This book aims at contributing to the study of chronic diseases and the debate on compression versus expansion of morbidity by applying a new methodology that integrates demography and epidemiology and by deriving indicators of cardiovascular disease history. The long-term impact of four risk factors on cardiovascular disease is investigated using the 48 years follow-up of the original Framingham Heart Study cohort. Four disease types are distinguished: coronary heart disease (CHD), myocardial infarction (MI), congestive heart failure (CHF) and stroke. The risk factors that are studied are smoking, blood pressure, body mass index, and serum cholesterol. These risk factors at middle age are strong predictors of the incidence of cardiovascular disease and mortality at older age. The study shows that a healthy life style postpones cardiovascular disease and may reduce the lifetime risk of the disease.

The human life course offers an exciting new paradigm for multidisciplinary scientific research. The study of individual disease histories and risk factors during different stages of life leads to new insights in the impact of lifestyle on health at older age. The adoption of the new paradigm necessitates longitudinal data and appropriate data analytic techniques. This book demonstrates that a combination of the multistate life table and multivariate techniques of event history modeling provides an effective way to describe, explain and predict disease histories.

Abdullah Al Mamun received MSc in Statistics, University of Dhaka, and MSc in Population Studies, University of Groningen. He worked as a senior research officer (1995-1999) at International Centre for Diarrhoeal Disease Research of Bangladesh (ICDDR,B). On the basis of this book he holds PhD in Demography from the University of Groningen, The Netherlands. Recently he was appointed as a Research Fellow at the School of Population Health, University of Queensland Medical School, Australia.
Life History of Cardiovascular Disease and Its Risk Factors
The book series Population Studies aims at disseminating results of research on population trends, in the broadest sense.

Series editorial board: Melinda Mills, Anton Oskamp & Harrie van Vianen.
In memory of Anton Kuijsten.

ISBN 90 5170 720 7
NUR 740

© Abdullah Al Mamun, 2003

Cover design: PuntSpatie, Amsterdam, Abdullah Al Mamun

Cover picture: The cover picture showing the distribution of arteries leading from the heart muscle. This drawing picture was an artistic grace and intelligent observation of human heart by Leonardo da Vinci.

All rights reserved. Save exceptions stated by the law, no part of this publication may be reproduced, stored in a retrieval system of any nature, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, included a complete or partial transcription, without the prior written permission of the publisher, application for which should be addressed to the publisher: Rozenberg Publishers, Rozengracht 176A, 1016 NK Amsterdam. E-mail: info@rozenbergps.com
Life History of Cardiovascular Disease and Its Risk Factors

Multistate Life Table Approach and Application to the Framingham Heart Study

Proefschrift

ter verkrijging van het doctoraat in de Ruimtelijke Wetenschappen aan de Rijksuniversiteit Groningen op gezag van de Rector Magnificus, dr. F. Zwarts, in het openbaar te verdedigen op maandag 30 juni 2003 om 14.15 uur

door

Abdullah Al Mamun

geboren op 1 januari 1970
te Dhaka, Bangladesh
Promotores: Prof. dr. ir. F.J. Willekens
Prof. dr. J. P. Mackenbach

Beoordelingscommissie: Prof. dr. ir. D. Kromhout
Prof. dr. D. Post
Prof. dr. L.J.G. van Wissen
To my parents & my small one Mehjabin
# Table of contents

List of figures
List of tables
Preface

## 1 General introduction

1.1 Introduction
1.2 The changes in mortality and morbidity
1.3 Integration of morbidity and mortality: a general framework
1.4 Setting the scene
  1.4.1 Cardiovascular disease
  1.4.2 Cardiovascular disease risk factors
  1.4.3 Basic model of cardiovascular disease
1.5 Research objectives
1.6 Outline of this book
References

## 2 Multistate life tables in public health

Abstract
2.1 Introduction
2.2 Prevalence and incidence measures
  2.2.1 Prevalence
  2.2.2 Incidence
  2.2.3 Prevalence versus incidence
2.3 The Sullivan method
2.4 The multistate life table
  2.4.1 Conceptual issues
  2.4.2 The Markov model
  2.4.3 The equations
  2.4.4 The introduction of covariates
  2.4.5 Confidence interval estimation
  2.4.6 Input data for the multistate life table
  2.4.7 Output from the multistate life table
2.5 A critical assessment of the multistate life table
References
3 A multistate life table analysis of cardiovascular disease life history

Abstract
3.1 Introduction
3.2 General approach
3.3 Data and methods
   3.3.1 Overview of the Framingham Heart Study
   3.3.2 States and events
   3.3.3 Estimation of occurrence-exposure rates
   3.3.4 Multistate life table construction
3.4 Results
   3.4.1 Observed event occurrences in the FHS
   3.4.2 Males and females combined
   3.4.3 Male-female differences
3.5 Discussion and conclusion
References
Technical Appendix

4 Handling missing data in risk factors in repeated measurements

Abstract
4.1 Introduction
4.2 The missing data mechanism
   4.2.1 Substantive issues
   4.2.2 Problems
4.3 Missing data in the FHS
   4.3.1 Total missing cases in the FHS
   4.3.2 Missing risk factors
4.4 Missing value imputation
   4.4.1 Smoking status imputation
   4.4.2 Systolic blood pressure imputation
4.5 Validation of the methods
4.6 Discussion
References
## 5 Smoking career

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>128</td>
</tr>
<tr>
<td><strong>5.1 Introduction</strong></td>
<td>128</td>
</tr>
<tr>
<td><strong>5.2 Methods</strong></td>
<td>130</td>
</tr>
<tr>
<td>5.2.1 Data source</td>
<td>130</td>
</tr>
<tr>
<td>5.2.2 Smoking status</td>
<td>131</td>
</tr>
<tr>
<td>5.2.3 Missing value imputation</td>
<td>131</td>
</tr>
<tr>
<td>5.2.4 Model specification and input data</td>
<td>132</td>
</tr>
<tr>
<td>5.2.5 Multistate life table construction</td>
<td>133</td>
</tr>
<tr>
<td><strong>5.3 Results</strong></td>
<td>134</td>
</tr>
<tr>
<td>5.3.1 Changes in smoking status</td>
<td>135</td>
</tr>
<tr>
<td>5.3.2 Smoking episodes</td>
<td>135</td>
</tr>
<tr>
<td>5.3.3 Transition rates</td>
<td>136</td>
</tr>
<tr>
<td>5.3.4 Survival probabilities</td>
<td>138</td>
</tr>
<tr>
<td>5.3.5 Life expectancy</td>
<td>140</td>
</tr>
<tr>
<td>5.3.6 Lifetime probability of quitting smoking</td>
<td>141</td>
</tr>
<tr>
<td>5.3.7 Person-years lived by smoking status</td>
<td>142</td>
</tr>
<tr>
<td><strong>5.4 Discussion</strong></td>
<td>147</td>
</tr>
<tr>
<td>References</td>
<td>143</td>
</tr>
</tbody>
</table>

## 6 Life history of cardiovascular disease

**A comparison of smokers and non-smokers**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>149</td>
</tr>
<tr>
<td><strong>6.1 Introduction</strong></td>
<td>150</td>
</tr>
<tr>
<td><strong>6.2 Data and method</strong></td>
<td>150</td>
</tr>
<tr>
<td><strong>6.3 Results</strong></td>
<td>153</td>
</tr>
<tr>
<td>6.3.1 Disease incidence and death</td>
<td>153</td>
</tr>
<tr>
<td>6.3.2 Life table outcomes</td>
<td>154</td>
</tr>
<tr>
<td><strong>6.4 Discussion</strong></td>
<td>161</td>
</tr>
<tr>
<td>References</td>
<td>163</td>
</tr>
<tr>
<td>Technical Appendix</td>
<td>164</td>
</tr>
</tbody>
</table>
# 7 Cox and Gompertz regression models

An assessment with empirical estimates

## Abstract

### 7.1 Introduction

### 7.2 Models and data

#### 7.2.1 The Cox regression model

#### 7.2.2 The Gompertz regression model

#### 7.2.3 Data source

### 7.3 Transition rates comparison

#### 7.3.1 Model without covariates (null model)

#### 7.3.2 Model with covariates

### 7.4 Life table outcomes

### 7.5 Discussion

## References

# 8 A multivariate analysis of risk factors of cardiovascular disease and its subtypes

## Abstract

### 8.1 Introduction

### 8.2 Data and methods

#### 8.2.1 Data source

#### 8.2.2 Risk factor definition

#### 8.2.3 Estimation of relative risk and model specification

### 8.3 Results

#### 8.3.1 Disease incidence

#### 8.3.2 Death

### 8.4 Discussion

## References

## Technical Appendix
9 Risk factors and compression of cardiovascular morbidity

Abstract

9.1 Introduction

9.2 Data and methods

9.2.1 Data and risk factor exposure

9.2.2 State space and model

9.2.3 Estimation of transition rates

9.2.4 Life table construction

9.3 Results

9.3.1 Univariate analysis

9.3.2 Multivariate analysis

9.4 Discussion

References

10 Discussion and conclusion

10.1 Introduction

10.2 Summary of findings

10.3 Major contribution of this research

10.4 Implications for public health

10.5 Some future prospects

10.6 Concluding remarks

References

Samenvatting
List of figures

Figure 1.1  Mortality (hypothetical), morbidity (hypothetical) and disability (hypothetical) survival curves
Figure 1.2  Established adult risk factors
Figure 1.3  Development of cardiovascular disease
Figure 1.4  Basic (multistate life table) model structure of cardiovascular disease
Figure 2.1  Conceptualization of incidence versus prevalence measurement
Figure 2.2  Interrelationship of incidence, prevalence and mortality
Figure 2.3  Lexis diagram
Figure 3.1  Multistate life table model of cardiovascular disease
Figure 3.2  Estimation of occurrence-exposure based on the observational plan
Figure 3.3  Implied prevalence of disease in the life table population
Figure 3.4  Survival of a cohort of cardiovascular disease-free 40-year-old population
Figure 3.5  Implied prevalence of a history of cardiovascular disease by sex
Figure 3.6  Survival probabilities free of cardiovascular disease by sex
Figure 3.7  Life table percentage of lifetime spent in a state by sex
Figure 3.8  Male-female differences in the person years lived in each disease state in the age interval $x$ to $x+1$
Figure 4.1  Missing cases (%) in given exam
Figure 4.2  Observed mean of blood pressure by age, FHS
Figure 4.3  Scenarios of the age specific mean values of blood pressure
Figure 5.1  Model specification
Figure 5.2  Transition rates by smoking status
Figure 5.3  Survival probability in a non-smoking state and survival probability in a smoking state of a cohort of 10-year-old males and females of FHS
Figure 5.4  Differences of the person-years lived: non-smokers – smokers
Figure 6.1  Survival curves illustrating the probability of surviving and surviving free of cardiovascular disease (CVD), coronary heart disease (CHD), myocardial infarction (MI), or stroke
Figure 6.2  Difference in life years lived with cardiovascular disease, coronary heart disease, myocardial infarction, stroke, or congestive heart failure, by age (smokers – never smokers)
Figure 7.1  Comparison of age-specific death rates: Cox, Gompertz and observed occurrence-exposure rates, by sex
Figure 7.2  Comparison of age-specific rates of transition from NO-CVD to CVD: Cox, Gompertz and observed occurrence-exposure rates, by sex
Figure 7.3  Comparison of age-specific rates of transition NO-CVD to death: Cox, Gompertz and observed occurrence-exposure rates, by sex
Figure 7.4  Comparison of age-specific death rates: Cox, Gompertz and observed occurrence-exposure rates, by sex and education

Figure 7.5  Comparison of age-specific rates of transition NO-CVD to CVD: Cox, Gompertz and observed occurrence-exposure rates, by sex and education

Figure 7.6  Comparison of age-specific transition rates of NO-CVD to death: Cox, Gompertz and observed occurrence-exposure rates, by sex and education

Figure 7.7  Survival curves illustrating the probability of surviving free of cardiovascular disease by age and educational status

Figure 7.8  Age-specific probability of cardiovascular disease free people at age 50 developing cardiovascular disease, by sex and education

Figure 9.1  The differences in the number of years lived with cardiovascular disease between optimal risk and high risk category (difference: optimal – high)

Figure 9.2  Survival curves illustrating the probability of surviving and surviving free of cardiovascular disease (CVD), coronary heart disease (CHD), and myocardial infarction (MI), by sex and risk profile

Figure 9.3  Differences of the number of years lived with diseases: optimal risk profiles – high risk profiles
List of tables

Table 2.1  Prevalence vs. incidence
Table 3.1  Framingham Heart Study original cohort: age-sex distribution
Table 3.2  Number of outcomes studied in FHS
Table 3.3  Life expectancy and residual life expectancy free of disease at specified ages, based on a population free of cardiovascular disease at age 40, FHS
Table 3.4  Lifetime risk of developing cardiovascular disease (%) for cardiovascular disease-free individuals at age 40
Table 4.1  An example of repeated measurement data (a hypothetical example of cholesterol level)
Table 4.2  Total number of exams missing
Table 4.3  Missing cases in given exam
Table 4.4  Smoking status by exam, FHS
Table 4.5  Systolic blood pressure by exam, FHS
Table 4.6  Smoking status of an individual
Table 4.7  Smoking status before and after imputation, FHS
Table 4.8  Blood pressure level of an individual before and after imputation, FHS
Table 4.9  Scenarios for smoking status- responses are imputed after random deletion
Table 4.10  Scenarios for blood pressure- responses are imputed after random deletion
Table 5.1  Smoking prevalence (%) at different time points of follow-up, FHS
Table 5.2  Total number of smoking and non-smoking episodes based on biannual reporting of smoking status over 48 years of follow-up of FHS
Table 5.3  Number of years that a person of a given age and sex may expect to spend in each smoking status, FHS
Table 5.4  Probability of quitting smoking
Table 6.1  Risk of cardiovascular disease (including sudden death) by smoking status, relative to never smokers (95% confidence intervals in parentheses)
Table 6.2  Risk of death (from given states) by smoking status, relative to never smokers (95% confidence intervals in parentheses)
Table 6.3  Risk of developing cardiovascular disease before age 70 or death for any person free of CVD at age 50. 95% confidence intervals are presented in parentheses
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4</td>
<td>The burden of cardiovascular disease in always smokers versus never smokers free of any cardiovascular disease at age 50. Life years lived with a history of cardiovascular disease. 95% confidence intervals are presented in parentheses</td>
</tr>
<tr>
<td>7.1</td>
<td>Lifetime probability that cardiovascular disease free people at age 50 develop cardiovascular disease, by sex and education (percentage)</td>
</tr>
<tr>
<td>7.2</td>
<td>Total life expectancy (SDLT) and life expectancy free of cardiovascular disease (MSLT)</td>
</tr>
<tr>
<td>8.1</td>
<td>Categorization of the risk factor exposures</td>
</tr>
<tr>
<td>8.2</td>
<td>Relative risk of cardiovascular disease and its subtypes (including sudden death) by single risk factor (95% confidence intervals in parentheses) during age 30-50</td>
</tr>
<tr>
<td>8.3</td>
<td>Relative risk of cardiovascular disease and its subtypes (including sudden death) by multiple risk factor (95% confidence intervals in parentheses) status during age 30-50</td>
</tr>
<tr>
<td>8.4</td>
<td>Relative risk of deaths (total death and post-disease death) by single risk factor (95% confidence intervals in parentheses) status during age 30-50</td>
</tr>
<tr>
<td>8.5</td>
<td>Relative risk of deaths (total death and post-disease death) by multiple risk factor (95% confidence intervals in parentheses) status during age 30-50</td>
</tr>
<tr>
<td>9.1</td>
<td>Regression coefficients and standard error of Gompertz regression models for the transition NO-CVD to CVD for males</td>
</tr>
<tr>
<td>9.2</td>
<td>Total life expectancy (LE) and residual life expectancy free from cardiovascular disease based on a population free of cardiovascular disease at age 50, by single risk factor status</td>
</tr>
<tr>
<td>9.3</td>
<td>Lifetime probability (%) of subjects without cardiovascular disease developing cardiovascular disease at age 50, by single risk factor</td>
</tr>
<tr>
<td>9.4</td>
<td>Total life expectancy and life expectancy free from diseases based on a population free of cardiovascular disease at age 50 by risk profiles at middle age</td>
</tr>
<tr>
<td>9.5</td>
<td>Lifetime probability (%) of developing cardiovascular disease, for the cardiovascular disease free person of age 50 by optimal and high risk profiles</td>
</tr>
</tbody>
</table>
Preface

Life is full of attachments and detachments through matter of choices and opportunities! After graduation in statistics from the University of Dhaka, I got the opportunity to join as a researcher at International Center for Diarrhoeal Disease Research of Bangladesh (ICDDR,B)- a center for health and population research. During my research at ICDDR,B, I looked for higher studies, got the opportunity to join several places but I made choice to join the Population Research Centre in Groningen. I received my ‘doctoraal’ (Master’s) in Population Studies in 2001. The Master’s thesis was the commencement of this dissertation. The subject of this dissertation is fascinating because of its broad research domain, at least from public health perspective. I explore the utility of multistate life table technique in public health research and investigate the cardiovascular disease life history, its risk factors and the compression of morbidity. The general approach is to link demographic and epidemiological methods to develop new public health and life course indicators, which are transparent and policy relevant. A new public health approach to the compression of morbidity debate is initiated that relates risk factor status and disease incidence.

This research- like a journey- has been exciting, dynamic and full of discoveries. There is a great feeling of excitement, as you never know what is coming. This study is a part of the ‘Compression of morbidity’ project financed by the Netherlands Organization for Scientific Research (NWO). It is a collaborative project between Department of Public Health, Erasmus MC, Rotterdam and the Population Research Centre, University of Groningen. When I started work on this research project, the glossary of compression of morbidity, cardiovascular disease and epidemiology was completely new for me. I was not sure if I had chosen the right track but after two and a half years I have no regrets at all. This is something where I could put my heart into it.

This book would not have been possible without the unflinching support and guidance from Prof. Frans Willekens and Prof. Johan Mackenbach. My earnest thanks and admiration go to both of them. Prof. Frans Willekens taught me multistate method and life history analysis during his course in the Masters program. His enthusiasm, devotion and willingness to share his knowledge have been of essential importance in motivating me to complete this study. During the preparation of this manuscript, I had several tough and stressful times with “deadline, red ink and ‘zin’ ”, however, I fortunately managed with the support and motivation from Prof. Willekens. His use of the Dutch word ‘zin’ (meaning or sentence) in my drafts has indeed encouraged me to formulate my arguments clearly and coherently to improve my writing skills. Prof. Mackenbach showed me the way to interpret the statistical or analytical results in epidemiologic or public health perspectives. His comments during the NWO meeting and the comments on
the draft version of this thesis made it readable and interpretable for the public health researchers. His sharp interpretation of analytical results has always been awe inspiring and finally made the birth of this book possible.

Prof. Willekens and Mrs. Maria Willekens have supported me in many areas including academic, family, and personal matters. Whatever and whenever I have problems, their door is open to me, their hands are receptive and their heart filled with happiness. My sincere gratitude goes to Mrs. Maria for her care, support and hospitalities at all stages of my stay here.

I would like to extend a special thanks and admiration to Anna Peeters and Luc Bonneux, from Erasmus University Rotterdam. From the beginning of this research we have been working together. Some of those hasty days of defining cardiovascular disease states, creating input data for multistate life tables using the Framingham Heart Study, data analysis … ended by a nice dinner offered by Prof. Frans. Both of you inspired me and helped me unrestrictedly to make me understand Framingham Heart Study and the cardiovascular disease process. Thanks Anna for all your comments, e-mails and editing. Luc your sharp comments and encouragement motivated me lots to write this book.

Without pensive and encouraging support at each NEDCOM (Netherlands Epidemiology and Demography Compression of Morbidity Research Group) group meeting and further e-mail communications, this study would not have been completed. I would like to express my sincere gratitude to all of my colleagues in the NEDCOM group. I am proud to be a part of this dynamic group in public health research. Thanks Wilma, Jan, Chris, Fanny and Anton for your valuable comments.

I am extremely grateful to the Framingham Heart Study coordinators for access to the Original Data-set, and in particular to dr. Paul Sorlie. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University. It should be noted however that the manuscript of this book was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI.

I express my appreciation to the members of my home organization ICDDR,B, from where I am on official leave now. Especially I would like to thank Professor Mahmud Khan, Dr. Abbas Uddin Bhuyia, Professor Lars Ake Persson and Professor David Sack.

Special mention and gratitude go to Jeroen van Ginneken, who showed me the way to reach PRC in Groningen University. We are always good friends from ICDDR,B. I am grateful to his inspiration and hospitality. My special thanks go to one of my best friends Rian Scheffer, who taught me how to behave and adapt
environmentally in the Dutch culture. She has always been a constant source of encouragement for me.

With my friends at PRC, Sabu Padmas (now at the University of Southampton) and Salut Muhidin (now at the University of Brown) to whom I am grateful for their very friendly reception on 24th of August 1999 that was my entry at PRC. From that time we are always best friends. Thanks go to Inge Hutter, Sergei Scherbov and Harrie van Vianen for their kind suggestions and exchanges of experiences. I had excellent time with my office mate Ganesh and Hideko, with whom I shared both professional and personal events. I also take this opportunity to thank Maaike, Tomás, Sarbani, Karen, Vladimir, Padma and Alinda who provide active company in PRC. I would also like to thank Jose Dias and Samir (now in Nepal) for the statistical discussions and humor.

There are many other people who have helped me during this study. The management support by Stiny Tiggelaar was indeed extraordinary. She helped me lots in many respects including reading Dutch letters, taking care of health insurance and many other things. Special thanks go to Annette who handled all the financial and logistic support. I am thankful to all my other friends and colleagues at the Faculty of Spatial Sciences.

I am grateful to Karen Laird Gribling for correcting the English, which made the book more readable and interesting. Finally, I extend my sincere thanks to the members of the reading committee, Prof. Dr. Doeke Post, Prof. Dr. ir. Daan Kromhout and Prof. dr. Leo J.G. van Wissen.

The whole credit goes to my parents! Nothing of my career or my character would have been possible without the great love and care of my parents, which they have always shown to me. Without their blessings, I am nothing. I am grateful to my parents who have taught me honesty, kindness, forgiveness and hard work. My elder brother, Ali Ahmed, has always been a source of strength in all situations and movements of need.

Still I can remember the day I came here, on a sunshine afternoon on 24th of August 1999. Just after 4 months my wife Mohsina Khatun (Moury) joined with me at PRC. From that time we have had a wonderful life (that was the first time we started living together). We passed breathtaking life and became parents to a lovely daughter on 26th of November 2000. The beginning of our baby was really threatening! She was in hospital for the first 38 days of her life. After a long fight she survived and now she walks, runs, talks …. keeps us busy always with lots of fun and joy. I received much supports from all of my friends, colleagues, relatives and others. All the support from all of you and God’s blessings made it possible to overcome all the hard times and begin another joyful life.

