   DISP=SHR

EXEC WNOTIFY
```
BIBLIOGRAPHY


- 149 -


37. Shurtleff, D. *Some Characteristics Related to the Incidence of Cardiovascular Disease and Death: Framingham Study, 18-year*