My dear Moury, your constant encouragement, support and sacrifice that not only helped me to accomplish this thesis, but also did so with love, humor, and tenacity, was really essential. My love and thanks to my wife, my lovely daughter
and my mother in law. The inspiration and love that I received from my mother in law helped me immensely to concentrate on this study, in particular during her 10-month stay in Groningen. Without their unreserved support and encouragement, I would have never finished this study. You all always kept me warm in so many ways. Finally, without the blessings of the Almighty all endeavors would not have been possible.

Groningen
April 23, 2003
1

General introduction

1.1 Introduction

The world celebrated the new millennium with more than 6 billion people. Significant improvements in population health have occurred in the past half-century, markedly increasing the number of older people in industrialized countries. Together, the rapidly expanding older population in developing countries and the older population in the industrialized countries, form a global, aging force that is increasing the burden on the public health system and on medical and social services. More than 65 percent of Americans aged 65 years or older, for example, currently have some form of cardiovascular disease (Kane, 2002) and 30 percent of deaths are due to cardiovascular disease. The longer people survive and the older they become, the more health becomes a dominant issue. Along with an aging population come alarming health problems such as chronic disease (i.e. cardiovascular disease) and disability. The focus of this study is on chronic disease, specifically cardiovascular diseases, as they cause much human suffering, create substantial threats to economies of individual countries, and contribute to health inequalities between countries and within populations worldwide (WHO, 2002).

Uncertainty about the future health of the elderly population is known in the international context as the compression versus expansion of morbidity. The term morbidity refers to a manifestation of ill health. The compression of morbidity hypothesis assumes that the length of life is fixed and chronic disease and related disability can be postponed to older ages (Fries, 1980). The opposite hypothesis, i.e. that of the expansion of morbidity, states that mortality reductions will produce more years with morbidity and related disability (Gruenberg, 1977; Kramer, 1980; Verbrugge, 1984; Olshansky et al., 1991). Both hypotheses, that of the compression and that of the expansion of morbidity have been debated in relation to ageing over the past decades. The debate has focused either on degenerative diseases (Barendregt and Bonneux, 1998) or disability (Crimmins et al., 1994; Nusselder, 1998) of the older population. Little attention has hitherto been devoted to cardiovascular disease and its risk factors. Cardiovascular disease is the major chronic disease. Cardiovascular disease is the number one killer in the world and it contributes to disability, diminished quality of life, and greatly increases health care
costs. Its burden on society is increasing (WHO, 2002). The subject of this book is the development of a model within the framework of the compression of morbidity theory for cardiovascular disease, its subtypes in the life history of cardiovascular disease and its risk factors. In the greater debate on compression or expansion hypotheses, the compression or expansion of specifically cardiovascular morbidity and its risk factors is a new concept in public health research.

The compression of morbidity hypotheses is investigated in relation to the life history of cardiovascular disease and its risk factors, by means of consistent application of the time-honored multistate life table (MSLT) technique to the longest follow-up study in history- the prestigious ‘Framingham Heart Study’. The methodology applied in this study can be described as a new public health approach to the compression of morbidity debate “coupling demographic techniques and the framework of epidemiology”.

Section 1.2 of this introductory chapter offers an overview of the compression versus expansion of morbidity hypotheses, with a description of the changes occurring over time in mortality and morbidity. A general framework for integrating morbidity and mortality is described in Section 1.3. In Section 1.4, we describe the cardiovascular disease process and its risk factors. The research objectives are formulated in Section 1.5. The organisation of this book is summarised in Section 1.6.

1.2 The changes in mortality and morbidity

Mortality is a fundamental factor in population dynamics (Omran, 1971). The overall mortality rate of human beings has declined radically throughout the world during the last half centuries. Life expectancy at birth has doubled, infant and child mortality has declined significantly and mortality has shifted to older ages. The survival of the old-age population has increased substantially since 1950 (Vaupel et al., 1998). At present, in the low mortality countries, a longer life is often taken for granted. The global life expectancy at birth has increased from a global average of 46 years in 1950 to 66 years in 1998 (Sen and Bonita, 2000). The twentieth century has seen a dramatic increase in the life expectancy of residents in the United States. The life expectancy at birth increased from 49 years in 1901 (U.S. Census Bureau, 1996) to 77 in 2000 (World factbook, 2001).

This remarkable increase in life expectancy has been guided by substantial changes in the age-at-death and cause-of-death patterns (Nusselder, 1998). These shifts in age and cause specific deaths are illustrated in the epidemiologic transition theory (Omran, 1971), which is an extension of the mortality component of the demographic transition (Mackenbach, 1994). The theory of epidemiologic transition has developed from the application of epidemiology, demography, and other social and health sciences to population dynamics. The focus of the theory is on the complex
changes in patterns of health and diseases; their demographic, socioeconomic, and biologic determinants; and consequences for population groups (Omran, 1971).

There are two established groups of researchers investigating mortality and life expectancy- one group, known as the promoters of the limited-life-span paradigm and another group known as the promoters of the mortality-reduction paradigm. The first group concludes that average life expectancy will not increase beyond 85 years of age (Keyfitz, 1978; Fries, 1980; Olshansky, et al., 1990; Olshansky and Carnes, 1994; Wilmoth, 1998). The arguments for this are provided by the evolutionary theory of senescence (e.g. substantial reductions in mortality rates at advanced ages are constrained by biological barriers), well known to researchers in the field of biodemography (Fries, 1983; Carnes and Olshansky, 1993; Olshansky and Carnes, 1994). The second group argues that the decline in mortality rates will persist and may even speed up, even at the most advanced ages (Vaupel and Gowan, 1986; Vaupel and Lundstrom, 1994; Manton, et al., 1991; Guralink et al., 1988; Schneider and Brody, 1983). Based on the enormous increase in life expectancy in the past, this group predicts further increases up to as high as 100 years or even more in the near future (Manton et al., 1991; Vaupel and Lundstrom, 1994). However, viewpoints differ as to whether there is a biological limit to life expectancy. Some researchers believe that biological limits exist, but that future advances in technology (e.g. biomedical research) may nevertheless boost life expectancy up to 100-125 years or even 150-200 years (Manton et al., 1991).

What was ignored in the debate was morbidity. Traditionally, improvement in population health status has been measured on the basis of the increase in life expectancy of that population. The reason is simple- data availability and simple estimation procedure. Nowadays, in the industrialized countries where improvements in life expectancy are mainly caused by mortality reductions from chronic disease in older ages, thought-provoking doubt exist as to whether longer life expectancy means a reduction in morbidity (Gruenberg, 1977; Kramer, 1980; Verbrugge, 1984; Olshansky et al., 1991). This doubt about the development of morbidity, in particular in relation to the elderly population, has led to the formulation of three hypotheses: the expansion of morbidity, compression of morbidity and dynamic equilibrium.

The expansion versus compression of morbidity debate originated with papers by Gruenberg (1977) and Kramer (1980). They pointed out that the present advances of medical technology allow us to save the frail and disabled from dying from complications, and therefore that mortality reduction will produce more years with morbidity and related disability. Their hypothesis is also called ‘failure of success’, and would further lead to a ‘pandemic of mental disorders and associated chronic disease and disabilities’. They based their argument on the fact that if the incidence of chronic disease (e.g. cardiovascular disease) and disability remains constant, but survival improves, the stocks of frail patients will increase.
The antithesis of expansion is compression, as is espoused by Fries (Fries, 1980; Fries, 1983; Fries, 1989). The main proposition of Fries is that the length of life is fixed. He argues that there is a natural limit to the life span—the so-called 'natural death' at around age 85. The process of rectangularization is viewed as a signal that life expectancy is reaching the maximum life span, preventing this from any further significant increase. Since the genetic potential of the human species is limited, mortality will cease to decline and the maximum limit be reached. Fries' second proposition is that chronic disease can be postponed by adopting a healthy life style. It is assumed that if morbidity can in reality be compressed into fewer years toward the end of a hypothetical 'full life span,' the quantity of disability over the whole life span will be abridged (Vita et al., 1998). Slowing down the onset of chronic disease and disability, while assuming a fixed length of life, produces a decline in the number of years with morbidity. This is called as 'compression of morbidity'. The notion of the compression of morbidity hypothesis has implications for the health care policies and costs and for the quality of lives of a population.

While the compression of morbidity hypothesis relates longer life (though limited) with an improvement in the healthfulness of life, and the expansion of morbidity hypothesis associates a longer life span with a fixed incidence, a third, intermediate view, known as the dynamic equilibrium hypothesis has been formulated by Manton (Manton, 1982). Manton argues that an equilibrium exists between life expectancy and the health and functioning of the elderly population. This hypothesis states that for many chronic diseases improved survival will come about by slowing the rate of progression of the primary disease process, i.e. the number of years lived with morbidity will increase, but the years lived with severe morbidity and disability will be relatively constant.

Numerous studies have tried to find empirical evidence for the expansion or compression of morbidity (Mathers et al., 1994; Perenboom et al., 1993; Robine, 1994; Wilkins, et al., 1994; Nusselder, 1998; Barendregt and Bonneux, 1998). Usually they operationalised expansion or compression in the generic term 'disability'. Some findings point towards expansion, some towards compression, and some towards neither. Several methodological papers have pointed out that the most commonly employed method (e.g. Sullivan method), that is based on prevalence-based information (Colvez and Blanchet, 1983; Wilkins and Adams, 1983; Bebbington, 1988; Crimmins et al., 1989; Rogers et al., 1989; Mathers, 1991; Robine et al., 1995; 1998; Crimmins et al., 1997; Hayward and Heron, 1999) is extremely complex. Compression and expansion depend on the kind of disease and kind of the disease epidemiology. At the very least, a distinction should be made between fatal and non-fatal disease (Manton, 1982; Olshansky et al., 1990). However, to our knowledge there is no empirical evidence as to whether cardiovascular morbidity and its risk factors compress or expand cardiovascular morbidity. In this study, we propose to operationalise the concept of compression...
in terms of the number of years lived with cardiovascular disease and cumulative incidence (absolute compression) in the presence or absence of risk factors using the multistate life table approach. Methodologically speaking, multistate life tables were felt to be a better choice for estimating compression or expansion than the prevalence-based Sullivan (Sullivan, 1971) type method (Rogers et al., 1989; Barendregt and Bonneux, 1998; WHO, 2000).

1.3 Integration of morbidity and mortality: a general framework

The main determinants of morbidity are the prevalence of a disease and the severity of that disease (Barendregt and Bonneux, 1998). The distinction between diseases or disorders, impairments, disabilities and the handicaps is the succession of events that may occur following disease. To describe the consequences of diseases, WHO (1980) introduced the concept of IDH: impairment, disability and handicap. Disease is defined as the abnormality that appears within individual (intrinsic situation). Impairments are “concerned with abnormalities of body structure and appearance and with organ or system function, resulting from any cause; in principle, impairments represent disturbances at the organ level”. Disabilities reflect “the consequences of impairments in terms of functional performance and activity by the individual; disabilities thus represent disturbances at the level of the person”. Handicaps are “concerned with the disadvantages experienced by the individual as a result of impairments and disabilities; handicaps thus reflect interaction with and adaptation to the individual’s surroundings”. However, in public health research, the health outcome of interest is the disability which a disease causes, or in other words disability is considered to be a dimension of disease. Thus chronic diseases such as cardiovascular disease bring disability into the life of the individual.

The general model of health transitions that allows a direct assessment of the health consequences of increasing survival was introduced by the World Health Organization (WHO) in 1984. This survival curve model provides a comprehensive framework and an analytical tool to integrate changes in mortality, morbidity and disability over the life course in a standard manner. It distinguishes between total survival, disability-free survival and survival without (chronic) morbidity (Figure 1.1). The survival curves used in the WHO models are determined by age-specific mortality, morbidity and disability and are calculated from the life table. In Figure 1.1, the area below the mortality curve represents total life expectancy (say, LE=80 years), the area under the disability curve represents disability-free life expectancy (say, DFLE=75 years) and the area under the morbidity curve is the life expectancy without morbidity (say, LEWOM=71 years). The difference between LE and DFLE measures life expectancy with a disability (LEWD: area between ‘mortality’
and ‘disability’ curves). The area between the mortality and morbidity curve mirrors the expected duration of life with morbidity (LEWM). The sum of complementary life expectancies is equal to total life expectancy, for example, LEWOM plus LEWM is equal to total LE.

The general framework of the WHO health transition models enables us to assess whether or not changes in mortality and morbidity result in compression or expansion of morbidity. The survival and health status depends on the risk factor status, which is ignored in this model. For instance, people with an optimal risk in adulthood survive longer and remain in good health compared to the high-risk group. Therefore, the effect of risk factors on disease incidence and mortality and the balance in compression or expansion of morbidity can be analyzed using the same framework. For those with optimal risk the survival curves (Figure 1.1) will shift upwards and for those with high risk it will shift downwards. In this study, we have operationalised this framework, thereby incorporating single as well as multiple risk factors.

Figure 1.1  Mortality (hypothetical), morbidity (hypothetical) and disability (hypothetical) survival curves

The health status of a population is measured by combining morbidity and mortality. An indicator that summarizes the population health status over the life course and addresses the discussion of compression and expansion of morbidity is health expectancy. Health expectancy is defined as the average number of years an individual is expected to be healthy if current mortality and health status trends continue to apply (Mathers and Robine, 1994). Since health expectancy is in fact the combination of a life expectancy with a health concept, there are as many possible
health expectancies as health concepts (Robine et al., 2000). We can thus compute indicators such as life expectancy “in good perceived health” according to WHO definition of health, or “disease-free” life expectancy, “impairment-free” life expectancy, “disability-free” life expectancy and “handicap-free” life expectancy. The “dementia-free” life expectancy, proposed by Ritchie, is a good example of disease-free life expectancy (Ritchie, 1991). The “illness-free” life expectancy proposed by Newman (1988) is another example of combining morbidity and mortality data. In the field of mental health, general indicators of mental health expectancy have been produced as well (Gispert et al., 1998). Through the use of longitudinal data sets and new methodological techniques, researchers have started to compute active life expectancies – the expected duration in years of functional well-being according to the performance of activities of daily living (ADL) (Katz et al., 1983). This measure not only measures how long a sub-population can expect to live, but also what fractions of the expected remaining lifetimes will be spent in independent or dependent statuses (Rogers et al., 1989). We can compute life expectancy “without significant risk factor damage” (Manton, 1989) or more simply without “risk” (Rowe, 1990) to assess what we call “successful aging” (Rowe and Kahn, 1997). However, in this study the focus is mainly on life expectancy without cardiovascular disease, including its different manifestations, in presence or absence of potential risk factors.

1.4 Setting the scene

1.4.1 Cardiovascular disease

The subject of this study is to model the cardiovascular disease history and its risk factors. Cardiovascular disease is an established chronic disease for the population of developed and developing countries. Chronic diseases are illnesses that are prolonged, do not resolve spontaneously, and are rarely cured completely. Cardiovascular disease refers to variety of diseases and conditions affecting the heart and blood vessels. The major cardiovascular diseases are coronary heart disease (CHD), myocardial infarction (MI), stroke and congestive heart failure (CHF). Cardiovascular diseases are preventable. This disease causes a significant burden in mortality, morbidity, and health care cost.

Despite the gradual decline in cardiovascular death rates over the last few decades, these diseases are and will remain the number one cause of death in industrialized countries. In this region, 48.6 percent of deaths were caused by cardiovascular disease in the year 2000 (WHO, 2002). By 2020, 46.4 percent deaths are expected to be attributable to CVD in this region. Moreover, CVD will soon become the main cause of death and disability in the developing world as well; by 2020, a third (33.8 percent) of all deaths are expected to be due to CVD (WHO,
2002). In the year 2000, 16.7 million people died from CVD, accounting for 30.3 percent of all deaths worldwide; more than half of these deaths were in developing countries (WHO, 2001). Not only is cardiovascular disease a considerable health burden (e.g. high morbidity and high mortality), it causes also a significant health care cost (i.e. economic burden), which will continue to grow as the population ages.

The ongoing economic and technological developments taking place in the developing countries will in all likelihood cause the pattern seen in developed world to be repeated. The epidemic of cardiovascular disease is expected initially to emerge in those who are wealthy and subsequently to spread to those who are less wealthy. Likewise, when the epidemic starts to slow, this will first become apparent among the affluent, with the disease continuing to have a high prevalence in those who live in poverty. The burden of CVD, although already high in developed world, is therefore expected to increase on a global scale as the developing countries start to contribute significantly to this (WHO, 2002). The presence of rising CVD underscores the imperative need to develop effective and appropriate prevention policies.

### 1.4.2 Cardiovascular disease risk factors

A risk factor is a condition that elevates or raises our chances of getting a disease. Two types of risk factors of cardiovascular disease might be distinguished - adult risk factors and risk factors in early life. These early and later risk factors may well interact with each other. The adult risk factors and their impact on cardiovascular morbidity and mortality are the focus of this study.

**Figure 1.2 Established adult risk factors**

<table>
<thead>
<tr>
<th>Unmodifiable traits that predict risk</th>
<th>Age</th>
<th>Sex</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behaviors that affect risk</td>
<td>Smoking</td>
<td>Alcohol consumption</td>
<td>Diet</td>
</tr>
<tr>
<td>Physiological conditions increasing risk</td>
<td>Hypertension</td>
<td>Hypercholesterolemia</td>
<td>Obesity</td>
</tr>
<tr>
<td>Cardiovascular endpoints</td>
<td>CHD</td>
<td>MI</td>
<td>Stroke</td>
</tr>
</tbody>
</table>

*Source: Stolley and Lasky, (1995)*
Most of the established adult risk factors for cardiovascular disease are shown in Figure 1.2. Some risk factors are modifiable and some are not. Age, sex and race are unmodifiable risk factors. The major behavioral risk factors are smoking, alcohol consumption, diet, psychosocial factors and physical exercise. Some of the physiological conditions such as hypertension, high cholesterol, obesity and diabetes are most important risk factors for cardiovascular disease.

The risk of cardiovascular diseases accumulates with age and is influenced by factors acting at all stages of the life span (WHO, 2002; Ben-Shlomo and Kuh, 2002). This life course epidemiology is new but growing. From this perspective, for instance, the development of cardiovascular disease starts during fetal development and is influenced by factors acting at different stages- infancy and childhood, adolescence and adult life (Figure 1.3). To explain the possible ways in which factors over the life course may act to cause chronic disease, so far only a few theoretical models have been advanced (Ben-Shlomo and Kuh, 1999; 2002). However, the most firmly established associations between cardiovascular disease and factors in life span are those between disease and the major known adult risk factors: tobacco use, obesity, physical inactivity, cholesterol, high blood pressure, and alcohol (Godlee, 1999; Elisaf, 2001). This study focuses solely on these adult risk factors.

Figure 1.3 Development of cardiovascular disease

Source: WHO (2002)

1.4.3 Basic model of cardiovascular disease

To estimate cardiovascular compression of morbidity and its risk factors we use multistate life table techniques. Multistate life table techniques have been used before to obtain consistent sets of data for chronic disease modeling (Hoogvenveen et al., 2000). In public health, MSLT methodology has also been used to estimate
the population health status and to address the question of compression or expansion of morbidity (Crimmins et al., 1994; Nusselder, 1998; Nusselder et al., 2000; Barendregt and Bonneux, 1998). The main feature of this model is that it considers both the increment (inflow) and decrement (outflow) according to various diseases states of the population.

The basic model structure of cardiovascular disease is presented in Figure 1.4. There are three states in this model: NO-CVD (i.e. free of CVD), CVD and dead. At age \( x \) an individual could be in NO-CVD state or in CVD state. At age \( x+1 \) that individual can make transition to another state (either incidence or remission) or remain in that state. Life tables are usually presented in one-year age intervals, beginning with an arbitrary value at age \( x \).

Figure 1.4    Basic (multistate life table) model structure of cardiovascular disease

We use the first 50 years of follow-up of the Framingham Heart Study (FHS) to construct multistate life tables. The *transition probability*, which is the basic parameter of the MSLT is estimated based on the *occurrence-exposure rate*, which represents the standard method. The occurrence-exposure rates are estimated directly from the data set. Estimation of the MSLT using occurrence-exposure rates has several advantages. *First*, the MSLT can be viewed from a life history perspective, allowing the reconstruction of the individual biography. *Second*, it provides us with a tool to analyze the changes in morbidity, disability and mortality rates within the framework of a single integrated model of population health and to summarize the information on changes in morbidity, disability and mortality into an integrated indicator of population health (Nusselder, 1998). Health expectancy is one of the best-known examples. *Third*, the observed occurrence-exposure rates enable the use of data in the presence of any type of censoring- left, right, double or
interval. Fourth, an MSLT can work as a tool to operationalize the conceptual model (Figure 1.1) to describe and visualize the association between changes in morbidity, disability and mortality and overall population health. Fifth, multiple covariates can be included in the construction of the MSLT. This means that we can combine MSLT with multivariate regression models. Sixth, this model can be used to clarify the debate of compression or expansion of morbidity hypotheses more accurately compared to the prevalence-based method.

We chose to consider the original cohort of the Framingham Heart Study to perform this study, because of the long-term follow up and accurate record keeping of the cardiovascular disease occurrences and deaths. The more impressive information is the 48 years consistent follow-up of the cardiovascular disease risk factors. This study offers a unique opportunity to assess the life history of cardiovascular disease and its risk factors. The FHS, combined with the robust multistate life table method offers an excellent means to pursue our object.

1.5 Research objectives

The objectives of this study are both methodological and substantive. We have two aims, namely to:

- Explore the utility of the multistate life table for public health, specifically, to operationalise the debate on compression or expansion of morbidity.

- Model the cardiovascular disease life history and its risk factors and to investigate whether the risk factors of cardiovascular disease and its sub-types lead to expansion or compression of cardiovascular morbidity.

To date, the question whether or not changes in population risk factors such as high cholesterol and smoking would lead to increased or decreased cardiovascular morbidity within the population has received little to no attention in the literature. The question of whether changes in population risk factor status leads to compression or expansion of cardiovascular morbidity is addressed using the first 48-years of follow-up of the original Framingham Heart Study cohort. We compare the cardiovascular life course between the optimal risk (e.g. normal body mass index) and high-risk category (e.g. obesity). We investigate whether cardiovascular risk factors (both single and multiple risk factors) at middle age compresses cardiovascular morbidity.

The compression versus expansion debate has grown more complex over the past decades. For instance, two recent studies reported that a population of non-smokers would experience more years lived with disability (Ferrucci et al., 1999; Martel et al., 2000), and one study reported that non-smoking compresses
morbidity both in absolute and relative terms (Bronnum-Hansen and Juel, 2001). In addition to furthering this complex debate, the present study, using the original Framingham Heart Study cohort allows us to measure the compression or expansion of disease-specific life expectancy, such as the life years lived with cardiovascular disease and life years lived free of cardiovascular disease in the presence or absence of risk factors.

We make an explicit distinction between prevalence-based and incidence-based measures. We explore the theoretical and practical aspects of the multistate life table approach and indicate the methodological innovation and utility of this method in public health research. In this study, the basic parameter for constructing the multistate life table is the occurrence-exposure rate. A method for calculating the occurrence-exposure rates using micro data is illustrated. We generalize the multistate life table, describing this in multiple covariate contexts i.e. bridging the multivariate regression models and multistate life table techniques. We illustrate the bootstrapping technique used to estimate the confidence intervals of the multistate life table outcomes.

Together with disease history, risk factor status in an important aspect in better understanding the complexity of the compression or expansion of morbidity issue. To better address the issue of compression of morbidity hypotheses, longitudinal data are needed. A recurrent problem with longitudinal data is that of missing values. We propose potential methods to impute missing values of the risk factors in a repeated measurement study. We selected smoking as an example of a risk factor career. The life history of smokers and nonsmokers is demonstrated by constructing a multistate smoking status life table.

To construct a life table (with or without covariates), we need age-specific transition rates. Transition rates can be derived directly from the data set (e.g. occurrence-exposure rates) or can be estimated by fitting regression models to the data. Two well-known models developed by Cox and Gompertz are considered and assessed with the aid of empirical rates and life table outcomes.

This study compares the burden of cardiovascular disease in terms of lifetime risk and life years lived with disease between long-time smokers and non-smokers. We address the question of whether non-smoking shortens the number of years lived with cardiovascular disease, i.e. leads to compression, given the competing forces of an increased risk of cardiovascular disease and increased mortality in smokers.

We focus on the risk factor status at middle age (age 30 to 50) and its impact in the older ages of life. One of our objectives is to investigate the association of the risk factor status at middle age and the incidence of cardiovascular disease and its subtypes, and post-disease mortality over a long-follow-up. Whether changes in population risk factors such as high cholesterol and smoking lead to increased or decreased cardiovascular morbidity within the population has previously not been
examined in the literature. This question is addressed in the present study by integrating multistate life table and the multivariate regression models. Performing both univariate (i.e. single risk factor) and multivariate (multiple risk factors) analysis allows us to investigate whether risk factor status at middle age compresses cardiovascular morbidity.

1.6 Outline of this book

This book is structured into ten chapters, of which Chapter 1 is the general introduction. In Chapter 2, we review the multistate life table technique and emphasize the potential utility of this method in public health research. The multistate life table is the basic tool that used throughout this study to measure the cardiovascular disease history and address the question of compression or expansion of morbidity. We explicitly differentiate between the prevalence and incidence measure of population health status. A concise overview of the mathematical construction of a multistate life table is given in this part of the study. We describe the equations, thereby focusing on micro data instead of macro data. We generalize the multistate life table by describing it in multiple covariate contexts. The input data and the possible output from the multistate life table are also illustrated. We demonstrate the possible methods to calculate the confidence intervals around the multistate life table measures. The fundamental difference between the multistate life table approach and the Sullivan method is described as well. Overall, this provides an overview of the theoretical and practical aspects of the multistate life table method and indicates the methodological innovation and utility of this method for public health research.

Chapter 3 demonstrates a method to produce occurrence-exposure rates using micro data. The construction of the multistate life table is illustrated step-by-step. Here, we transfer the well-described epidemiological measures into time based public health policy measures, such as the life time probability of cardiovascular disease, life years lived with cardiovascular disease and life expectancy free of cardiovascular disease. The indicators derived in this chapter of the study are integral to appropriate health planning and assessment of the potential population health value of various treatment and prevention strategies.

In Chapter 4, we propose potential methods for imputing missing values of risk factors in repeated measurement studies. In a longitudinal or panel study, missing data is very common. To construct the risk career (e.g. smoking career), the imputation of such missing values is essential. Without imputation, we may not be able to capture the event that we would like to relate with the risk factor status at a nearby point in time. We consider two risk factors- smoking status and systolic blood pressure. Smoking status is selected to illustrate the imputation method for
categorical risk factors. The systolic blood pressure is selected to exhibit the
imputation method for continuous risk factors.

Chapter 5 illustrates how a risk career can be reconstructed. We have taken a
novel approach to analyze the “career or pattern over time” of risk factors, starting
with a very common risk factor smoking. Multistate life tables are developed to
capture different episodes of smoking. The described approach is a novel one, in
that it demonstrates the importance of explicitly taking into account the changes in
risk factors throughout life.

In Chapter 6, we compare the burden of cardiovascular disease in terms of
lifetime risk and life years lived with disease between smokers and non-smokers.
The basic question answered in this chapter is whether smoking compresses
cardiovascular disease morbidity. We constructed multi-state life tables describing
transitions through various cardiovascular diseases for smokers and non-smokers
observed during 20 biannual observations in the Framingham Heart Study. To
estimate the confidence interval of the multistate life table outcomes we use
bootstrapping technique.

Chapter 7 provides as assessment of the well-known and widely used transition
rate models of Cox and Gompertz. We compare the age-specific transition rates
and the life table outcomes to know how the Gompertz and Cox models fit the
empirical results, in presence or absence of covariates.

In Chapter 8, we perform both the univariate and multivariate regression
analysis of the cardiovascular risk factors to investigate the association of the risk
factor status at middle age (during age 30 to 50), the incidence of cardiovascular
disease and its subtypes, and post-disease mortality over a long follow-up. We
estimate, in long follow-up, how the effect of an adult risk factor is changed if one
adjust this effect for the confounding influence of other factors.

In Chapter 9, we examine whether cardiovascular risk factor status compresses
cardiovascular disease morbidity. We focus on the important CVD risk factors at
middle age and their impact on CVD and mortality at later ages of life. We compare
mainly the optimal and high risk. Initially, we explore the compression of morbidity
hypotheses for each single risk factor status separately. Finally, the combination of
several risk factors (i.e. multifactorial effect) and the compression of cardiovascular
morbidity are investigated. In the multivariate case, the optimal risk profile is
defined as a non-smoking individual, on average with optimal blood pressure
(BP<120), optimal cholesterol level (SCL<200) and optimal BMI (BMI<25)
between the ages of 30 and 50. If an individual is smoker, has high blood pressure
(BP>140), high cholesterol (SCL>240) and obese (BMI>30) is considered to
belong to the high-risk group. Empirical evidence is presented on the basis of
which the impact of risk factors on cardiovascular life course can be assessed.
Chapter 10, the final chapter, summarizes the major findings and presents the implications of the findings. We synthesize the utility of this research in chronic diseases modeling. The chapter concludes with suggestions for future research on the life history of cardiovascular disease, its risk factors and compression of morbidity.

References


Abstract

The main objective of this chapter is to review the multistate life table (MSLT) method and to emphasize the potential utility of this method in public health research. During the past decade, various public health researchers have applied this method as a means to measure health expectancy, which is the best-known index for summarizing the health status of the population and addressing the issue of 'compression' and ‘expansion’ of morbidity. In this chapter, we describe the theoretical aspects of the multistate life table approach that relate to public health status measurement. We distinguish between prevalence- and incidence-based measurements, and conceptualize these from the public health point of view. A concise overview of the mathematical construction of multistate life tables is given in this chapter. We describe the mathematical equations, focusing on micro data instead of macro data. We discuss population-based and status-based measures of health expectancy. We generalize the multistate table, describing it in multiple covariate contexts. The input data construction and the possible output from multistate life table are also illustrated. We demonstrate possible methods to calculate the confidence intervals around multistate life table statistics. The fundamental difference between the multistate life table method and the Sullivan method is described as well. We found that the MSLT method provides a more accurate assessment of health expectancy in a population. Overall, this chapter gives an overview of theoretical and practical aspects of the multistate life table and indicates the methodological innovation and utility of this method in public health research.
2.1 Introduction

One of the most important and lively current debates in the study of public health and mortality revolves around the idea that as improvements in survivorship and life expectancy continue, the health of those individuals benefited by these improvements may deteriorate (Fries, 1980; Singer and Manton, 1994). This improvement of survivorship raises the question of whether or not an increase in the proportion of older people will result in an increase in the prevalence of chronic disease and disability i.e. expansion of morbidity. On the other hand, public policy debates in the areas of health and medical care have also emerged in recent years around the question of whether improvements in life style and medical technology will delay the onset of chronic illness and disability and result in a compression of morbidity at older ages (National Research Council, 1988). For modeling the life history of cardiovascular disease and addressing the issue of compression or expansion of cardiovascular morbidity, we chose to make use of the time-honored ‘multistate life table’ method. The objective of this chapter is to review the multistate life table, describe it from public health perspectives and investigate its utility and applicability to measure population health status.

Over the past three decades, public health researchers have focused on the development and application of health measures that combine mortality and morbidity data. Efforts to collapse mortality and morbidity into a single measure to provide summary measures of population health stretch back far into the past (Sander, 1964; Chiang, 1965; Moriyama, 1968; Sullivan, 1971a; Katz et al., 1983; Rogers et al., 1990; Preston, 1993; Crimmins et al., 1994; WHO, 2000). The volume of work from members of the Réseau de Espérance de Vie en Santé (REVES) offers an indication of the activity in this field. The Sullivan method (see Hauet, 1997; Mamun, 2001) is one of the methods that is widely used in the public health domain to measure health status. The index is attractive for public health researchers, because this method is very simple and the data access is easy. There are, however, serious limitations to this method, such as the fact that it is unable to take into account any sudden change in health status (incidence) and re-entry into the life table population. A method that does not have these limitations is the multistate life table. The main advantage of this method is that it can accommodate both the increment and decrement according to various (health) statuses of the population. In this chapter, we explicitly distinguish between the concept of prevalence and incidence.

---

1 REVES: International Network of Healthy Life Expectancy. REVES is an international organization of researchers and health planners from universities, governments, and international agencies dedicated to understanding the use of health expectancy as an indicator of population health and as a tool for health planning.
In public health, Katz and co-authors (1983) introduced multiple decrement life tables. They found that this method gives better estimates than the Sullivan method. This method allows several exits from the healthy population: through death, dependency in carrying out activities of daily living, or institutionalization. But this method does not include differential mortality by health status and transition from disability states to healthy states (Saito et al., 1999). That is, this method considers change in health states from healthy to disability and death, from which no return is allowed. In practice, however, people can recover from disabilities or from diseases (acute or chronic) that affect their activities of daily living, and can be released from institutions. “This clearly produces an estimate of active life expectancy that is biased downward” (Saito et al, 1999). The multistate life table, an extension of basic life tables, allows more complexity to enter the analysis: people can enter, as well as exit a population and can move back and forth across a variety of states within a population. In 1989, Rogers and colleagues constructed the multistate life table to calculate the active life expectancy. They adopted the multistate life table technique to incorporate both the decrements and increments of several interacting sub-populations (based on the activities of daily life) to measure the active life expectancy. This was the first time the MSLT method was applied to the field of public health, although many researchers (Rogers et al., 1990, Crimmins et al., 1994; Saito et al., 1999; Laditka and Hayward, 1999; Mathers and Robine, 1996; Barendregt et al., 1997; Nusselder, 1998; Peeters et al., 2002; Mamun, 2001) subsequently went on to use this method.

Most of the researchers compared the multistate method with the Sullivan method using the same data set. There is no conclusion as to which method is the best one. However, methodologically and theoretically, the multistate method is preferable (Mathers and Robine, 1996; WHO, 2000). The key issue of the debate centers on the use of prevalence proportions versus incidence rates in the estimation of health expectancies. The MSLT method is data-demanding in nature. The Sullivan method is widely used, as it is very easy to calculate and solely takes prevalence information into account. The Sullivan method is based on the status data, while the multistate method is based on flow data. For the theory of compression or expansion of morbidity and in the health dynamics of an individual life cycle, the emphasis should be on the multistate method (Barendregt et al., 1997).

A key component of the multistate method is its relaxation of the assumption of unidirectional health changes over the life course. The multistate method allows individuals, as they age, to experience the onset of health problems and the recovery from those problems. No assumptions are made about the hierarchy of health status changes. Thus, persons can experience multiple and recurrent health problems over their lives. This method deals with these transitions from one state to another: from alive to dead, healthy to unhealthy, unhealthy to recovery,
employed to unemployed, and so on. Some of the transitions may occur only once, whereas other transitions can be repeated. Some transitions may preclude other types of transitions or may make further transitions possible (Willekens, 1987). This method often offers answers to questions such as: what is the probability/rate that healthy men of 65 become unhealthy by age 75? Or what is the probability that healthy women of 50 suffer from cardiovascular disease by age 65? How does the probability of being in a health state differ in terms of duration? What is the average amount of time spent by an individual in good health or with disability in his or her life? What is the lifetime probability of developing cardiovascular disease? What are the differences between smokers and non-smokers in the number of years spent with disease? At the generic level, what is the impact of cardiovascular disease risk factors on life expectancy? A simple way to understand the events and relations involved and indeed to begin to answer these questions in public health research is the MSLT approach. This method is capable of providing estimates of the expected number of transitions over a person’s lifetime or within a specified period of time. Thus, researchers can gain some sense of the health trajectories experienced by persons in the population.

The multistate method is very attractive because it can capture the natural course of a disease, and can encompass patients who are cured or have intermittent disease-free periods (Barendregt, 1994). The multistate model is used to chart the disease process or disease history and to detect the pattern of similarities or differences by basic explanatory factors such as sex, smoking habit etc. Recently, the multistate method has been used to model the population health and resource use (Niessen, 2002). A multistate life table can describe the transition of life and sojourn time in various stages of life. Using this dynamic approach the reconstruction of an individual biography is possible, even based on the elementary event (two states and one event) process. This method affords a unique opportunity to reconstruct the life history and to construct a ‘synthetic biography’ (Willekens, 1987).

The multistate method originated in the early 1970s in multiregional demography (Rogers 1973; 1975). Schoen (1975) and Schoen and Land (1979) applied it to the analysis of marital status. Since that time, the field has expanded and developed, as multistate methods have been found to be both powerful and flexible enough to capture the movement of the complex behavior of a cohort (Mills, 2000). Willekens (1987), following the work of Rogers (1973, 1975), Schoen (1975, 1988a) and others, argues that “the life table has made the transition from a method for estimating the length of life to a technique for describing the structure of life”. The key to the development of this method was the use of the Rogers’ mathematical demography, especially the application of matrix algebra in the life table.
The application of the multistate method includes diverse subject matters related to the family (Bongaarts, 1987; Yi, 1991), labor force participation (Willekens, 1980; Lalu, 1992), migration (Rogers and Willekens, 1986), spatial population distribution (Willekens and Rogers, 1978), marital status life table (Willekens, 1987), union formation and dissolution (Mills, 2000), contraceptive use (Islam, 1994), voting status (Land et al., 1985), active life expectancy (Rogers et al., 1989, 1990; Branch et al., 1991; Land et al., 1994; Crimmins et al., 1994; Hayward et al., 1998; Liu et al., 1995; Brouard and Robine, 1992; Izmirlia et al., 1997), mortality and morbidity projections (Rusnak et al., 1992) and health status and life expectancy (Crimmins et al., 1994; Nusselder, 1998). The majority of these studies used vital statistics or census data, for estimating the parameters of multistate life tables. Recently, researchers from variously different disciplines adopted this technique to measure health status (especially for aged population) by calculating the health expectancy or active life expectancy. However, from a public health point of view the explicit explanation of this method is inadequate. Here, we describe and explain the multistate method from the public health point of view. We present the theoretical issues that are essential to construct an MSLT. A concise overview of the mathematical construction of an MSLT is given in this chapter. The MSLT in multiple covariate contexts is discussed. From the technical point of view, we also describe the input data construction and the representation of output from the MSLT. For the statistical competence, we have described the available methods to calculate the confidence intervals of the outcomes from MSLT. The MSLT described in this study is a type of descriptive transition rates model with and without covariates. The multistate model is used to chart the cardiovascular disease process or disease history.

This chapter consists of the following sections. In Section 2.2, the distinction and interrelationship between prevalence and incidence has been made explicit. The Sullivan method and its limitations are discussed in Section 2.3. The multistate life tables are illustrated from a public health point of view in Section 2.4. The conceptual issues of MSLT are discussed in Subsection 2.4.1. In Subsection 2.4.2, the theoretical basis of the multistate life table, i.e. the Markovian property, is talked about. The MSLT equations are derived and discussed in Subsection 2.4.3. It includes the derivation of transition probabilities and equations for life table statistics. MSLT in covariate contexts is discussed in Subsection 2.4.4. The estimation of confidence intervals is described in Section 2.4.5. The input data to construct MSLT and the output from MSLT are described in Subsections 2.4.6 and 2.4.7. We conclude this chapter with a critical assessment of MSLT in public health.
2.2 Prevalence and incidence measures

Epidemiologists and demographers describe the magnitude of health problems either in terms of prevalence or incidence, or a combination of both. Incidence and prevalence are two basic measures of disease occurrences in populations (Selvin, 1991; Young, 1998). Prevalence (or proportion) reveals how many cases exist in a population at a given time. Incidence is usually expressed as the number of new cases occurring within a population at risk (for example, of contracting a disease) over a period of time. For epidemiological purposes, the occurrence of cases of disease must be related to the “population at risk”. The concepts of incidence and prevalence are discussed here.

2.2.1 Prevalence

Prevalence refers to the proportion of a total population that has a defined health problem or disease. It is usually measured by surveying a particular population containing people with and without the condition of interest. The prevalence is usually reported as percentage. Usually, prevalence is of two types: point prevalence and period prevalence. We identify another type of prevalence, duration prevalence. Any measures of prevalence are proportions—such they are dimensionless and should not be described as rates (Friis and Sellers, 1999).

Point prevalence
According to Selvin (1991), point prevalence (often called prevalence “rate”) is the number of affected individuals in a population at a specific point in time divided by the size of the population under consideration. For example, the point prevalence proportion of coronary heart disease in the population age 65 and above in a specific country is the number of existing cases divided by the number of people above age 65 in the country on a specific date.

Period prevalence
Selvin (1991) defines a period prevalence (also usually called a “rate”) as the number of affected individuals in a population plus a count of new cases over a defined period of time divided by the size of the population under consideration. The numerator is the combination of incidence and prevalence (point prevalence) and the denominator is the same as that in the point prevalence proportion measure. This measure is used less since it combines both incidence and point prevalence into a single number, which can be difficult to interpret.
**Duration prevalence**

We also identify a type of prevalence known as duration prevalence. This measure is used in some life table calculations. It is the proportion of time that a person is affected or a group of people are affected. It is a ratio of two durations that can be formulated as follows:

\[
\text{Duration prevalence} = \frac{\text{The duration with a disease or impairment}}{\text{The total duration under observation}}
\]

### 2.2.2 Incidence

Incidence measures the development of a disease or health problem in a population i.e. the frequency of NEW cases in the population at risk during a specified period. The term incidence refers to the absolute number of new cases and the term ‘incidence rate’ refers to the relative number of new cases (i.e. the new cases related to the population at risk). Incidence can be measured in two ways: (i) incidence proportion and cumulative incidence and (ii) incidence rate or incidence density.

**Incidence proportion and cumulative incidence**

For a given interval of time, we can express the increase in incidence number per unit increase in population size. If we measure size at the beginning of the interval and no one enters the population (immigrates) or leaves alive (emigrates) after the beginning of the interval, such a rate becomes the proportion of the people who experience events among those who entered the interval. This quantity is called the incidence proportion (Rothman and Greenland, 1998), which may also be defined as the proportion of a closed population at risk that becomes diseased within a given period of time. This quantity is often called the cumulative incidence (Miettinen, 1976 in Rothman 1998). The incidence proportion can be measured as:

\[
\text{Incidence proportion} = \frac{\text{Number of new cases during stated period}}{\text{Number of persons at risk at beginning of period}}
\]

The numerator relates to events. The denominator is the population at risk (risk set) of the event of interest. It consists of people who do not have the disease and who can have the disease. An example is the so-called attack rate\(^2\) used in epidemic investigation. For example, if we start with 245 susceptible persons and 6 develop the disease over the study period, the risk during that study period = 6 / 245 = .0245, or 2.45 percent. The cumulative incidence is the proportion or

---

\(^2\) Note that the attack rate is not a rate in the usual sense. It is a proportion instead.
probability of healthy individuals who contract the disease during a certain period. An often-used measure is lifetime incidence. Since incidence proportion requires the follow-up of individuals over time, it is a longitudinal measure of disease frequency. Incidence proportion is often expressed in terms of a population multiplier. For example, an incidence proportion of .0245 may be expressed as 24.5 per 1000.

**Incidence rate or incidence density**

Public health usually deals with large populations such as a city, state/province or country where it is not feasible to establish a disease-free group to be followed in time or calculate the person-time of observation for each person. Sometimes, measurement of incidence is complicated by changes in the population at risk during the period when cases are ascertained, for example, through births, deaths, or migrations. Therefore, instead of using the number of people at the start of the observation period as the denominator, we can define a period of observation and determine for each person the actual time period at risk, from the beginning of the period to the time the disease is detected. Or in the case of a person who does not become sick at all, to the end of the period of observation. The duration at risk is calculated by adding together the periods during which each individual member of a population is at risk during the measurement period. The duration is usually expressed in terms of *person-months* or *person-years* (i.e. is called person-time at risk). According to Rothman and Greenland (1998), the number of new cases of disease (i.e. incident number) divided by the person-time is the incidence rate of the population over the period:

\[
\text{Incidence rate} = \frac{\text{Number of new cases during a period of time}}{\sum_{\text{persons}} \text{Time spent free of disease during the same period}}
\]

The incidence rate is also defined as incidence density (Miettinen, 1976a in Rothman, 1998). This measure has also been called the person-time rate, force of mortality or morbidity, hazard rate and disease intensity, although the latter three terms are more commonly used to refer to the theoretical limit approached by an incidence rate as the time interval is narrowed down zero (Rothman and Grenland, 1998).
2.2.3 Prevalence versus incidence

Prevalence and incidence are interrelated. Prevalence is dependent on the incidence and the duration of disease. In a stable situation this relation may be expressed (Rothman and Greenland, 1998) as:

\[ P = \frac{I \times D \times (1 - P)}{1 - P} \]

where \( P \) = prevalence (i.e. not free from disease), \( I \) = incidence and \( D \) = duration.

The left side of the equation is odds (ratio of a probability to its inverse). The denominator on the left side of the equation is the part of the population that is free from the disease. For rare diseases, where \( P \) is low such that \( (1-P) \approx 1 \), then the following approximation may be used:

\[ P = I \times D \]

This relationship states that prevalence varies directly with both incidence and duration. For instance, if the incidence proportion is 10 new cases each month per 10,000 individuals and the duration of observation is 5 months, then the prevalence in this population is 50 cases per 10,000 individuals. This simple relationship, that prevalence is the multiplication of incidence proportion and duration, holds under rather strict steady-state conditions (Selvin, 1991). In realistic situations, these steady-state conditions rarely occur since diseases generally have complex incidence/prevalence dynamics influenced by background characteristics (race, age, sex, medical care etc). Despite only a few people in a group becoming ill each year, if the disease is chronic, the number of diseased people will mount and the prevalence will be relatively high in relation to the incidence. Whereas, if the illness is of short duration, because of either recovery or death or if there is migration of diseased persons from the area, the prevalence will be relatively low.

The differences between prevalence and incidence measurement are conceptualized in Figure 2.1 and summarized in Table 2.1. Incidence is measured over a period of time and prevalence is measured at a point in time. Prevalence differs from incidence in that it refers to status rather than change of status or event. The numerator is the count of existing cases rather than new cases (events). Incidence is about becoming, whereas prevalence is about being something or having something (Young, 1998). Prevalence is always a proportion and is dimensionless, whereas incidence could be either proportion or probability based on the risk set (cumulative incidence, dimensionless) or rate based on occurrence-
exposure (incidence density, dimension is per-unit of person-time). Prevalence data are stock data; incidence data are transition data (Flow).

Figure 2.1 Conceptualization of incidence versus prevalence measurement

The risk set is defined as the total number of individuals at risk of experiencing the event under study during a unit interval, accounting for censoring. In the multistate life table we use probability or rate to measure the risk level. We use risk set (as denominator) to measure the probability of an event. The probability is defined as the number of events to the risk set. Measuring risk level using rate we use exposure (as denominator) instead of risk set. The rate is defined as the number of events per unit of exposure. It is the occurrence-exposure rate i.e. occurrence/exposure. The basis is the duration of exposure or duration at risk. The concept of a rate differs from a probability. The rate used for dynamic analysis depends on the risk period or exposure time while the probability depends on risk set. A probability does not incorporate a direct reference to time whereas a rate is a measure of change per unit of time. A probability is a unitless value between 0 and 1. Rate has a unit and takes any value. Therefore, the estimated parameters of MSLT might vary depending on whether the estimation procedure is based on the risk set or exposure time.
Table 2.1  Prevalence vs. incidence

<table>
<thead>
<tr>
<th></th>
<th>Prevalence</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>Entities</td>
<td>Events</td>
</tr>
<tr>
<td>Source population</td>
<td>At risk to be a case</td>
<td>At risk to become a case</td>
</tr>
<tr>
<td>Time</td>
<td>Static (point)</td>
<td>Dynamic (interval)</td>
</tr>
<tr>
<td>Data</td>
<td>Stock</td>
<td>Flow</td>
</tr>
</tbody>
</table>


Mortality is the incidence of death. The interrelation of incidence, prevalence, and mortality is depicted in Figure 2.2. The mortality rate is a specific kind of incidence rate where the incidence or event is death. The key concept is a change in status, from healthy to sick, from alive to dead. From Figure 2.2, we can say that each new (incident) case enters a prevalence pool and remains there until either recovery or death.

Figure 2.2  Interrelationship of incidence, prevalence and mortality

The prevalence of disease in the population is an indicator of the current stock of health. It is the result of past rates of disease incidence, disease progression, and survival. Over time, the prevalence of disease may change because of increases or decreases in risk factors for various diseases and because of an increasing ability to treat diseases, thus delaying their progression to disability and death. The relative size of the change in incidence rates, recovery and death rates will determine change in disease prevalence (Crimmins and Saito, 2000). Prevalence is often used as an alternative to incidence in the study of rare chronic diseases such as multiple sclerosis, where it would be difficult to accumulate large numbers of incident cases. Again, however, care is needed in interpretation.
2.3 The Sullivan method

The Sullivan method originated in 1964, when a researcher, Sander, proposed a health indicator that combined information on mortality and morbidity. The first example of such an indicator was published in a report of the US Department of Health Education and Welfare in 1969. It contained preliminary estimates of disability free life expectancy calculated using a method devised by Daniel Sullivan (Sullivan, 1971a). He developed a life table technique that collapses both mortality and morbidity into a single composite index of disability free life expectancy. The method uses the observed prevalence of disability at each age in the current population at a given point in time, to divide the years of life lived by a period life table cohort at different ages into years with and without disability (Mathers and Robine, 1997).

A large number of studies presented in the REVES network employed the Sullivan method to calculate estimates of healthy life expectancy (Colvez and Blanchet, 1983; Katz et al., 1983; Wilkins and Adams, 1983; Bebbington, 1988; Crimmins et al., 1989; McKinlay et al., 1989; Rogers et al., 1989; Mathers, 1991). Robine et al., (1995) calculated disability-free life expectancy for over 30 countries using the Sullivan method. Robine et al., (1998) examined trends in disability-free life expectancy in France in the period 1981 to 1991, and concluded that the proportion of life lived free from significant disability increased during the period studied. A similar conclusion was reached by Crimmins et al., (1997) who used the Sullivan life table model to examine trends in life expectancy in the United States from 1970 to 1990. They found evidence of morbidity compression in the 1980s. Hayward and Heron (1999) recently used the Sullivan approach to examine racial heterogeneity in active life expectancy in the United States and they found substantial differences in total and active life expectancy across racial groups.

Up to the second half of the 1980s, most researchers used the Sullivan method to calculate the healthy expectancy. The method uses information on the prevalence of health states in the population. The estimates of active life expectancy for the USA had also been made using double decrement life tables (Katz et al., 1983) with two competing events: disease and death. The Sullivan method reflects the current health status of a real population that is adjusted for current mortality levels. It relies on the input of prevalence measures of the age-specific proportions of the population health conditions, generally gathered in cross-sectional surveys, and on information taken from a period life table (Hauet, 1997). The method is an attractive one, with its simple calculation, low data demand, relative accuracy and easy to interpret outcome.

The Sullivan method resembles the method used in the construction of life tables of working life developed by Wolfbein in 1949 and later on used by others for the construction of tables of working life in studying labor force participation.
The method relies on restrictive assumptions such as the unimodality of the labor force participation curve, entrance into labor force only at ages before the peak and retirement only at ages after the peak, and independence of mortality from labor-force status (Willekens, 1980). These assumptions were implicit in working-life tables that had been published in a large number of countries. After 30 years Hoem and Fong (1976) and Willekens (1980) calculated a multistate table of working life that had no restrictive assumptions. The basic difference in their method is its focus on flows instead of stocks. The Sullivan method uses prevalence, or stock data instead of incidence, or flow data. In the Sullivan method any number of unhealthy states can be adapted—e.g., short term and long term disability, with and without disease, severe and less severe disability. Details of the Sullivan method and its consequences are described elsewhere (Hauet, 1997; Mamun, 2001).

The Sullivan method and the MSLT will give the same estimate of health expectancy if actual transition rates have been stable for a very long time (Barendregt et al., 1994; WHO, 2000). Otherwise, health expectancies calculated by the Sullivan method lag some way behind what is actually happening and consequently, they may indicate a compression or expansion of morbidity when in fact the reverse is taking place (Barendregt et al., 1994). Laditka and Hayward (1999) point out that the development of the multistate method provides demographers with a means to model the dynamics of health in the population in a
more realistic fashion than the Sullivan method. In our study, we make extensive use of multistate life tables to describe the life course of cardiovascular disease.

2.4 The multistate life table

Of the models used by mathematical demographers, the life table model is the most representative. It follows a group of people, born during the same period, experiencing transitions between two or more states over time and age. In the simplest situation, i.e., in the case of a conventional life table, there are two states, namely alive and dead. The transition is irreversible from the former to the latter; the state of death is an absorbing state. The extension of this ordinary life table is the multiple decrement life tables, which distinguishes more than one final absorbing state. A well-known example is cause of death. This model cannot deal with transient states, i.e., states that can be entered and left again. Such states can easily be handled using the multistate life table (MSLT) or increment-decrement life table (IDLT). The MSLT models allow individual members of the life-table population to move into states (increment) and out of states (decrement). These movements or transitions are the fundamental concepts of the MSLT framework.

The MSLT approach can be viewed from two perspectives. The first is the population (macro) perspective: the life table is seen as a description of a stationary population. The second is the life history (micro) perspective: the life table can be viewed as describing the life course or life history of members of a synthetic cohort i.e. the cohort biography (Willekens, 2002). The life history perspective is a micro perspective since it focuses on transitions at the individual level. In the micro perspective, multistate models describe transitions in life and sojourn times in the various stages of life. It visualizes the structure of life. For instance, age-specific transition rates and transition probabilities produce a life structure, and can answer the question of, how t life structure changes when transition rates or probabilities change. The life history perspective is also a longitudinal perspective that focuses on the life course development of cohort members (intra-cohort development) (Willekens, 1987). In the macro perspective, the MSLT describes the dynamics of interdependent sub-populations, where the sub-populations are defined on the basis of personal attributes (health status, place of residence, marital status, employment status etc). The MSLT described in this study is a transition rate model with and without covariates. The multistate model is used to chart the cardiovascular disease process or disease history.

2.4.1 Conceptual issues

A life table is a systematic way to keep track of the experiences (e.g., disease) of a group of people. The experiences of this group can be gathered from birth until
death or over a short period of time. Several conceptual issues that are important in constructing life tables are described here. The issues are: types of MSLT, state and state space, events and transitions, risk and exposure, observation period, and exact age and age in completed years.

Types of MSLT

There are different types of MSLT. The life table can be nonhierarchical or hierarchical, population-based or status-based, uni-radix or multiradix. A hierarchical model is when all states, except the initial state, have only one possible transition into the state. Hierarchical models deal with cases moving sequentially through successive states without reentry. In nonhierarchical models reentry into a given health state is possible. The preference for a nonhierarchical model is determined by the fact that a regaining of health functioning, for example, reentering the non-disabled or non-diseased state, has shown to be a significant force of increment (Rogers, et al., 1989; Mathers and Robine, 1997).

MSLT measurement can be population-based or status-based. The population-based multistate model deals with the entire population irrespective of initial (health) states. It is an unconditional measure, which describes the life history of an average person of a given age, irrespective of the person’s health status at that age. The health status-based measures depend on the health status of the person for which the measure is calculated (Willekens, 1987). Both the population-based and status-based measures are possible for the same data set. For example, the Longitudinal Study of Aging has been used to estimate life expectancy with and without disability of persons who were disabled at age 70, which is status-based, and of the total 70 year-old population, which is population-based (Crimmins et al., 1994). The choice between population-based and status-based depends on the research objectives and research questions. If the primary focus is on change of the health status of the population and on addressing the compression of morbidity discussion, a population-based model is used (Crimmins et al., 1994).

Based on the observation period\(^3\), an MSLT can be of two types: cohort-life table or period-life table. Cohort refers to a group of persons who experience a particular event (e.g. heart stroke) during the same period of time. If the event experienced is birth, then the corresponding cohort is called a birth cohort. The period life tables deal with the study of different cohorts at the same point in time i.e. the cross-section of the population. For example, the life table constructed based on the period from 1968 to 1978 of the Framingham Heart Study (FHS) is a period life table. Depending on the research questions, research objectives and the availability of data we use cohort life table or period life table. As observed in the foregoing, the MSLT could be uniradix or multiradix. In a uniradix life table the initial cohort is concentrated in one state (e.g. healthy or never married) while in the

---

\(^3\) The period during which the events are recorded
multiradix life tables, the initial cohort is allocated to several states. Life tables are also classified according to the length of age interval in which the data are presented. A complete life table contains data for every single year of age. In an abridged life table, information is given only for broader age intervals, e.g. 5-year age groups.

**State and state space**

*State* is defined as a specific attribute of an individual at a given age and time. Attributes are individual characteristics or traits. Persons with a given set of attributes are said to occupy a given state. In general, the attributes are objective and we can measure them. They may also be subjective and refer to values, attitudes, opinions, perceptions, or evaluations (Scott and Alwin, 1998). In modeling, a category of an attribute or a combination of categories of a set of attributes is referred to as a state (Willekens, 1999). For instance, if a person is healthy at exact age 50, he or she occupies the healthy state at that age. At the same age, he or she might be married, may have two children, highly educated and have a high-ranking job. All together constitute his or her attributes at age 50 and healthy is a specific attribute that is a specific state as well.

The number of states is finite and the state variable is a discrete variable. Therefore, the possible number of states can be represented by a discrete or categorical variable. States are classified as recurrent and non-recurrent. A non-recurrent state is entered only once and a recurrent state allows multiple entries. Those states from which exits are possible are called transient states, and the states from which no exit can be made are called absorbing states. In a multistate framework, recurrent and non-recurrent states can be combined to examine transitions between several states, which can be either transient or absorbing (Rajulton, 1999).

The collection of all possible states constitutes the state space. For example, to analyze changes in health status, a state space may consist of two states: healthy and unhealthy. To analyze changes in marital status, a state space may consist of the following five states: never married, cohabitation, first marriage, divorced/widowed and second marriage. In this example, the states never married, first marriage, and second marriage are non-recurrent (because these states can be entered only once), while cohabitation and divorced or widowed are recurrent (they can enter more than once).

**Events and transitions**

A change in attribute is called an event and an event is an outcome of some process. For instance, infection, recovery, entry/exit labor market, graduation, migration, marriage, birth of a child, are life events. These events are associated with underlying causal mechanisms. An event can also be thought of as a change that
places an individual in a new status, which is different from the previous status, he or she was in before the change took place. From a demographer’s point of view, “A vital event is a major change in an individual’s status which leads to a change in composition of the population” (Pressat, 1985). From an epidemiological point of view, recovery from a disease or incidence of a disease is an event. This general definition of an event enables us to visualize events as transitions between statuses (Rajulton, 1999). MSLT is the method that reconstructs the life history based on the given information.

An event is a transition from an origin state to a destination state. For instance, transitions may be from unhealthy to health (recovery), from never migrant to migrant, from single to cohabiting, from second to third births, or unemployed to employed. Events can be non-renewable or renewable i.e. some of the transitions can be experienced only once, whereas other transitions can be repeated. For example, migration and marriage are renewable events. First marriage, first birth, death are all non-renewable events because they only happen at the most once in a lifetime. Some transitions may prevent particular types of transitions or may make further transitions possible. Any transition may be defined by the state before and the state after the transition (Willekens, 1987). A sequence of events can be simply considered as shifts between successive states in the state space (Rajulton, 1992).

An individual experiences an event at a specific age, at a specific length of time since an earlier event, or at a specific calendar time (Rajulton, 1992). Sequence and timing of events are important for the multistate method. Events can be measured by different time scales (see Willekens, 1999; Rajulton, 1999). A sequence of events and stages of life constitute a career, for instance, marital career, fertility career. When two or more careers run simultaneously they are parallel careers. A set of interdependent or parallel careers constitutes the life course. In this study, we construct the life course of cardiovascular disease.

**Risk and exposure**

Risk and exposure are two other concepts important for understanding the multistate method. An individual is exposed or at risk of an event if he or she has a chance of experiencing the event. For example, only healthy individuals are at risk of disease, only married persons are at risk of marital dissolution and only fecund couples are at risk of conception. Risk is conceptualized in a variety of ways. Being at risk means that there is a chance of experiencing an event, i.e. the probability of an event is not zero. If the probability of occurrence is zero it means the individual is not at risk. Life table analysis is a risk analysis.

Exposure is defined as being exposed to the risk of an event; and exposure time is the duration at risk. For example, in case of CVD, the exposure time is from entry into the survey (provided the person is without CVD) to the moment of CVD or end of observation. People generally differ in risk levels and duration at
risk (Willekens, 1999). Risk indicators measure the risk level. The risk indicators are counts, probabilities, rates, odds, and relative risks. Exposure analysis engages measurement and estimation of the length of exposure and the level of exposure for those exposed (Willekens, 2002).

**Observation period**

Multistate life tables reconstruct the individual biography based on partial observation. The partially observed information is gathered during an *observation period*. The observation period is the period during which the events are recorded. The MSLT can be viewed as describing the life histories from birth or from a given age \( y \) (reference age) to the death of the members of a synthetic cohort (Willekens, 1987). The individual biography may be displayed graphically in the Lexis diagram (Figure 2.3). The explanation and interpretation of this diagram is similar to the Lexis diagram presented for the marital status life table by Willekens (1987).

Figure 2.3  Lexis diagram

The Lexis diagram is a popular choice for demographic analysis because of its focus on both personal time (age) and historical time. The horizontal axis of the Lexis diagram represents historical or calendar time and the vertical axis gives the person’s age. Consider the lifeline of individual \( b \). The events are located on the
lifeline. For example, A (Angina Pectoris: AP), B (Myocardial Infraction: MI), C (Other cardiovascular disease: OD) and D (Death: D). On the lifeline b, four events are shown (A, B, C, and D). Figure 2.3 shows an age interval \((x, x+1)\), a time interval \((t, t+1)\) and a cohort interval \((t-x-1, t-x)\). The cohort consists of the group of people who experienced the initial event (birth) in the time interval \((t-x-1, t-x)\).

The timing of events and the age at the time of events may be illustrated in a Lexis diagram. In this regard, two central questions are of importance:

i). How is time or age measured?

ii). When is time or age measured?

When measuring age, it is important to note whether the time intervals are discrete. It may be measured at any point of time in the interval. Suppose an interval has the length of one year. Age is measured either at the exact time of the event, at the end of the year or at the beginning of the year in which the event occurs. In a retrospective survey, the measurement of *age in completed years*, at the end of the interval, is common. It is equivalent to measuring seniority\(^4\) in period difference. This difference is obtained by subtracting the year of occurrence of the initial event from the year of occurrence of the event under study (Willekens, 1987). The measurement of age at the beginning of the year is equivalent to the measurement of the year of birth. A person born between \(t-x-1\) and \(t-x\) is at time \(t\) between \(x\) and \(x+1\) years of age, i.e. of age \(x\) in completed years.

Implicit in the measurement of the timing of events in completed time units is the definition of the observation period. In the Lexis diagram, the observation period refers to the segment of the lifeline for which events are recorded. Events on parts of the lifeline not included in the segment remain invisible. The different ways of fixing the observation period leads to different observation plans (Willekens, 1987). We can present a few of these with reference to the Lexis diagram (Figure 2.3). We refer to read Willekens (1984) for the elaborate typology of observational plans.

Four observational plans may be distinguished, if we assume that the observation period is fixed (say, one year).

a. **Cohort-age (cohort) observation**

Cohort-age observational plan records for a person experiencing an event the cohort to which the person belongs and the *age in completed years* at the time of the event. The parallelogram PQSV in Figure 2.3 shows this observation plan. The observation period extends over two calendar years.

\(^{4}\) The seniority is the time elapsed since age \(y\), the age at which the event occurred that started the process (event-origin) (Willekens, 1987).
b. Period-cohort observation
This observational plan records the calendar year in which the event occurs as well as the cohort to which the person belongs. The parallelogram VQRS in Figure 2.3 shows this observational plan. The observation interval covers two age classes.

c. Period-age (period) observation
This observational plan records the calendar year in which an event occurs, as well as the age of the person in completed years at the time of the event. The square VQSW in Figure 2.3 represents this observation plan.

d. Age-period-cohort (APC) observation
For a person experiencing an event, an APC observational plan records the calendar year in which the event occurs, the year of birth of the person and his/her age in completed years at the time of the event. The triangle VQS in Figure 2.3 shows this observational plan. The observation period covers only one age, one cohort and one calendar year. The data that are recorded by an APC observational plan are commonly referred to as doubly classified data.

The ways in which the observation period is fixed affects the estimation of transition probability in the MSLT. The transition probability, from which all measures that are relevant in MSLT are derived, must be estimated from the data.

**Exact age and age in completed years**
Each of the life table functions refers to a specific age or age interval. In life table measurement, the word “age” is used very precisely, and the precision is emphasized by the addition of the modifier “exact”. For instance, when we say that a person is exact age 0, we mean that person was just born. When that person is exact age 5, that individual has lived exactly five full years (i.e. 5th birthday). In contrast we can say that someone “is 5 years old,” meaning that the person is between exact age 5 and exact age 6, i.e. turned 5 on the last birthday or aged 5 in completed years.

2.4.2 The Markov model

The application of probability theory, statistical inference and matrix algebra has shifted life table analysis from length of life (single decrement) to the structure of life (MSLT), from non-renewable event to renewable or repeatable events and from event of death to various types of events. In life table methodology, this change can be defined as a paradigm shift. Behind this shift, the underlying assumption is that a stochastic process generates the events (Namboodiri and Suchindran, 1987; Rajultan, 1992; Willekens, 1987; 1991) within the defined state space in the individual biography. Willekens (1991), described this shift
as situating demographic analysis as stochastic processes represents a distinct shift in life table theory from accounting or data driven approaches to one that addresses complexity and process.

The probabilistic structure of the MSLT estimated in this study is based on the Markov process with discrete state spaces. We assume that the occurrence of an event is an outcome of a random process. A random variable is defined by a set of possible values; which is associated with a probability. The MSLTs of this study can be described as time-inhomogeneous, finite-space, continuous-time Markov models. In technical terms, the time-inhomogeneous property means that rates of transition can vary between age intervals (Schoen, 1988a). The finite state space of the model is assumed to contain \( J \) \((j=1,2,\ldots,J)\) states, where \( J \) is greater than 1 and a positive integer. The \( J \)th state is an “absorbing” state (e.g. “dead” state, from which there are no decrements).

We define a stochastic process \( \{S(x): x \geq 0\} \) in the state space, where the continuous time parameter \( x \) indicates the exact age attained (Schoen, 1988a). For a population, \( S(x) \) denotes the individual’s position in the state space (i.e. state occupancy) at age \( x \). This information can be summarized in terms of a set of state probabilities, \( P\{S(x) = j\} \), where \( j \) is a specific state. The transition probability between two states is defined as:

\[
p_{ij}(x, x+1) = \Pr\{S(x+1) = j \mid S(x) = i\} \quad (2.1)
\]

where \( p_{ij}(x, x+1) \) represents the probability that a person is in \( j \) at \( x+1 \) provided he or she is in \( i \) at \( x \). The transition intensities from state \( i \) to state \( j \) are defined as

\[
\mu_{ij}(x, x+\Delta x) = \lim_{\Delta x \to 0} p_{ij}(x, x+\Delta x) / \Delta x \quad \text{for } i \neq j \quad (2.2)
\]

where, \( \Delta x \) is a small time interval.

Equations (2.1) to (2.2) denote a Markov process, since the occurrence of an event of interest depends directly on only the occurrence of the state occupied and the person’s age. In state space terminology, it means that a transition from one state \( i \), the origin state, to another state, \( j \) the destination state, depends only on the origin state. This characteristic is known as the Markovian property and constitutes the basis of the construction of MSLT.

A Markov process, therefore, ignores the pathway through which the preceding state was reached. The Markov model has been applied to a variety of social phenomena such as labor mobility, attitudes change, collective violence and in the last two decades to the development of the multistate demography (Rogers, 1975; Schoen and Land, 1979; Willekens and Rogers, 1978; Hoem and Funck-Jensen, 1982; Schoen, 1988a; Crimmins et al., 1994).
2.4.3 The equations

To construct an MSLT we usually have information for a segment of life course. The event history data contains the exact time of transitions, to the month or week or day. In this study, we construct an MSLT using individual level information. First we estimate the transition probabilities. After estimating the transition probabilities other life table statistics are rather easy to calculate. The presentation of equations in MSLT is split up into two main parts. In the first part (A), we describe the estimation procedure of transition probabilities. The second part (B), describes the equations to construct an MSLT.

A. Estimation of transition probabilities

In constructing the MSLT, the first step is to estimate the age-specific transition probabilities from the data. The transition probabilities determine the changes in state occupancies. They are the ultimate parameters of the multistate models (Willekens, 2002). The estimation of these parameters depends on the observational plan. Since the life table provides information on the status of cohort members at consecutive exact ages, the ideal observational plan for life table analysis is the cohort (i.e. cohort-age) observational plan (Willekens, 1987). The formulas to estimate the transition probabilities based on the cohort data are shown in this study. We refer to Willekens (1984) for the estimation of life table transition probabilities from other types of observational plans.

In a cohort observational plan, transitions or events are recorded by age in completed years and by year of birth of the persons experiencing the transition or event (Willekens, 1987). Using micro data, there are two alternative approaches to estimate the transition probability (I) risk set approach and (II) occurrence-exposure rates.

(I) Risk set approach

In the risk set approach, the transition probability is the ratio of the number of events to the population at risk during an age interval. The population at risk is defined in Section 2.4.1 of this chapter. The number of people in state $i$ at exact age $x+1$ is defined as:

$$k_i(x+1) = k_i(x) + entry_i[x,x+1] - exit_i[x,x+1] - death_i[x,x+1]$$ (2.3)

where $k_i(x)$ is the number of people who reach exact age $x$ in state $i$, $entry_i[x,x+1]$ is the number of people who enter in state $i$ in age interval $x$ to $x+1$; $exit_i[x,x+1]$ is the number of people who leave the state $i$ (either make the transition to other states or (right) censored) in the age interval $x$ to $x+1$ and $death_i[x,x+1]$ is the number of deaths in state $i$ within that age interval. Assuming that the censoring
and new entries are uniformly distributed over the age interval $x$ to $x+1$. Therefore, the risk set at age $(x+1)$ is:

$$R_i(x+1) = k_i(x+1) - \frac{1}{2} c_i[x,x+1] + \frac{1}{2} f_i[x,x+1]$$

(2.4)

where, $c_i[x,x+1]$ is the number of people who are right censored within the age interval $x$ to $x+1$, $f_i[x,x+1]$ is the number of people who enter into observation between $x$ and $x+1$, and $R_i(x+1)$ is risk set. It is assumed that they entered in the middle of the age interval and remained in the risk set for half of the age interval.

The transition probability from origin state $i$ to destination state $j$ is defined as:

$$p_{ij}[x,x+1] = \frac{D_{ij}[x,x+1]}{R_i[x,x+1]}$$

(2.5)

where, $D_{ij}[x,x+1]$ is the number of transfers (e.g. CVD) from the origin state $i$ to destination state $j$ during the age interval $x$ to $x+1$.

In matrix notation, the $J$ by $J$ matrix of transition probability can be written as:

$$P[x,x+1] = \begin{bmatrix}
    p_{i1}[x,x+1] & p_{i2}[x,x+1] & \cdots & p_{ij}[x,x+1] \\
    p_{i2}[x,x+1] & p_{i2}[x,x+1] & \cdots & p_{ij}[x,x+1] \\
    \vdots & \vdots & \ddots & \vdots \\
    p_{ij}[x,x+1] & p_{ij}[x,x+1] & \cdots & p_{jj}[x,x+1]
\end{bmatrix}$$

(2.6)

Each column in the $P[x,x+1]$ matrix sums to 1.

**(II) Occurrence-exposure rate approach**

The observed transition rates are estimated by dividing the number of occurrences to the duration of exposures within the age interval $x$ to $x+1$. Following equation 2.2, the observed occurrence-exposure rate from state $i$ to state $j$ can be defined as

$$M_{ij}[x,x+1] = \frac{D_{ij}[x,x+1]}{PY_i[x,x+1]}$$

(2.7)

where, $D_{ij}[x,x+1]$ represents the observed number of transfers from state $i$ to state $j$ between the ages of $x$ and $x+1$, and $PY_i[x,x+1]$ is the risk period or person years in state $i$ within the age interval $x$ to $x+1$. Because of the Markovian assumption the observed rate applies to all persons aged $x$ to $x+1$ in state $i$, regardless of the state they were in at exact age $x$ (Schoen, 1988a). Now the age-specific occurrence-exposure rates can be written as a $J$ by $J$ matrix.
with \( M_{ij}[x, x+1] \) representing the observed rate of transition from state \( i \) to state \( j \) during the interval from \( x \) to \( x+1 \). The diagonal element is expressed as:

\[
M_{ii}[x, x+1] = -\sum_{j \neq i} M_{ij}[x, x+1]
\]

where, \( j \)th state is absorbing state. Each column in the \( M[x, x+1] \) matrix sums to 0.

Defining the matrix of life table transition rates, \( m[x, x+1] \), in the same fashion, we arrive at the orientation equation

\[
m[x, x+1] = M[x, x+1]
\]

where, \( m_{ij}[x, x+1] = \frac{d_{ij}[x, x+1]}{L[x, x+1]} \).

\( L[x, x+1] \) is the number of person years spent by the members of the cohort (in the life table) in state \( i \) between exact age \( x \) and \( x+1 \), \( d_{ij}[x, x+1] \) represents the number of transfers from state \( i \) to state \( j \) between the ages \( x \) and \( x+1 \) and \( m_{ij}[x, x+1] \) is the life-table transition rate. We assume that the observed occurrence-exposure rate, \( M_{ij}(x) \) is equal to life table rate \( m_{ij}(x) \). The configuration of observed \( M(x) \) is similar to that of life-table \( m(x) \).

There are different alternatives to transfer the transition rates to transition probabilities, of which the linear and exponential methods are most common. The methods differ in the distribution of transitions over the age interval. The linear and exponential methods are briefly discussed here.

**Linear assumption**

The assumption of uniform distribution of the events over the observation interval and of births of cohort members over the interval \( t-x-1 \) to \( t-x \) leads to a piecewise linear survival function. That is, transitions occur on average in the middle of the interval. Assuming the distribution of transition is uniform within the interval, the formula yields an estimate of \( P[x, x+1] \) suggested by Rogers and Ledent (1976) and Willekens et al., (1982):

\[
P[x, x+1] = [I + \frac{1}{2} M[x, x+1]]^{-1} [I - \frac{1}{2} M[x, x+1]]
\]

where \( P[x, x+1] \) is the transition matrix defined in equation 2.6, \( M[x, x+1] \) is the occurrence-exposure rates same as in equation 2.8 and \( I \) is an identity matrix. This
linear approach is applied to construct the multistate smoking status life tables (Chapter 5).

**Exponential assumption**
The exponential method assumes that the forces of transition or intensities are constant within the age intervals. The occurrence-exposure rates are converted to probabilities for use in the life table by the following equation:

\[
P(x, x+1) = \exp[-M(x, x+1)]
\]  

(2.11)

Where \( P(x, x+1) \) and \( M(x, x+1) \) are the same matrices as defined before. The relation is derived from the Kolmogorov equation, which is a system of differential equations. This constant rate approach is applied in Chapter 3, Chapter 6 and Chapter 9 of this study. However, when the age interval is small the results from both the methods are equivalent.

Under very general conditions, the linear method is to be preferred on the grounds of simplicity and ease of calculation (Palloni, 2001). However, it is acknowledged that when the underlying risks are decreasing rapidly the assumption of linearity leads to a fair amount of inaccuracy (Schoen, 1988a). That may even lead to impossible negative values when some of the transition rates are very large (Hoem and Funck-Jensen, 1982; Nour and Suchindran, 1984). Therefore, the exponential method, or alternatively the so-called “mean duration of transfer method” (Schoen, 1988a), are to be preferred on the grounds of consistency. In this study, the exponential method is used for most of the applications to get more consistent estimates.

**B. Construction of life table**

Once we have the transition probabilities from the data, we can construct the life tables. Here we briefly illustrate the construction of a life table following the description of life table number of survivors, number of person years (i.e. sojourn time in a state), and life expectancy.

**Life table number of survivors or survival probability (state occupancies)**
The life table number of survivors is the number of people surviving at the beginning of the age interval \( x \) to \( x+1 \) or at exact age \( x \). As with a standard life table, the calculation begins by the specification of a radix for the first age interval. The value of the radix is arbitrary. In demographic literature, it is usually considered to be 100,000 or 10,000 or 1000. If the beginning value is 1, it indicates the survival proportion or survival probability. Throughout this study, we use beginning value 1, i.e. the survival proportion or survival probability. Sometimes the radix population also starts from the observed prevalence in each of the health states from the
LIFE HISTORY OF CARDIOVASCULAR DISEASE AND ITS RISK FACTORS

starting age of that life table (Crimmins et al, 1994). Tuma and Hannan (1984), who refer to the survival probability as the survivor function make a similar distinction. Following Rogers (1975), the relationship between survivorship values and the transition probabilities can be written as follows:

\[ l(x+1) = P(x, x+1)*l(x) \] (2.12)

The survivorship vector \( l(x+1) \), denotes the proportion of survival in a state at exact age \( x+1 \).

**Number of person years lived**

The time spent in each state between two exact ages by a cohort member may be estimated from the life table number of survivors or survival proportion. Assuming that the events that occur in a unit interval are uniformly distributed over the interval, the average sojourn time in years spent in \( i^{th} \) state between ages \( x \) and \( x+1 \) can be approximated by the following expression:

\[ L_i[x, x+1] = \frac{1}{2}[l_i(x) + l_i(x+1)] \] (2.13)

Assuming exponential gross flow functions, the sojourn time spent in \( i^{th} \) state between ages \( x \) and \( x+1 \) can be approximated by the following expression (Chiang, 1984):

\[ L_i[x, x+1] = \frac{l_i(x) - l_i(x+1)}{-\ln[l_i(x+1)/l_i(x)]} \] (2.14)

The number of times a member of the cohort may expect to move from state \( i \) to state \( j \) between exact ages \( x \) and \( x+1 \) is \( m_{ij}(x)L_i(x) \), where, \( m_{ij}(x) \) is the occurrence-exposure rate and \( L_i(x) \) is the duration of exposure in state \( i \) between ages \( x \) and \( x+1 \), irrespective of the state occupied at exact age \( x \) or at any prior age (i.e. population-based measure). For details of this calculation, both for population-based and health status-based measures, we refer to Willekens (1987).

To close the life table we need to assume that the population was stationary above some high age, \( \omega \). This enables us to set up the following equation (Palloni, 2001):

\[ L[\omega, \infty] = l(\omega, \infty)[M(\omega, \infty)]^{-1} \] (2.15)

In our application, we estimate \( M(\omega, \infty) \) from the observed life expectancy at that age using the equation \( M(\omega) = e(\omega)^{-1} \). For example, the life expectancy of the
Massachusetts population between 1989-91 is 4.55 years at age 90, so $M(90+)=1/4.55=0.22$.

**Total number of years lived**
The total number of years lived in state $i$ beyond age $x$ is equal to

$$T_i(x) = \sum_{t=x}^{\omega} L_i(t)$$

(2.16)

where $\omega$ is the highest age or age group. $T_i(x)$ depends on the $(\theta,x)$ survival probability and the time spent in each state beyond age $x$ by a person who reaches age $x$.

**Life expectancy**
The last key measure of the synthetic cohort’s life cycle disease experiences is the expectation of life in each disease or health status. The life expectancy in state $i$ at exact age $x$ is:

$$e_i(x) = \frac{T_i(x)}{l_(x)}$$

(2.17)

where $l_(x) = \sum_i l_i(x)$ is the number of survivors at exact age $x$.

This life expectancy is a population-based measure. It does not depend on the state occupied at age $x$.

The total life expectancy is:

$$e_+ = \sum_i e_i(x)$$

(2.18)

The proportion of the remaining lifetime spent in state $i$ is $\frac{e_i(x)}{e_+(x)}$.

**Population-based and status-based measures**
As we discussed earlier (in subsection *Types of MSLT*), the MSLT can be population-based or status-based. Population-based MSLTs show the life expectancy and health status of the entire life table population by age. Status-based MSLTs indicate the life expectancy by status at $x$, the reference age. The health status-based life table measures depend on the state occupied at the reference age. Their estimation is only meaningful for reference ages at which at least two states are occupied (Willekens, 1987). In status-based measures, a reference age $y$ is selected at which at least two health states are occupied. The minimum reference age, which satisfies this condition, is generally only a few years more than the minimum age at which the transition (such as cardiovascular disease, incidence) starts. In our application, we construct population-based life tables in Chapter 2,
Chapter 6 and Chapter 7. For details of the mathematical construction and interpretation of population-based and status-based life tables we refer to Willekens (1987).

### 2.4.4 The introduction of covariates

Most studies that use the MSLT model and that consider covariates, stratify the population by covariate status and construct life tables for each stratum separately. For example, we construct separate life tables for males and females in order to identify their different cardiovascular life histories (e.g. Chapter 3). To discover the cardiovascular life history of smokers and non-smokers, we construct MSLTs for smokers and non-smokers separately (e.g. Chapter 6). However, to obtain estimates of the net or partial effect of a specific covariate on life table measures, for instance-life expectancy without cardiovascular disease, we also often need to control for other covariates. Therefore, the introduction of covariates in MSLT methodology has potential demographic, epidemiological and sociological applications beyond the computation of traditional “health expectancy” for a group of people.

As we have mentioned earlier (Chapter 1), one of the major aims of this study is to construct multistate life tables in a multiple-covariates context i.e. bridging the multivariate regression model and the multistate life table. In multiple covariate contexts, the age-specific transition rates or transition probabilities are predicted using multiple regression models. Later on these adjusted rates or probabilities for a group of people with specific characteristics are transferred into life table estimates. The definition of specific characteristics depends on the research objectives and interest. For example, multistate life tables are constructed for the high and low risk profiles of cardiovascular disease (Chapter 9). Once we estimate the transition rates or probabilities, the construction of life tables follows the same rules as those formulated in part B.

Recently, multistate life tables with more than one covariate have been applied in the field of work and retirement (Hayward et al., 1990), active life expectancy (Land et al., 1994) and life cycle model of labor force inequality (Hayward and Lichter, 1998). They estimated transition rates for the MSLT model using a log-linear modeling approach:

\[
\ln \mu_i(x) = \exp(\beta_0 + \beta_{1i} z_1 + \beta_{2i} z_2 + \ldots + \beta_{pi} z_p)
\]

\[
= \exp(\beta Z')
\]

where \(\beta\) is a row vector of \(p+1\) regression coefficients. \(Z\) is a \(p\times1\) column vector of \(p\) covariates. The covariates are all discrete variables. The effects of the covariates on the transition rates, as specified in equation 2.19 can be given an interpretation...
like that of typical conventional hazard regression models. The effect of \( z_{k}, \ k=1,2,3,\ldots,p \), when expressed as \( \exp(\beta_k) \) is greater than 1, the transition rate is amplified by the covariate \( k \), whereas if \( \exp(\beta_k) \) is less than 1, the transition rate is depressed by variable \( k \) (Land et al 1994).

The model specified in equation 2.19 was used for the panel data where transition is assumed to occur in the middle of the two panels or intervals (i.e. interval censored) or at a discrete time point i.e. in the absence of the exact timing of an event or censoring. This type of model is called piecewise exponential or discrete time hazard model (Hannan, 1984; Land et al 1994; Hayward et al, 1998; Crimmins et al, 1994). This model was suggested and discussed conceptually by Hannan (1984). Following Hannan’s suggestions and adapting the nonparametric discrete-time approach to hazard regression to the multistate, multiple-covariate context, as described by Trussell and Hammerslough (1983) and by Guilkey and Rindfuss (1987), Hayward and Grady (1990) constructed an MSLT. Gill (1992) provided a survey of this approach and studies of associated statistical issues.

When the exact dates of transitions into and out of the states in continuous event histories are known, the estimation of hazard regression models at individual level with the computation of group-specific multistate life tables is directly applicable (Land et al., 1994). In our study, we use the FHS to construct the MSLT with multiple regression contexts. The FHS recorded disease incidence and mortality at the exact time, i.e. we have continuous event histories. Therefore, we construct hazard regression models at individual level to estimate the adjusted age-specific transition rates to construct the MSLT in covariate contexts. We use parametric and semi-parametric regression models. We describe these hazard regression models and estimations in Chapter 7.

2.4.5 Confidence interval estimation

In statistical analysis, we are not only interested in obtaining a point estimate of a statistic but also an estimate of the variation around this point estimate, and a confidence interval for the true value of the parameter. Traditionally, we depend on the central limit theorem and normal approximations to obtain standard errors and confidence intervals. These techniques are valid only if the statistic, or some known transformation of this, is asymptotically normally distributed (Efron and Tibshirani, 1993; DiCiccio and Efron, 1996). Therefore, we cannot use traditional methods if the normality assumption is not valid. For the same reason, the traditional method is not valid with MSLT. In this study, we describe a non-parametric bootstrapping approach to estimate the 95 percent confidence intervals of the multistate life table statistics.

Prior work on estimating confidence intervals for the MSLT model is relatively limited (Hayward et al., 1999). There are two basic approaches to measuring the CIs
of the outcomes of the MSLT model: analytical and numerical. Analytically, *delta method* is used to estimate the CIs (Ferrucci, 1999). For the numerical approach, *bootstrapping* is used (Calhoun, 1997; Laditka and Wolf, 1998; Hayward et al., 1999; Barendregt and Peeters, 2003). In this study we describe the bootstrapping method, as this procedure is easier to estimates CIs compared to analytical method.

The bootstrap method allows us to obtain an approximation of the distribution of an estimator in the absence of any priori information about the true distribution of the estimator or the original data (Efron and Tibshirani, 1993). It is a resampling technique, which can be parametric or non-parametric. The parametric bootstrapping method employs epidemiologic knowledge about the distribution of the transition variables in the MSLT and number of cases and exposure time or person years at risk (Barendregt and Peeters, 2003). For example, the occurrence of cardiovascular disease is assumed to follow a Poisson distribution. In the non-parametric bootstrap method, $N$ new samples, each of the same size as the observed data are drawn from the observed data with replacement. The statistic is calculated for each new set of data, yielding a bootstrap distribution for the statistic. The fundamental assumption of this method is that the observed data are representative of the underlying population (Efron and Tibshirani, 1993). For a more detailed description of the bootstrapping method and its applicability, see Efron and Tibshirani (1993); Chernick, (1999); DiCiccio and Efron (1996). By resampling observations from the observed data, the process of sampling observations from the population is mimicked (Efron and Tibshirani, 1993). To estimate the CI of the MSLT outcome, a random sample (with same population size) with replacement from the distributions around all variables is drawn, and the model outcomes calculated. If this procedure is repeated at sufficient length (e.g. 2000 times), a distribution of the outcomes will result, from which a CI can be calculated (Hayward et al., 1999).

Non-parametric bootstrapping can be regarded as the gold standard (Hayward et al., 1999; Barendregt, 2002), leading us to choose to focus on this method. Also, when using regression to smooth the transitions by age, non-parametric bootstrapping would be the method preferred. The basic steps of this method are described here, and are similar to the methods described by Hayward et al., (1999).

**Step 1:** Draw a sample from the original data and calculate observed occurrence-exposure rates or statistically estimate transition rates for each possible transition in the state space.

**Step 2:** Use the observed occurrence-exposure rates or the parameter estimated from the statistical models to calculate a set of predicted rates (e.g. smoothed rates) for each possible transition and from these calculate the life table using the equations described in part B for each sample.
Step 3: Repeat step 1 and step 2 a large number of times (e.g. 2000).

Step 4: Calculate summary statistics (e.g. bootstrap standard deviation of 2000 replications) based on the sampling distribution of life table estimates.

Step 5: CIs are estimated from the summary statistics (e.g. the bootstrap standard deviation is used to estimate the standard deviation of the life table estimates).

Repeating steps 1 to 5, we can estimate the confidence intervals of any outcomes we estimate from the MSLT. For an application of this method we refer to Chapter 6 of this study.

2.4.6 Input data for the multistate life table

MSLT calculation depends on what types of data the researcher is going to use. The input data could be from macro to micro levels. Macro-level data refers to grouped or aggregated or tabulated data, such as the number of events during a year by the mid-year population. Micro level data refers to the information obtained from an individual, such as the smoking status of an individual at age 45. When calculating health expectancy in public health research we usually use the micro-level data from a prospective panel study to construct the life table.

Since MSLT calculations stem largely from Roger’s (1975) extension of matrix algebra and its formula as a stochastic process, Schoen (1988a) states that with the exception of a few (e.g., Espenshade, 1983), most multistate projects have used census or vital statistics data. This could be attributed to the fact that at the time of MSLT inception and also during its development in the late seventies and early eighties, the event history data were not available. As a result the majority of the MSLT models have been applied to period data from census or vital registration system to estimate the occurrence-exposure rates, \( M(x) \). Rogers (1975; 1995), Willekens (1980; 1987), Schoen (1988a), Namboodiri and Suchindran (1987) documented details of these issues.

For this study, we focus on micro-level information from a prospective panel study the Framingham Heart Study. Using this longitudinal data, three opportunities arise. First, unlike most census or vital statistics registration data, FHS contains the exact time of transitions (i.e. not only the aggregated number of events over the year). If we know the exact timing of events, we can easily calculate the exact age at the time of events and the exact exposure time. Second, due to the continuous recording of the events (e.g. CVD) we are provided with continuous information for a substantial segment of the life courses. This enables us to
calculate the occurrence-exposure rates or transition probabilities directly from the different sequences or pathways in the prospective data set. The transition probabilities may be estimated directly, due to the fact that the risk set of those making a transition from state $i$ to state $j$ is known (Mills, 2000). After the estimation of the transition probabilities, it is easy to extend these calculations to produce age-specific transition probability matrices (Rajultan, 1992) which is the basis for the construction of MSLT. Third, due to time to time recording of the risk factor status over 48 years of follow-up, the FHS study gives us the opportunity to investigate the risk factor career using the multistate life table technique.

The first step in constructing the MSLT is to prepare the input data. The input data and the construction of MSLTs using macro level information are described in LIFEINDEC (Willekens, 1979), SPA (Willekens and Rogers, 1978), SPACE (Rogers, 1995) and LIPRO (Van Imhoff, 1994) and a few other packages. They are usually applied in multiregional demography. One point that is rarely discussed in the published literature is how to obtain the input data from the micro-level survey data set to produce the life table. Rajulton (1999) developed the package LIFHEHIST for constructing MSLTs with the help of micro data. This package constructs MSLT based on the risk set approach. The major limitation of this package is its inability to calculate the occurrence-exposure rate needed to construct an MSLT. Hence, we cannot use this package to construct an MSLT if left censoring or delay entries are present in the data. Recently, Willekens (2002) extended his LIFEINDEC to SURVEYLIFE, which allows the construction of MSLTs using micro data. However, most of the available software is neither sufficiently general nor flexible to handle a very broad class of applications or to implement alternative solutions.

For this study, we have constructed MSLT based on the long time follow-up data. In most of the cases, we combine general software packages such as S-PLUS, SPSS, STATA, EXCEL that are conducive to mixing preprogrammed routines (such as matrix inversion) with user defined subroutines. The construction of input data using SPSS syntax command is described in Chapter 3 of this study. The construction of input data and subsequently the construction of the MSLT, along with the confidence intervals of estimated life table parameters, are described in Chapter 6. The basic input data for all applications are the exact occurrence-exposure rates.

### 2.4.7 Output from the multistate life table

Output from the MSLT depends on the research question and objective of the study. In general, survival probabilities, sojourn times, expected years of life spent in each health or disease state and number of transitions are the major valuable output data obtained from the MSLT. The MSLT provides insight into the
processes of health change and health state structure of the estimated years in specified states. This output makes the multistate method so valuable.

The multistate life table enables an estimation of the expected magnitude of flows between health states for exact age \(x\) and above to be made. For example, we can calculate the expected number of moves into and out of the “cardiovascular disease” state above a certain age or between ages. These numbers are useful descriptors of individuals’ experiences in health status over the life cycle.

Another important output from the MSLT connects the prevalence proportion and the incidence rate. The prevalence proportion is a product of the incidence rates and the stock of persons in each status, which in turn is a product of temporal prior behavior that has left its stamp on the composition of the population (Schoen, 1988b). This approach allows the depiction of the underlying processes that determine the prevalence proportions of health status at given age.

We illustrate the connections between prevalence proportions and incidence rates by adopting a simple life table model consisting of two “alive” states- healthy (1), diseased (2), and an absorbing state of death (3). The implied duration prevalence (IDP) (i.e. synthetic duration prevalence) of disease among persons aged \(x\) to \(x+1\), \(IDP_2[x,x+1]\), can be expressed in terms of life table notation, as follows (Schoen and Woodrow, 1980):

\[
IDP_2 = \frac{L_2[x,x+1]}{L_2[x,x+1] + L_4[x,x+1]} \tag{2.20}
\]

Where \(IDP_2\) is duration prevalence and \(L_i[x,x+1]\) refers to the person-years lived in the health state \(i\) during the interval \(x\) to \(x+1\). Similarly, the implied prevalence (IP) (i.e. state probability of the synthetic cohort) at exact age \(x\) in terms of life table notation can be defined as:

\[
IP_2 = \frac{l_2(x)}{l_2(x) + l_1(x)} \tag{2.21}
\]

Where \(IP_2\) is point prevalence.

Another important outcome of the MSLT can be produced using the number of person-years function (i.e. \(L_i[x,x+1]\)). We can estimate the life years lost to disease or number of years lived with disease of the life table population. For example, if we are interested in knowing at which age or until which age smokers spend more time with cardiovascular disease compared to non smokers, we can estimate the differences in the number of life years spent with disease by smoking status using the following equations-

\[
LYSD = L_{CVD}[x,x+1] \text{ of smokers} - L_{CVD}[x,x+1] \text{ of non-smokers} \tag{2.22}
\]
Where, LYSD indicates the differences of the life years spent with disease, \( L_{CVD}[x; x+1] \) is the number of years lived with cardiovascular disease.

### 2.5 A critical assessment of the multistate life table

In this chapter, we described the multistate life table and its utility in public health. We illustrated several conceptual issues that are basic to the construction of the MSLT. The MSLT has been explained from a public health point of view, especially as a means to measure the chronic disease life history. To measure health status, we illustrate how the mathematical theory of multistate regional demography can be applied to calculate health expectancy. The fundamental difference between the MSLT approach and the Sullivan method is described. The MSLT method focuses on flows; the Sullivan method focuses on stocks. In the Sullivan method, only changes in the stock (net flows) of the population by health status and age are considered. The MSLT provides a more accurate assessment of the health expectancy in a population. We have also illustrated some of the research questions, questions in public health that can be best answered from a life course perspective. There are several methodological features of this study that may serve as an added contribution to the field of public health research, more especially to summarize the population health status and address to debate of expansion versus compression of morbidity. Some of the points can be summarized as nine basic points.

*First*, the debate between the Sullivan method and multistate method centers on the use of prevalence versus incidence measures of public health. Therefore, the concept of prevalence and incidence is explicitly discussed in this chapter. The precise meaning and exact application of prevalence or incidence measurement is important to get the standard estimation of a population health status. The prevalence is the basis for the construction of the Sullivan method and the incidence is the root for the multistate method.

*Second*, the major strength of the multistate method is its ability to capture the implications of age-related declines and improvements in health (i.e. inflow and outflow). This provides a more accurate assessment of the expected life cycle health experiences of the average person in the population. Moreover, this method allows the explicit assessment of how disability and mortality processes contribute to the structure of population health, or the changing prevalence of health problems associated with age (Crimmins, et al., 1994). These strengths have prompted some scholars to urge for the adoption of the multistate method as the methodological “standard” means of calculating active life expectancy in public health. The Sullivan method yields an inaccurate portrayal of the timing and volume of a cohort's disability and hence could mislead the theory on compression of morbidity. This is because Sullivan method does not take into account the recovery from a particular disability that results from some intervention measures. This method only deals
with the prevalence in population. In addition, the Sullivan method also can produce what appears to be a counterintuitive relationship between changes in disability prevalence proportions and active and inactive life expectancy (Laditka and Hayward, 1999). During periods of rapidly declining mortality, Sullivan based calculations of active life remain stable relative to the growth in overall life expectancy (Bebbington, 1992; Crimmins et al., 1997).

Third, the multistate life table equations are described in a simplified way. Since we use micro data to construct an MSLT- the derivation of transition rates or probabilities is an issue of discussion. Two concepts, the occurrence-exposure rates and the observed transition probabilities are discussed. The concept of a rate differs from a concept of a proportion due to the fact that the former is a dynamic concept that depends on the risk period or exposure time. We distinguished between the risk set and the occurrence-exposure rates approach. In this study, the occurrence-exposure rates are used as the basic input data to construct the MSLT. The occurrence-exposure rates can deal with any censoring (left, right or double) that occurs in the data. Therefore, a multistate life table constructed based on the occurrence-exposure rates is more accurate than the risk set approach.

Fourth, we discussed both population-based and health status-based measures. Population-based measure describes the potential life cycle events for the whole population; status-based tables can be used to compare the perspective life cycles of those who reach specified ages in different health states. It is only in the multistate perspective that a status-based life-table analysis can be carried out, because such an analysis requires simultaneous consideration of all status-specific cohorts.

Fifth, using the multistate life table method, population heterogeneity can be dealt with in a better way than by adopting other methods. In a mortality context, Vaupel (2002) points that heterogeneity should be taken into consideration because not everyone benefits the same way from mortality improvements. That is also true for the chronic disease process. In the MSLT context, the heterogeneity problem is resolved by stratifying the population into sub-populations or groups on the basis of important attributes such as sex, socioeconomic status and risk factors. During the life course, memberships of the groups shift and alter, for instance, by changing the health status. Neither are members of the same group homogeneous; they might differ in many ways and the differences are likely to affect their chances for survival, disease process and other aspects of life.

Sixth, the introduction of multiple covariates in the context of the MSLT is a recent innovation i.e. the combination of multistate life table models with event history models has a more general appeal beyond the ordinary MSLT. For instance, we can link the MSLT of cardiovascular disease and risk factors of CVD and measure the adjusted life expectancy with disease and without disease. That results in the multivariate-multistate life tables. This type of model may be useful for health intervention (e.g. treatment exposure) and policy making. From a life course
perspective, this method will facilitate the integration of the risk factor exposure at different stages (see Figure 1.3) of life and their impact at later ages of life.

Seventh, to evaluate the significance group differences in life table estimates we discussed the non-parametric bootstrapping method. Generally, this approach allows us to estimate variances around the MSLT functions, thereby permitting formal statistical tests for group differences. Using this procedure the statistical competence of the MLST estimates could be addressed.

Eight, a Markov model ignores past experiences and examines the effect of the most recent experiences. For the chronic disease modelling, the duration of a disease is important. In that situation we need an extended Markov model, called the Semi-Markov model. The Semi-Markov model considers duration dependence i.e. the occurrence of an event depends on the length of duration between two successive events. Most-real world problems are not fully Markov in nature- they are often non-stationary, history-dependent and/or not fully observable. In order to solve such problems, we need to extend the Markovian properties that are history dependent.

Ninth, it would appear from the literature that most of the public health researchers have shown that the multistate life table approach is data demanding. To estimate the transition probability either we need to conduct a longitudinal survey (at least two rounds or waves) or collect information from vital statistics or census in at least two time points. In the multiregional demography, MSLTs were also constructed based on the information collected in vital statistics or census (e.g. Rogers, 1975). Day by day, the number of longitudinal studies is increasing in developed countries. Therefore, the utility of the MSLT approach to measure the public health status would not be problematic in terms of data accessibility. For the developing country there could be some alternatives. We may also obtain individual health status on the basis of calendar records, which is very simple and less costly. This calendar recording can be applied to construct the period MSLT. If we measure the health status-based on the activities of daily living (ADL) or self-perceived health or self-reported health, then the calendar history recording will be an alternative to longitudinal study. In that case sampling design will not affect the MSLT results.

In this chapter, we drew an explicit distinction between prevalence and incidence based measures of population health status and shed light on the utility of the multistate life table to public health questions regarding population health status. To address this debate, researchers agreed that when the required data is available, the multistate life table approach provides a better solution than does the prevalence based method. The conceptual issues, mathematical equation, generalization, statistical competence, input and output of the multistate life table as discussed in this chapter are consistently applied throughout this book.
References


LIFE HISTORY OF CARDIOVASCULAR DISEASE AND ITS RISK FACTORS


MULTISTATE LIFE TABLES IN PUBLIC HEALTH


A multistate life table analysis of cardiovascular disease life history

Abstract

We demonstrate a method for producing occurrence-exposure rates from micro data and for using occurrence-exposure rates to construct a multistate life table (MSLT). Using epidemiological data, we measure the burden of cardiovascular disease and its sub-types in terms of life course indicators such as lifetime probability of cardiovascular disease, life expectancy free of cardiovascular disease and life expectancy with cardiovascular disease. We use the 48 year cardiovascular life history of the original Framingham Heart Study (FHS) population, aged 28-62 at study onset, and followed up between 1948 and 1998. We construct the life table for the total population and for males and females separately, to analyze sex differences in the cardiovascular disease life history. We found that at age 40, a FHS participant can expect to live an average of another 38.5 years, of which 84 percent free of cardiovascular disease. At age 50, a male can expect to spend 25 percent of his remaining life expectancy with the disease; a female, 18 percent. The lifetime probability of developing CVD is 67 percent for males and 55 percent for females (60 percent for total population). The number of years lived with disease after age 80 is higher for females compared to males. There is a disparity in the life trajectory of cardiovascular disease between males and females. Although middle-aged males spend more time with CVD, the burden of CVD at later ages of life is higher for females. The method and results we present here are simple and transparent to enable meaningful conclusions to be drawn about the potential burden of cardiovascular disease life history on both total population and male-female separately.
3.1 Introduction

The previous chapter provided a theoretical and mathematical description of the multistate life table and its potential utility in public health. In the present chapter, we have constructed a multistate life table to analyze the life history of cardiovascular disease. The epidemiology of cardiovascular disease has been comprehensively investigated and described during the past half century. Many prospective studies have enabled researchers to identify and quantify the major risk factors for cardiovascular disease (Dawber et al., 1957; Doyle et al., 1957; Chapman et al., 1957; Drake et al., 1957). Numerous studies have confirmed that altering these risk factors causes a reduction of event rates (Pignone et al., 2000; Chalmers, 1999; Hooper, 2000; Wilson, 2000; Yusuf et al., 2001). A recent study using 50 years of follow-up in the Framingham Heart Study (FHS) has found that the incidence of heart failure has declined among females but not among males and that survival after the onset of heart failure has improved in both sexes (Levy et al., 2002). Cardiovascular disease has been identified as a leading cause of disability and premature mortality (Murray and Lopez, 1996). Despite the recognition of cardiovascular disease (CVD) as a paramount health problem in public health, surprisingly little evidence is available on trajectories of CVD and its different manifestations.

Over the past few years, interest in conceptualizing the disease etiology within a life course framework has grown (Kuh, 1997; WHO, 2002). A recent review study by Ben-Shlomo and Kuh (2002) concludes that ‘A life course approach is paradoxical as on the one hand it is intuitively obvious (do we really need research to demonstrate risk accumulation?), and yet on the other hand is empirically complex (do we really have much evidence in support of these models?).’ In this chapter, in which demographic techniques are applied to the epidemiologic information, an initial attempt is made to measure the burden of chronic disease, more specifically cardiovascular disease history, on human life course.

Demographic models of cardiovascular disease morbidity and mortality are very limited at best. Most of the demographic models have combined both morbidity and mortality by focusing on so-called health expectancy (e.g. active life expectancy) (Rogers et al., 1989; Crimmins et al., 1994). The underlying cardiovascular disease etiology of active life has been largely ignored by demographers in general (for exceptions see Hayward et al., 1998; Barendregt and Bonneux, 1998). However, Manton and Stallard (1988) proposed what may be the first demographic model of chronic morbidity and mortality. Much like the model developed here, Manton and Stallard’s approach began by following individuals who are free of (cardiovascular) disease. They were thus able to observe the onset of a given fatal chronic disease, perhaps the onset of another fatal disease and the ending of the disease experience,
which is was death. The substantive aspect of our study is directly based on Manton and Stallard’s conceptual work. We followed individuals as they experienced cardiovascular disease and its subtypes, and observed subsequent disease mortality and morbidity experience. The backbone of such a model is the age schedule at death and disease. Parsimonious models estimating life histories from the age schedules of various events (such as migration, or entry in the labour market) are a time-honored tradition in demography, laid down in the multi-state life table. We present the cardiovascular life history of the 4998 subjects making up the population of the original Framingham Heart Study cohort. This population was aged 28-62 at study onset and followed up between 1948 and 1998.

The objectives of this chapter are two-fold. First, we present a method to obtain observed occurrence-exposure rates from micro data. The observed occurrence-exposure rates are the basic input to construct a MSLT. Second, applying this demographic technique to the FHS cohort, we have derived several life course indicators, such as lifetime risk of CVD, the expected number of years lived with CVD, the expected number of years lived without CVD and the differences in the number of years lived with and without CVD. These population measures, derived from a health-based life history, are essential both for precise assessment of changing health care needs and for intervention preferences.

This chapter is an extension of our previous study (Mamun, 2001; Peeters et al., 2002). From the methodological point of view, there are three basic differences between this chapter and previous works. First, in our previous study, multistate life tables were constructed based on the direct transition probabilities estimated from the micro data. We relied in that study on the concept of risk set. In the present study, the multistate life table is constructed based on the exact occurrence-exposure rates, which give a more accurate estimation of the life table parameters compared to the risk set approach. We demonstrate a method to estimate occurrence-exposure rates using micro data. The novelties of the latter approach are that the exact risk period is counted and left censoring is controlled. Second, the mortality and disease incidence used here were taken from the first 48 years of follow-up of the Framingham cohort; in the previous study we used the first 40 years of follow-up of the same cohort. We calculate the life table for the total population and also estimate the differences in cardiovascular life history between male and female participants of the Framingham Heart Study. Third, we provide a step-by-step demonstration of the multistate methods and estimation procedures.

This chapter is organized as follows. Section 3.2 offers a general overview of the approach that is used in this chapter. Section 3.3 describes the data and methods that are used to describe the life trajectory of cardiovascular disease. In this section, we have provided an overview of the original Framingham Heart Study cohort, after which we go on to distinguish between states and events, to specify different state space and transitions occurring with cardiovascular disease. We
illustrate how occurrence-exposure rates are estimated and finally, show how an MSLT is constructed. The results are described in Section 3.4, where mainly the cardiovascular disease history for the total FHS cohort is described. Finally, we explore the differences in cardiovascular disease life trajectory between males and females based on the life table implied prevalence, the proportion surviving, the proportion of time spent in a particular state, the lifetime probability of events and number of years lived. Section 3.5 closes this chapter with a discussion and conclusion.

3.2 The general approach

In this chapter, we used multistate life table techniques to describe the life history of cardiovascular disease. First, however, a brief description of the cardiovascular disease process and the modeling framework is given.

Cardiovascular disease process

Much information has been disseminated in the past half century regarding the cardiovascular disease process. Cardiovascular disease is a general term for diseases of the heart and blood vessels. Cardiovascular disease typically includes coronary heart disease (heart attack or myocardial infraction and angina pectoris), stroke, congestive heart failure, rheumatic fever and rheumatic heart disease, and congenital cardiovascular defects (American Heart Association, 2002). Heart disease, or coronary heart disease, is a disease caused by atherosclerosis, which leads to the narrowing of the lumen of arteries in the heart (American Heart Association, 2002). It is likely to produce angina pectoris (heart-related chest pain) or a heart attack. A stroke occurs when an artery supplying blood to the brain is blocked causing cells, or even an entire area of the brain to die. The damage resulting from an insufficient supply of blood may include a loss of mental function, muscle function, vision, sensation, or speech, depending on the area of the brain affected (National Stroke Association 2002). A heart attack occurs when an artery feeding the heart is blocked, allowing the cells in the part of the heart usually supplied with oxygen and nutrients by that artery to be damaged and even die. Heart attacks are most often caused by the blockage of an artery already narrowed by atherosclerosis; it can also result from an artery that is blocked because of contractions, i.e. the artery goes into spasm (American Heart Association, 2002).

Cardiovascular disease is of multifactorial etiology. Genetic susceptibility, risky behaviors and age play a primary role in the etiology of CVD (Center for Disease Control, 2002). Modifiable risk factors include high blood pressure, high blood cholesterol, obesity, smoking, diabetes, and physical inactivity and irremediable risk factors are age, genes and family history. The etiology of some of the risk factors is
complex and still unknown. For example, genetic factors and cardiovascular disease etiology are still unresolved (WHO, 2002). For details of cardiovascular disease etiology we refer to WHO (2002); American Heart Association (2002), National Stroke Association (2002); Center for Disease Control (2002).

The modeling framework

Useful public health measures of CVD occurrences are (WHO 1980; 1984): disease free survival, disease free life expectancy or healthy life expectancy, years of life saved or lost, disability adjusted years of life saved and so on. In the public health literature, the life table method is one of the best-known models to measure all these indicators. The life table method is an illustration of duration analysis (Willekens, 1991) where the time clock could be individual time (age), process or sojourn time and calendar or historical time. For chronic disease modeling, individual age is the most important time clock. The life table is based on age-specific occurrence-exposure rates.

The construction of a life table requires information on the dates of some basic variables: time of entry into the study, observation window, date of experiencing the event of interest or failure date, date of censoring or last follow-up or cutoff date. Theoretically, the date of entry marks the beginning of the exposure of an individual or unit under study to the risk of experiencing the event of interest (Namboodiri and Suchindran, 1987). For instance, the first examination date of the FHS original cohort marks the beginning of the risk of cardiovascular disease (provided there is no left censoring present). The observation of an individual ends with occurrence of the event of interest, competing event or the discontinuation of observation. The occurrence of an event unrelated to the event of interest or the discontinuation is known as right censoring.

In the single decrement life table, the attrition is due to a single cause, namely death (Namboodiri and Suchindran, 1987). If censoring exists, attrition can be either the result of having experienced the event of interest or censoring. For the application of life table techniques in the FHS, censoring concerns the right end of the observation window. The occurrence-exposure approach can easily control for censoring.

We have constructed a number of different life tables. In the standard life table, an individual moves from alive to dead. In the multi-state disease model, an individual may exit through death but may also move from “no cardiovascular disease” to disease to death. We use the Russian Doll model or Matroishka model for a series of life tables of increased complexity. The different models divide the disease states up further into more specific disease states, but all states are hierarchically related. Transition into a disease state represents the entry into a disease. In this way, we created a series of multi-state models of the cardiovascular disease process.
To extract the input data for the construction of the life tables using FHS, we made three assumptions:

i. The FHS population is assumed to be homogeneous within each age and sex category.
ii. Censoring is independent of the event of interest.
iii. Exact age at the time of event or censoring is known.

3.3 Data and methods

3.3.1 Overview of the Framingham Heart Study

The Framingham Heart Study (FHS) is considered as the pioneer study into the cardiovascular disease process. At the time the FHS started in September 1948, doctors and even researchers had no idea why 1 in 4 men aged 55 or older developed heart disease (Brink, 2001). William Kannel, senior investigator and director of the FHS from 1966 to 1979 mentioned that “when we started, we were getting death certificates saying that patients had died of acute indigestion”. At the time, the FHS started with one main aim and two subsidiary aims (Dawber and Moore, 1952). The main aim was to secure epidemiological data on arteriosclerotic and hypertensive cardiovascular disease. The two subsidiary aims were (i) to secure data on the prevalence of all forms of cardiovascular disease in a representative population sample and (ii) to test the efficiency of various diagnostic procedures. The first aim was the major thrust of the study, the other two were viewed as by-products.

Under the direction of the National Heart Institute (NHI), in 1948, the Framingham Heart Study (FHS) embarked on an ambitious project in public health research. Now the institute is known as the National Heart, Lung, and Blood Institute (NHLBI). The Framingham Heart Study is now conducted in collaboration with Boston University.

The researchers in this study recruited 5,209 men and women between the ages of 28 and 62, from the town of Framingham, Massachusetts, and began the first round of extensive physical examinations and lifestyle interviews that they would later analyze for common patterns related to CVD development. The original study cohort consisted of respondents of a sample of 2 out of 3 adults, 28 to 62 years of age, who were residing in Framingham, Massachusetts in 1948 (Dawber and Moore, 1952). Of the original cohort of 5209 men and women, 1095 known were alive as of February 1998. Table 3.1 represents the age and sex distribution of the original cohort of the FHS at entry time in 1948 and as of February 1998, respectively.

The subjects have continued to return to the study every two years since 1948, for a detailed medical history, physical examination, and laboratory tests. For those participants who have moved out of the Framingham area and have not come back
for examinations, there is no clinical exam data available. However, telephone and/or mail contact data is maintained on nearly everyone, as a result of which morbidity and mortality information is available. Hospital records and death certificates are obtained. Family and doctors are queried. Event information is therefore quite complete. Also, many persons do come back for examinations even though they have moved out of the area. The participants are so committed to the study that when they come to the Boston area for a visit, they will often call the study to schedule an examination (Paul Sorlie, Personal communication, FHS, 2003).

In 1971, the study enrolled a second-generation group consisting of 5,124 individuals of the original participants' adult children and their spouses. They participated in similar examinations, and are known as the offspring cohort. This landmark study has even started recruiting a third generation since 2001 (National Institute of Health, 2001). In our analysis, we considered the original cohort.

Although the Framingham cohort is primarily white, the importance of the major CVD risk factors identified in this group have been shown in other studies to apply to other racial and ethnic groups, even though the patterns of distribution may vary from group to group (Leaverton et al., 1987). In the past 50 years, the study has produced approximately 1,000 articles in leading medical journals. Until recently, no demographic model has ever been applied, however. The concept of CVD risk factors has become an integral part of the modern medical curriculum and has led to the development of effective treatment and preventive strategies in clinical practice. The Framingham Heart Study continues to make important scientific contributions to public health, especially regarding diseases related to CVD by enhancing research capabilities and capitalizing on the inherent resources of this study. New medical innovations are evaluated and incorporated into ongoing protocols.

<table>
<thead>
<tr>
<th>Table 3.1 Framingham Heart Study original cohort: age-sex distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>29-39</td>
</tr>
<tr>
<td>40-49</td>
</tr>
<tr>
<td>50-62</td>
</tr>
<tr>
<td>100+</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

Source: (http://www.nhlbi.nih.gov/about/framingham/design.htm)

In the FHS, the following diseases were examined (Shurtleff, 1971): Cardiovascular disease (CVD) incorporating all the types of cardiovascular disease listed below:
• Coronary heart disease (CHD): myocardial infarction (MI), Angina pectoris (AP), and Coronary insufficiency (CI).
• Cerebrovascular accident (CA): Stroke (ABI, embolism, haemorrhage and other cerebrovascular accident), and Transient ischaemic attack (TIA).
• Congestive heart failure (CHF).
• Intermittent claudication (IC).

A brief definition of all these diseases is given in Technical Appendix 3.1.

In this study we refer to those who are free of cardiovascular disease as “NO-CVD”. We exclude the subjects with pre-existing CVD at 1st examination (139 respondents) from the population at risk of developing CVD and those who were lost to follow-up (72 respondents). The pre-existing CVD at 1st examination was identified by any one of the following diagnoses at examination 1: definite angina pectoris, definite history of myocardial infarction, definite myocardial infarction by electrocardiogram, doubtful myocardial infarction by electrocardiogram, definite coronary insufficiency by electrocardiogram and history. After excluding 211 (139+72) cases, we found a total of 4998 subjects in the original FHS cohort. Details of the criteria for the original cohort are described in the data manual of the FHS (Cupples et al., 1987; Shurtleff, 1971).

3.3.2 States and events

The life history of an individual is viewed as a sequence of states. Events signify the transitions from one state to another (Willekens, 2002). A state is defined as a specific attribute of an individual. At each age, a person has particular attributes and occupies a particular state. For instance, if a person has no CVD at exact age 32 he occupies the state “NO-CVD” at that age. The number of states a person may occupy is finite. Hence the state variable is a discrete variable. All possible states constitute the state space. The number of persons in a given state is referred to as the state occupancy. State occupancy is a stock variable.

A change of attribute (state) is defined as an event. An event implies a transition between states in the state space. For example, if an individual is healthy at exact age 34 and suffers an MI before exact age 35, the change of attribute from healthy to MI is an event. The types of events depend on the state space. We consider the different state spaces separately. The number of events or transitions during a given interval is a flow variable.

We specified different state spaces and transitions, starting from a simple state space and ending with a complex one. Some of the state spaces and the associated transitions are presented in Figure 3.1. Others are given in Technical Appendix 3.2. We constructed life tables for all the specified models in Figure 3.1. The models
specified in Technical Appendix 3.2 demonstrate the more complex state space and associated transitions. For their application, we needed a large sample population. The FHS was too short.

The first state space is a simple two-state model: the state space is \{alive, dead\}(Figure 3.1(a)). The second state space consists of three states (Figure 3.1 (b)). In the three-state model, we divided the state of alive into two states: NO-CVD (i.e. alive without CVD) and CVD (i.e. alive with CVD). The third state is death, which is an absorbing state. The state space is \{NO-CVD, CVD, Dead\}. Individuals could pass from NO-CVD to death, spending no time in the state cardiovascular disease, or they could first transit through this state. We assumed that there would be no back transition from CVD to NO-CVD i.e. no recovery. In the third model, we decomposed the state of CVD into two states: CHD and other CVD (OCVD). The state space is: \{NO-CVD, CHD, OCVD, Dead\} (Figure 3.1(c)). In this model, OCVD is CVD other than CHD. While back transition could occur in the population this was not taken into consideration for the purpose of our analysis.

Three other models were constructed with the same design except with myocardial infarction, congestive heart failure and stroke instead of coronary heart disease.

Figure 3.1  Multistate life table model of cardiovascular disease
3.3.3 Estimation of occurrence-exposure rates

As we discussed earlier (Section 2.4.3 in Chapter 2) there are two approaches to estimating the transition probabilities. There is the risk set approach that provides direct transition probability and the occurrence-exposure rates. The occurrence-exposure rates are also known as event rates, hazard rates, transition rates or instantaneous rates, depending on the field of study. Occurrence and exposure i.e. cumulative waiting time at risk are two key concepts in rate calculation. This method, according to which a key position is assigned to occurrence-exposure rate, is sometimes referred to as the person-years approach. This person-years approach is widely used in the field of demography to analyze event history data. In the field of epidemiology, Breslow and Day (1987) have used the person-years approach to describe incidence rates. In this approach, the clock begins for each individual at the start of an episode or time of the onset of risk for a specified transition. Both the fixed and time varying variables could be incorporated in this data file. The construction is best introduced by an example using the observational plan depicted in Figure 3.2.

In previous studies calculating MSLTs (Mamun, 2001; Peeters et al., 2002), we estimated transition probabilities directly from the micro data based on the risk set, where the risk set was the number of people who were at the risk of experiencing an event of interest (Namboodiri and Suchindran, 1987). This risk approach is an approximation of the occurrence-exposure rate approach since the denominator is the population at risk (i.e. risk set) instead of the exact risk period of the event of interest. Using the rate method, we can calculate the transition probabilities more accurately. The use of occurrence-exposure rates is a guarantee for the correct estimation of probabilities (Willekens, 2002). Therefore, instead of a risk set, we calculated the occurrence-exposure rate using the observational plan shown in Figure 3.2. Since the transition probabilities are calculated from occurrence/exposure rates, and not from a risk set, the estimated outcomes in this study could be different compared to our previous study.

The Framingham Study has maintained continual follow-up of the participants. In the data available to us, the follow-up is for 48 years, from mid-1948 to mid-1998. We have the exact time (in number of days) of survival or censoring time for each participant from exam 1 until exam 24. There were 72 respondents who were lost to follow-up at different points in time. We excluded these from our analysis. In the end we were left with 4998 (5209-139-72=4998) respondents. To estimate the occurrence-exposure or occurrence-risk set, we needed to define the

---

1 Days are transformed into exact years using the equations:
   \[ \text{Year} = \text{Exam1} + \frac{\text{Survtime}}{365.25} \]
   where ‘Year’ is the exact year of survival or censored, ‘Exam1’ is exact year of first exam and ‘Survtime’ is the number of days (survived or censored) since Exam1.
observational plan. The observational plan gave us a clear picture of the starting, and ending times of a follow-up. We calculated the occurrence-exposure rates based on the observational plan for 48 years of follow-up in the FHS, as illustrated in Figure 3.2.

To depict the timing of the occurrences of an event or censoring and the number of exact person-years contributed from entry to exit time, 4 hypothetical participants entering into the study at the exact time of 1948.7" were taken as an example. These were followed until they experienced events or were censored at the end of the observation time in 1996.5. We have shown the occurrence-exposure rates and risk set in two time scales: (A) calendar period and (B) age. Since life table estimation uses age as the time scale, we estimated occurrence-exposures rates by age.

Consider individual I. He entered the study at age 60.7 y at the first exam held in 1948.7. He died at exact age 75.0 y, which was in 1961.0. Individual II entered the survey at age 35.3 in 1948.7, experienced cardiovascular disease at age 52.2 y (i.e. the age of entry into the CVD state) in 1965.6 and died at age 66.6 y in 1980.0. Individual III entered the study at age 41.5 y and experienced no event of any kind during 47.8 y of follow-up. He is defined as right censored at age 89.3 y in 1996.5. Lastly, individual IV entered the study at age 28.9 y, experienced CVD at age 48.6 y in 1968.4 and died at age 70.2 y in 1990. All four individuals contribute a total of 98.7 y (14.3 + 16.9 + 47.8 + 19.7 = 98.7 y) to the NO-CVD state before making the transition to CVD or death (i.e. considering a multiple exit to CVD or death) and spend 36.0 y (14.4 + 21.6 = 36.0) in the CVD state before making the transition to death.

As age-specific occurrence-exposure rates were required as input for the MSLT, we used age as the time scale. For the construction of the life table, we used a one-year age band. In Figure 3.2 (B), however, we used a 10-year age band for illustrative purposes. Individual II and individual IV together contributed 13.6 y (individual II, 3.6 y and individual IV, 10.0 y) of exposure time to the age band 28.9 y-38.9 y. Similarly, individual II, III and IV contributed a total of 27.1 y (II 10 y, III 7.4 y and IV 9.7 y). Individual IV experienced CVD at age 48.6 y and contributed 0.3 y to age group 38.8 y-48.9 y. The person-years of observation obtained on these participants (in this case 98.8 y for being in NO-CVD state and 36.0 y for CVD) were the same regardless of whether calendar period or age was used as the time scale.

---

\[ \text{The day, month and year can be calculated from the date in exact years. For instance, 1948.7 y can be transferred to day, year and month as follows:} \]
\[ \text{Calendar Year} = \text{trunc(1948)} = 1948 \]
\[ \text{Month} = \text{trunc}(0.7*12)+1=9 \]
\[ \text{Day} = \text{TRUNC}((0.7*12+1-\text{TRUNC}(0.7*12+1))*30.347)+1=13 \]
\[ \text{i.e. 1948.7 y} = \text{September 13, 1948 (for details we refer to Mamun, 2001)} \]
Figure 3.2  Estimation of occurrence-exposure based on the observational plan

**A. Calendar period as time scale**

<table>
<thead>
<tr>
<th>Participant</th>
<th>age at study entry</th>
<th>Person years contributed</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (60.7 y)</td>
<td>Died 1961.0y</td>
<td>NO-CVD: 14.3y</td>
</tr>
<tr>
<td>II (35.3 y)</td>
<td>CVD 1965.6y</td>
<td>NO-CVD: 16.9y</td>
</tr>
<tr>
<td>III (41.5 y)</td>
<td>Died 1980.0y</td>
<td>CVD : 14.4y</td>
</tr>
<tr>
<td>IV (28.9 y)</td>
<td>CVD 1968.4y</td>
<td>NO-CVD: 47.8y</td>
</tr>
</tbody>
</table>

**B. Age as time scale**

<table>
<thead>
<tr>
<th>Participant</th>
<th>age at study entry</th>
<th>Person years contributed</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (60.7 y)</td>
<td>Died at age 75.0y</td>
<td>NO-CVD: 14.3y</td>
</tr>
<tr>
<td>II (35.3 y)</td>
<td>CVD at age 52.2y</td>
<td>NO-CVD: 16.9y</td>
</tr>
<tr>
<td>III (41.5 y)</td>
<td>Died at age 66.6y</td>
<td>CVD : 14.4y</td>
</tr>
<tr>
<td>IV (28.9 y)</td>
<td>CVD at age 48.6y</td>
<td>NO-CVD: 47.8y</td>
</tr>
</tbody>
</table>

Participants in risk set

<table>
<thead>
<tr>
<th>NO-CVD: 13.6y</th>
<th>27.1y</th>
<th>13.3y</th>
<th>18.2y</th>
<th>16.1y</th>
<th>10.0y</th>
<th>0.4y</th>
<th>98.7y</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO-CVD: 72.9y</td>
<td>98.9y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVD: 0.3y</td>
<td>16.7y</td>
<td>17.7y</td>
<td>1.3y</td>
<td></td>
<td></td>
<td></td>
<td>36.0y</td>
</tr>
</tbody>
</table>

Person years contributed
This age-specific person-years information can be used to blend the aggregate-level occurrence (i.e. event-count) and exposure (person-years) into aggregate contingency tables. In order to clarify the multistate life table calculation using occurrence-exposure rates, the basic algorithms are described here. A 3-state CVD model of the FHS original cohort was chosen as an example, as this was the simplest MSLT among those presented in this chapter. The calculation of observed occurrence-exposure rates using SPSS syntax is presented in Technical Appendix 3.3. The observed occurrences, exposures and occurrence/exposure rates for the basic cardiovascular disease model (Figure 3.1(b)) are presented in Technical Appendix 3.4.

In the FHS, some transitions could occur simultaneously. For example, some persons experienced CVD and death on the same day. We assumed that they had experienced two events at the same time. The first event was then the transition to CVD and the second was death. For the first event, the exposure time ran from entry into observation to the event. For the second event, the exposure time was zero.

How to obtain the input data from the micro data set to actually produce the multistate life table is something that is rarely discussed in the published literature. The basic data matrices contain two types of information: the first is the number of events or transitions from each origin state to the destination state, the second is the exposure time. The advantage of this input data is that it is simple to calculate the MSLT using Excel worksheet.

### 3.3.4 Multistate life table construction

We constructed a multistate life table that starts at age 40. At age 40, everyone was in the NO-CVD state. As the multistate life tables that we constructed were hierarchical, the life table equations presented here are in simple equations instead of the matrix notation that we described in Chapter 2 of this study. A step-by-step description of the construction of the life table is described in the following.

**Transition probabilities**

When constructing the MSLT, the first step was to estimate the age-specific transition rates or probabilities from the data. The empirical transition rate was estimated by dividing the number of occurrences by the duration of exposure during the age interval \( x \) to \( x+1 \). The occurrence-exposure rate from state \( i \) to state \( j \) was calculated using equation 2.7 in Chapter 2.2. The occurrence-exposure rates were converted to probabilities by assuming that the occurrence-exposure rates remained constant within an age interval (Schoen, 1988a). The total probability of leaving the NO-CVD state (\( q_{no-cvd} \)) in a unit interval (one year) was calculated from the death rate (\( M_{no-cvd,d} \)) and the transition from NO-CVD to CVD (\( M_{no-cvd,cvd} \)) using:
\[ q_{\text{no-cvd}}[x, x+1] = 1 - \exp(-M_{\text{no-cvd,d}}[x, x+1] - M_{\text{no-cvd,cvd}}[x, x+1]) \]

The probability of transit from the NO-CVD state to the CVD state \((q_{\text{no-cvd,cvd}}[x,x+1])\) while taking the competing risk into account (Manton and Stallard, 1988) was calculated by:

\[ q_{\text{no-cvd,cvd}}[x, x+1] = q_{\text{no-cvd}}[x, x+1] \left( \frac{M_{\text{no-cvd,cvd}}[x, x+1]}{M_{\text{no-cvd,cvd}}[x, x+1] + M_{\text{no-cvd,d}}[x, x+1]} \right) \]

Similarly, the probability of transit from the NO-CVD state to death \((q_{\text{no-cvd,d}}[x,x+1])\) was calculated by:

\[ q_{\text{no-cvd,d}}[x, x+1] = q_{\text{no-cvd}}[x, x+1] \left( \frac{M_{\text{no-cvd,d}}[x, x+1]}{M_{\text{no-cvd,cvd}}[x, x+1] + M_{\text{no-cvd,d}}[x, x+1]} \right) \]

The transition probability to death from CVD was estimated by:

\[ q_{\text{cvd,d}}[x, x+1] = 1 - \exp(-M_{\text{cvd,d}}[x, x+1]) \]

**Survival probability at exact age \(x\)**

The total survival probability at age \(x\) is denoted by \(l(x)\) (see Section 2.4.3). We used radix 1 at age 40. We assumed that at age 40, everybody was free of CVD, therefore, \(l_{\text{no-cvd}}(40) = 1\) and \(l_{\text{cvd}}(40) = 0\). The sum is the total probability of surviving at age 40 i.e. \(l(40) = l_{\text{no-cvd}}(40) + l_{\text{cvd}}(40) = 1 + 0 = 1\). The probability of surviving in state NO-CVD and CVD at age \(x+1\) was calculated by:

\[ l_{\text{no-cvd}}(x+1) = l_{\text{no-cvd}}(x) * [1 - q_{\text{no-cvd}}[x,x+1]] \]

and

\[ l_{\text{cvd}}(x+1) = l_{\text{cvd}}(x) [1 - q_{\text{cvd,d}}[x,x+1]] + l_{\text{no-cvd}}(x) * q_{\text{no-cvd,cvd}}[x,x+1] \]

**Person years lived or exposure time**

The number of person years or duration at risk between the ages of \(x\) and \(x+1\) in the state NO-CVD or CVD was calculated using equation 2.14 in Chapter 2. Schoen (1988a) discusses the advantages of the exponential approach. However, if a cell frequency is empty (i.e. \(l_{\text{cvd}}(x) = 0\) or successive values of two cells are equal (i.e. \(l_{\text{cvd}}(x) = l_{\text{cvd}}(x+1)\)), application of this method is problematic. In the first case, the denominator remains undefined and in the second case both numerator and denominator become zero. As a result, the person-years value in that age interval remains undefined. Therefore, in that extreme situation, we assume the uniform
distribution of the events and apply the linear formula given in equation 2.13 of Chapter 2.

In our application, \( L_{cvo}^{[40, 41]} \) was estimated using linear approximation instead of exponential. Life tables are constructed from age 40 onwards. They were closed at age 90 using the Massachusetts life expectancy at age 90 for 1989-91 (males 3.93 years, females 4.76 years, total population 4.55 years) (Centers for Disease Control and Prevention, 1989-91). We assumed that \( M_{no-cvo}^{[90+]} = M_{cvo}^{[90+]} = 1/e(90+) \). For males, for instance, \( e(90+) = 3.93; M_{no-cvo}^{[90+]} = M_{cvo}^{[90+]} = 1/3.93 = 0.25 \); \( L_{no-cvo}^{[90+]} = L_{cvo}^{[90+]} / 0.25 \) and \( L_{cvo}^{[90+]} = L_{no-cvo}^{[90+]} / 0.25 \). We assumed that mortality rates beyond age 90 were same for CVD and NO-CVD subjects.

**Total person years lived**

The total number of person years lived in the state NO-CVD and CVD beyond age \( x \) was calculated taking the sum of the values of person years lived or exposure time beyond age \( x \). The formula is:

\[
T_{no-cvo}(x) = \sum_{t=40}^{90} L_{no-cvo}(t, t+1) \quad \text{and} \quad T_{cvo}(x) = \sum_{t=40}^{90} L_{cvo}(t, t+1).
\]

**Life expectancy**

The population-based life expectancy was calculated using the following formula:

Free of cardiovascular disease, \( e_{no-cvo}(x) = T_{no-cvo}(x) / l(x) \)

and with cardiovascular disease, \( e_{cvo}(x) = T_{cvo}(x) / l(x) \),

where \( l(x) = l_{no-cvo}(x) + l_{cvo}(x) \). This is the average number of years an individual who is alive at age \( x \) may expect to stay in a state beyond age \( x \).

In Technical Appendix 3.5, a basic cardiovascular multistate life table for the FHS cohort is given.

### 3.4 Results

The cardiovascular disease histories are presented in three Sections. In Section 3.5.1, the observed event occurrences in the FHS are given. In Section 3.5.2, the cardiovascular life history of the FHS cohort (males and females combined) is described and in Section 3.5.3, the differences in the cardiovascular disease history of males and females are explored. The last two Sections contain the results of the life table calculation. We have mainly presented the implied prevalence, survival probability, life expectancy free of disease and with disease, and the differences in the years spent with disease.
3.4.1 Observed event occurrences in the FHS

At the onset of the study, the 4998 cardiovascular disease-free Framingham Heart Study members ranged between the ages of 28 and 62; 45 percent were male. Over 48 years of follow-up, 57 percent (64 percent male and 52 percent female) of this cohort developed cardiovascular disease and 77 percent (83 percent male and 72 percent female) died (Table 3.2). Of the original cohort, 37 percent developed coronary heart disease (male 47 percent and female 30 percent), 22 percent suffered an acute myocardial infarction (30 percent male and 16 percent female), 17 percent developed congestive heart failure (18 percent male and 17 percent female) and 16 percent suffered a stroke (15 percent male and 17 percent female). It is a well-established fact that cardiovascular disease mortality and incidence probabilities increase with increasing age, as does mortality from other causes. The quantification of the population burden of different diseases or determination of how this is distributed across age groups is not intuitive. The creation of multistate tables enables analyses to be made of the lifetime probabilities of disease, proportion of survival and expected sojourn times in the various health states.

### Table 3.2 Number of outcomes studied in FHS

**A** Total number of deaths, and number of deaths from cardiovascular disease during period of observation

<table>
<thead>
<tr>
<th></th>
<th>All deaths</th>
<th>NO-CVD to death</th>
<th>CVD to death</th>
<th>CHD to death</th>
<th>MI to death</th>
<th>CHF to death</th>
<th>Stroke to death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>1847 (83%)</td>
<td>615 (28%)</td>
<td>1232 (55%)</td>
<td>915 (41%)</td>
<td>596 (27%)</td>
<td>373 (17%)</td>
<td>301 (14%)</td>
</tr>
<tr>
<td>Females</td>
<td>1982 (72%)</td>
<td>832 (30%)</td>
<td>1150 (42%)</td>
<td>678 (24%)</td>
<td>370 (13%)</td>
<td>400 (14%)</td>
<td>400 (14%)</td>
</tr>
<tr>
<td>Total</td>
<td>3829 (77%)</td>
<td>1447 (29%)</td>
<td>2382 (48%)</td>
<td>1592 (32%)</td>
<td>966 (19%)</td>
<td>773 (15%)</td>
<td>701 (14%)</td>
</tr>
</tbody>
</table>

**B** Transition from NO-CVD to CVD and its subtypes states

<table>
<thead>
<tr>
<th></th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>CHF</th>
<th>Stroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1415 (64%)</td>
<td>1036 (47%)</td>
<td>672 (30%)</td>
<td>402 (18%)</td>
<td>334 (15%)</td>
</tr>
<tr>
<td>Female</td>
<td>1428 (52%)</td>
<td>838 (30%)</td>
<td>439 (16%)</td>
<td>469 (17%)</td>
<td>465 (17%)</td>
</tr>
<tr>
<td>Total</td>
<td>2843 (57%)</td>
<td>1874 (37%)</td>
<td>1111 (22%)</td>
<td>871 (17%)</td>
<td>799 (16%)</td>
</tr>
</tbody>
</table>
3.4.2 Males and females combined

**Implied prevalence**
The implied prevalence (IP) or state probability of the synthetic cohort in terms of life table notation was defined in equation (2.21) in Chapter 2. The implied prevalence at different ages reveals how different states evolve over the life course. They relate to the synthetic biography generated by the multistate life table. The prevalence of the life table population in each of the disease states is shown in Figure 3.3. This simple measure has intuitive appeal. The proportion of respondents in various states at each exact age illustrates how survivors who are not in the disease state (i.e. NO-CVD) at exact age 40 in the original cohort of FHS would be distributed over the state space at successive one-year intervals.

The proportion of the life table survivors in the disease states increases regularly with age, while the proportion with NO-CVD decreases sharply after middle age, either by death or by making the transition to CVD. Some 54 percent of the 90 year-old life table survivors were found to have experienced cardiovascular disease, more than half of which was coronary heart disease, the rest being other cardiovascular disease. Figure 3.3 illustrates the steady progression from NO-CVD to CVD.

![Figure 3.3: Implied prevalence of disease in the life table population](image-url)
Survival probability

The survival probability of a cohort of cardiovascular disease-free 40 year-old persons is demonstrated in Figure 3.4. The light area represents the proportion of a cohort that is alive at \( x \) and free of cardiovascular disease and dark area represents the proportion that has a history of cardiovascular disease (Figure 3.4(a)). Cardiovascular disease states are differentiated in Figure 3.4(b) (using Figure 3.1(c)), into history of coronary heart disease (dark area) and other cardiovascular disease (patterned area). Similarly, three other figures (Figures 3.4(c), 3.4(d), 3.4(e)) demonstrate the proportion of the cohort surviving with myocardial infarction, congestive heart failure and stroke.

These survival curves indicate survival with and without cardiovascular disease and its different subtypes, given that the cohort population was free of cardiovascular disease at age 40. As expected, the survival of both the NO-CVD and CVD life table proportion decreases as age increases. For instance, of the people free of CVD at age 40, nearly 50 percent will survive until age 80, and at age 80, one third will have cardiovascular disease.

Figure 3.4 Survival of a cohort of cardiovascular disease-free 40-year-old population

(a) CVD
(b) CHD

(c) MI
Life expectancy

The added value of the multi-state life table lies in its ability to synthesize the consequences of age-specific incidence rates and to calculate life expectancies in specific disease states. This is achieved by adding up the measure of years lived without a history of disease and with a history of disease, indicating more accurately the potential public health burden of the disease. Total life expectancy (LE) and residual life expectancy free of disease from specified ages, based on a population free of cardiovascular disease at age 40 is presented in Table 3.3. The number in parentheses represents the proportion of time spent without the relevant
cardiovascular disease subtype. Total life expectancy at the age of 40 was 38.5 years, which is consistent with a relatively healthy population (Table 3.3) (Leaverton, 1987). At age 50, a participant of FHS cohort can expect to survive 29.32 additional years, of which 23.11 (79 percent) free of cardiovascular disease and the residual 6.21 (21 percent) years with cardiovascular disease. Of the residual life expectancy at age 50, 4.27 (15 percent) years are spent with coronary heart disease. Less time is spent with MI (7 percent), congestive heart failure (3 percent) and stroke (4 percent). Although it is possible for individuals to be in more than one disease state at any point in time, we have not specifically modelled this co-morbidity.

Table 3.3  Life expectancy and residual life expectancy free of disease at specified ages, based on a population free of cardiovascular disease at age 40, FHS

<table>
<thead>
<tr>
<th>Age</th>
<th>Total LE</th>
<th>Life expectancy (proportion in %) free of a history of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>40</td>
<td>38.48</td>
<td>32.27(84)</td>
</tr>
<tr>
<td>50</td>
<td>29.32</td>
<td>23.11(79)</td>
</tr>
<tr>
<td>60</td>
<td>21.40</td>
<td>15.45(72)</td>
</tr>
<tr>
<td>70</td>
<td>14.30</td>
<td>9.35(65)</td>
</tr>
<tr>
<td>80</td>
<td>8.57</td>
<td>5.03(59)</td>
</tr>
</tbody>
</table>

3.4.3  Male-female differences

The number of transitions for all the examined cardiovascular events is consistently higher in males than in females (Table 3.2) but the greater longevity of females means that the burden of disease can be higher for females. In this section, therefore, the emphasis is on comparing the burden of cardiovascular disease between males and females in respect of the implied prevalence, life expectancy (in proportion of time), lifetime probability and number of years lived with a history of cardiovascular disease and its subtypes.

Implied prevalence

The life table prevalence of cardiovascular disease and its subtypes by sex are presented in Figure 3.5. Males have a higher prevalence of cardiovascular disease than females at all ages. This difference is even higher for coronary heart disease. At age 75, the prevalence of CVD is around 10 percentage points higher for males compared to females (30% vs. 20%). It is nearly 20 percent higher for coronary heart disease. The main reason for this very considerable difference is that males
experience more myocardial infarctions than females. For the other subtypes of cardiovascular disease, the difference in implied prevalence is smaller.

Figure 3.5  Implied prevalence of a history of cardiovascular disease by sex
Cohort survival
Male and female survival (in percentages) of the life table cohort free of cardiovascular disease is presented in Figure 3.6. Over time, females not only survive longer but also survive longer free of cardiovascular disease compared to males (Figure 3.6). The differences in male-female survival start immediately after age 40 (nearly at age 42) and reach 21 percent (maximum) at age 78. At age 90, the difference is still around 11 percent.

Life expectancy
The multistate life table estimates the expected number of years lived in a different state. We have presented the life expectancy in a state by sex (Figure 3.7). The results are presented in terms of the life table percent of lifetime spent in a
particular state. At age 40, a male can expect to survive 81 percent of his remaining life free of cardiovascular disease and a female of that age can expect 86 percent. At age 70, a man can expect 13 percent less years free of cardiovascular disease life expectancy than a female. Throughout life, females can expect to remain free of cardiovascular disease longer than males. Male-female differences are also substantially higher for coronary heart disease and myocardial infarction. Males and females can expect nearly an equal percentage of lifetime in the disease states of congestive heart failure or stroke.

Figure 3.7  Life table percentage of lifetime spent in a state by sex

![Graphs showing life table percentage of lifetime spent in a state by sex.](image)
Lifetime probability of disease
The life table offers a simple method for the calculation of lifetime risks, automatically accounting for competing causes of morbidity and mortality. The lifetime risk of developing cardiovascular disease is presented in Table 3.4. For the synthetic cohort derived from transition rates within the Framingham Cohort, the lifetime probability, at age 40, of developing any cardiovascular disease, including sudden cardiovascular death, is 67 percent for males and 55 percent for females (60 percent for the total population). For a 40 year old male and female without cardiovascular disease, the lifetime probability of developing coronary heart disease is 50 percent and 33 percent respectively, while the probability of an acute myocardial infarction is 33 percent and 17 percent, respectively. The lifetime probability of developing congestive heart failure for NO-CVD men and women at age 40 is 20 percent vs. 18 percent. One in five NO-CVD 40 year-old women and one in six NO-CVD 40 year-old men will suffer a stroke at some point in time. The higher lifetime probabilities of stroke in females are largely caused by the greater female life expectancy. The probabilities at age 40 of developing coronary heart disease, stroke or congestive heart failure before the age of 70 are all greater in males than females (32% vs. 16%, 6% vs. 5%, or 8% vs. 5%, respectively).
### Table 3.4

Lifetime risk of developing cardiovascular disease (%) for cardiovascular disease-free individuals at age 40

<table>
<thead>
<tr>
<th></th>
<th>Probability of developing disease before age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within lifetime</td>
</tr>
<tr>
<td>CVD</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>67</td>
</tr>
<tr>
<td>Female</td>
<td>55</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
</tr>
<tr>
<td>CHD</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>50</td>
</tr>
<tr>
<td>Female</td>
<td>33</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
</tr>
<tr>
<td>MI</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>33</td>
</tr>
<tr>
<td>Female</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
</tr>
<tr>
<td>CHF</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>20</td>
</tr>
<tr>
<td>Female</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
</tr>
<tr>
<td>Stroke</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16</td>
</tr>
<tr>
<td>Female</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
</tr>
</tbody>
</table>

### Differences of the years spent with disease

Not only do women survive longer, they can expect to spend more time free of cardiovascular disease. They moreover have less lifetime risk of developing cardiovascular disease. However, the burden of the disease could be higher as they survive longer, with or without disease. The male-female differences in number of person years lived in a disease state are presented in Figure 3.8. Before age 80, male exposure time to cardiovascular disease is higher compared to that of the female, which is substantially higher in the age interval between 50 and 75. At older ages (usually after age 80) females spend more time with cardiovascular disease compared to males. Throughout most of life, the differences in the number of person years lived with myocardial infarction is higher for males compared to females. Although male-female differences for the other cardiovascular subtypes (stroke and congestive heart failure) are negligible before age 80, they are higher for females thereafter.
3.5 Discussion and conclusion

In this chapter, we presented an analysis of the life history of cardiovascular disease of a white American population: the Framingham Heart Study Original Cohort. The published literature rarely discusses how to obtain the input data from micro data needed to produce a multistate life table. This study has demonstrated how the basic input data, i.e. occurrence-exposure rates, can be obtained. The enormous impact of cardiovascular disease on the human life course is translated into life years lost to disease and life years lived with a history of disease. We distinguish the life expectancy with and without cardiovascular disease, lifetime risk of developing the disease and the difference between males and females in number of years lived in a disease state.

From the age of 40, nearly two-thirds of the men (67 percent) and more than half (55 percent) of the women were shown to develop cardiovascular disease within their lifetime. Lloyd-Jones (1999) reported that one in two men and one in three women would develop coronary heart disease from the age of 40, which is consistent with our study (males 50 percent and females 33 percent). In addition, one out of six men and one out of five women will at some point suffer a stroke. One in three men and one in six women will suffer an acute myocardial infarction at some time. We have shown that the greater longevity of women is the primary...
cause of both their greater lifetime probability of stroke and the greater number of years of life lost for an equivalent disease, as compared to men.

Cardiovascular disease not only reduces life expectancy but is also a major cause of morbidity. Its potential contribution to population morbidity is highlighted, with 21 percent of this synthetic cohort’s residual life expectancy from the age of 50 spent with the cardiovascular disease. These results exhibit the utility of transforming epidemiological data into time-based health policy measures. While epidemiological data enables the number of coronary heart disease events and deaths to be predicted, the multistate life table technique enables estimation of the overall potential burden of specific diseases in terms of years of life lost to and lived with disease. This collective effect of differences in disease incidence and mortality probabilities cannot be intuitively estimated but is important for health care development.

Previous analyses have reported the burden of cardiovascular disease to be a loss of approximately 15,300 years of life and approximately 3,000 years lived with disability in countries such as the USA and Western Europe (Murray and Lopez, 1996). However, more detailed analyses have hitherto not been available. The advantages of the present analysis are the range of cardiovascular disease sub-types; long time follow-up and age groups examined the accuracy of disease definitions and the internal consistency of the various transition rates within the Framingham Heart Study.

One of the model’s strengths is that it symbolizes the relationships within a single, homogeneous historical cohort. However, one of its major limitations is also derived from this property. Because of the long follow-up and the broad age range at inception of the cohort, the forces of mortality and disease incidence by age are a mixture of cohort and period effects. At younger ages, the cohorts are exposed to the higher mortality of the older periods; at older ages, the cohorts are exposed to the lower mortality of more recent periods. As a result, transition rates for the intermediate ages are derived from a number of different periods. In addition, coronary heart disease case-fatality and incidence rates changed significantly during this period in the USA (Rosamond et al., 1998; McGovern et al., 1996; Sytowski, 1990). Because of the advancement of medical technology and diet intake, the population is surviving longer with cardiovascular disease; as a result, we expect the life expectancies with cardiovascular disease presented here to be less than those for current low mortality populations. An analysis of life tables constructed solely using the period between approximately 1970-1990 indicated that total life expectancy and life expectancy with cardiovascular disease at the age of 50 were 1.0 and 0.07 years higher than those presented here (Peeters et al., 2002). Therefore, our results suggest that the total and sex specific life expectancies by total cardiovascular disease and its subtypes presented here are an approximation of those experienced by a similar population today.
Another limitation is that the current model structure is primarily of use for descriptive rather than interventional analyses. Here, a unidirectional transition is used as the simplest way to confine all time spent with a history of cardiovascular disease without the creation of further mixed disease states. Addition of back-flow to the model would require age and sex specific transitions from all disease states. Our data does not permit any back flow from a disease state to NO-CVD state (i.e. CVD to NO-CVD). But some repeated transitions from one disease-state to another are possible. For instance, in model 3(c) the OCVD and CHD could be communicable states. However, when the required disease state is the end state, the results from that model would not be influenced by the reverse transitions. For instance, in model 3(c), we were interested in estimating life table parameters for CVD or CHD or CHF or MI or Stroke but not for the OCVD state. Some of the models are structured in Technical Appendix 3.2, which includes many transitions that are possible theoretically. We could not apply them since the life table method requires more power (sample size) than was available from the original Framingham cohort. While this structure is appropriate for the descriptive analyses presented here, a more biological pathway would be preferred (for example allowing transitions from coronary heart disease to congestive heart failure transitions or remission from angina pectoris) for any interventional analyses. While we have demonstrated different model structures of cardiovascular disease process in Technical Appendix 3.2, we applied a more consistent model 3(b) in the rest of the study (Chapter 6, Chapter 7 and Chapter 9).

Entering the 21st century, one of the most important dilemmas both developed and developing societies will face is how to maximize the health of the elderly. In this regard, cardiovascular disease intervention is one of the major targets for improving population health. Chronic diseases are long-term illnesses that are rarely cured. These diseases can become a significant health and financial burden to not only those persons who have them, but also their families and the nation’s health care system. Chronic conditions such as heart disease, diabetes and arthritis negatively affect the quality of life, contributing to declines in functioning and an inability to remain in the community. The method and results presented here is a simple and transparent one to enable meaningful conclusions about the potential burden of cardiovascular disease life history on both the total population and on the male-female population separately.
References


Technical Appendix 3.1

Glossary

**AP**: Angina Pectoris.

**MI**: Acute Myocardial Infarction. Acute manifestation of CHD: acute blocking of heart vessel, leading to death of the muscle dependent from that vessel. May be silent, and pass unnoticed, but will then cause typical ECG changes. Called **MI** in the FHS data-files.

**CHD**: Coronary Heart Disease. Identical to **IHD** (Ischaemic Heart Disease). Atherosclerotic disease of the vessel walls of the heart, which leads to chronic manifestations of ischaemia: Angina Pectoris (AP) or ECG manifestations, and acute manifestations: Acute Myocardial Infarction, (MI) or unstable angina pectoris (called Coronary Insufficiency, CI, in FHS).

**CHF**: Congestive Heart Failure. Failure of the heart muscle to deliver sufficient power to pump blood, and deliver all the needed oxygen. May or may not be a consequence of CHD. Is not yet estimated as part of the FHS-CHD model.

**CI**: Coronary Insufficiency. Outside Framingham a forgotten diagnosis. Seems equivalent to unstable angina pectoris and epimural infarctions. Is classified as an acute event by FHS.

**CVA**: Cerebrovascular accidents. The equivalent of CHD in the brain (although there are several types of stroke: thrombotic blocking due to a thrombus or an embolism (see further), or acute bleeding through rupture. Leads to acute brain infarctions (ABI, atherothrombotic BI), paralysis or death. To avoid misunderstanding with cardiovascular disease, I consistently use ‘stroke’. It is one of the main causes of disability in human populations. One or more strokes may also cause dementia.

**CVD**: All cardiovascular diseases. Include in this dataset CHD, Congestive Heart Failure (CHF), Stroke (in the dataset called CVA, cerebrovascular accidents), PAD (Peripheral Arterial Disease, in the dataset called Intermittent Claudication, IC).

**HA**: Heart Attacks. Includes MI, CI, death from CHD. Unclear what happens to sudden cardiac death which is not CHD.

**Ischaemia**: Lack of sufficient oxygen in the (muscle) tissues. Causes pain and/or ECG changes.

**PAD**: Peripheral arterial disease. The equivalent of CHD and Stroke in the legs (or potentially all body-parts). Called **IC** in FHS: intermittent claudication, which is the equivalent of angina pectoris: ischaemic pains by use of the muscles. This is a relatively rarer cause of death. If it causes death, it is through rupture of the aorta (the main body artery). Blood clots, formed on the damaged vessel walls, may travel through the arteries and cause acute blockings in the brain (causing a stroke)

**Stroke**: See CVA. Acute blocking of brain vessels, with often irreversible disabling consequences.

**TIA**: Transient Ischaemic Attack. Equivalent of unstable angina pectoris in the brain. ‘Small’ stroke, leading to temporary (< 24 hours) but not definitive disability.
Technical Appendix 3.2

Multistate model of cardiovascular disease

A. NO-CVD (i.e. without other states in model) to OCVD (to) OCVA (to) CVM (hard stroke; ABI, embolism, haemorrhage and other CVA) to death. OCVD is first CVD events other than CVA. OCVA is first CVA events other than CVM (i.e. TIA)

B. NO-CVD (i.e. without other states in model) to OCVD (to) CHD (to) CVM to death (could replace CHD with MI for direct effects of changes in MI)
C. NO-CVD (i.e. without other states in model) to OCVD (to) CHD (to) CHF to death (could replace CHD with MI for direct effects of changes in MI)

D. NO-CVD (i.e. without other states in model) to OCVD (to) CVA (to) CHF to death (could replace CVA with CVM for direct effects of changes in stroke)
E. NO-CVD (i.e. without other states in model) to OCVD (to) MI (to) CVM (to) CHF to death

F. NO-CVD (i.e. without other states in model) to OCVD (to) MI (to) CVM (to) CHF to death, incorporating back-flow
Technical Appendix 3.3

SPSS-syntax: estimation of occurrence-exposure

Get file =C:\a\b.sav. /* open the individual level data file with list of following variables

Variable labels

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pid</td>
<td>'Cohort general exam data random ID'</td>
</tr>
<tr>
<td>Ybirth</td>
<td>'Exact year of birth relative to 1900'</td>
</tr>
<tr>
<td>Ycvd</td>
<td>'Exact year of CVD relative to 1900'</td>
</tr>
<tr>
<td>Ydth</td>
<td>'Exact year of death relative to 1900'</td>
</tr>
<tr>
<td>Sex</td>
<td>'1=Male &amp; 2=Female'</td>
</tr>
<tr>
<td>Lost</td>
<td>'Lost of follow-up'</td>
</tr>
<tr>
<td>Age1</td>
<td>'Completed age in years at exam1'</td>
</tr>
<tr>
<td>Eage</td>
<td>'Exact age in years at exam1'</td>
</tr>
<tr>
<td>Endstudy</td>
<td>'Exact year of ending study time relative to 1900'</td>
</tr>
</tbody>
</table>

*Event counts: occurrences

Compute Hcvd=trunc(ycvd-ybirth). /* Age at transition from no-cvd to cvd
If (missing(Ycvd)) Hd=trunc(ydth-ybirth)./* age at transition from no-cvd to death
If (Ycvd>0) Cvdd=trunc(ydth-ybirth). /* age at transition from cvd to death

*Age-specific occurrences in tabular form

Tables

| Format zeros missing(\') /Tables(labels) |
| BY sex>( Hcvd + Hd + Cvdd) |
| /Statistics count ((f5.0) 'Count') . /* Number of occurrences is counted separately for males and females

*Exposure time: no-cvd to cvd or death

If (ycvd>0) Agecvdfr=(ycvd-ybirth). /* Exact age of transition from h to cvd
If (missing(ycvd)) Agecvdfr=min((ydth-ybirth),(endstudy-ybirth)). /* Exact time to death or censored (without experiencing CVD)
Compute Agecvd=trunc(averagecovdf). /* Transition from h to cvd at completed age
If (age1=agecvd) Tcvdage1=agecvdfr-age1. /* Fraction of time contributed to exam1
If (not(age1=agecvd)) Tcvdage1=age1+1-agecovdf. /* Fraction of time contributed to exam1
If (age1=agecvd) Tcvdagef=agecvdfr-age1. /* Fraction of time contributed to last age
If (not(age1=agecvd)) Tcvdagef=agecvdfr-agecvd. /* Fraction of time contributed to last age
If (age1=28) Tcvd0=tcvdage1.
Vector Tcvd(74). /* Total number of age groups by one-year age band

Loop Cvdring=1 to 74.
  If (age1=(cvdring+28)) Tcvd(cvdring)=tcvdage1.
  If (age1<(cvdring+28) and agecvd>(cvdring+28)) Tcvd (cvdring)=1.
  If (agecvd=(cvdring+28)) Tcvd (cvdring)=tcvdagef.
End loop.

* Aggregate outfile for h-cvd or h-d transition

Aggregate outfile='C:\FHS_2002_24exam\chap3_h_cvd_d.sav' /break=sex /sumt28 to sumt102=sum(tcvd0 to tcvd74).
*Aggregate outfile command combines groups of cases into single summary cases and creates a new aggregated data file. Cases are aggregated based on the value of one or more grouping variables. Cases are grouped together based on the values of the break variables.

**Exposure time cvd to death**

If (ycvd>0) Cvddf1=min((ydth-ybirth), (endstudy-ybirth)). /* Exact age of transition from Cvd to death or censored

Compute Cvdd=trunc(cvddf1). /*Cvd to death or censored at completed age

If (agecvd=cvdd) Tcvdd1=cvddf1-agecvdfr. /* Fraction of time contributed to agecvd

If (not(agecvd=cvdd)) Tcvdd1=agecvd+1-agecvdfr. /* Fraction of time contributed to agecvd

If (agecvd=cvdd) Tcvddf=1-cvddf1. /*Fraction of time contributed to last age

If (not(agecvd=cvdd)) Tcvddf=cvddf1-cvdd. /*Fraction of time contributed to last age

If (agecvd=28) Tcvdd0=tcvdd1.

Vector Ttcvdd(74). /* total number of age groups by one year age band

Loop cddring=1 to 74.
    If (agecvd=(cddring+28)) Ttcvdd(cddring)=tcvdd1.
    If (agecvd<(cddring+28) and cvdd>(cddring+28)) Ttcvdd(cddring)=1.
    If (cvdd=(cddring+28)) Ttcvdd(cddring)=tcvddf.
End Loop.

Aggregate outlife='C:\FHS_2002_24exam\chap3_cvd_d.sav'/break=sex /sumt28 to sumt102 =sum(ttcvdd0 to ttcvdd74).

Get file ='C:\FHS_2002_24exam\chap3_cvd_d.sav'.

Transfer the occurrences (from SPSS output) and exposures (getting file 'C:\a\chap3_h_cvd_d.sav' and 'C:\a\chap3_cvd_d.sav') tables

into (say) Excel work sheet. The occurrence and exposure rates are estimated for different age intervals with the number of people who experienced an event in a given age interval as the numerator and the total person-years (i.e. risk period) of observations of at-risk participants in that interval as the denominator.
### Technical Appendix 3.4

Occurrences, exposures and occurrence-exposure rates matrices: 3-state cardiovascular disease model (males and females combined)

<table>
<thead>
<tr>
<th>Age</th>
<th>No-CVD to CVD</th>
<th>No-CVD to D</th>
<th>CVD to D</th>
<th>No-CVD to CVD</th>
<th>No-CVD to D</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>5 2 1</td>
<td>1904.34</td>
<td>1904.34</td>
<td>11.81</td>
<td>0.003</td>
</tr>
<tr>
<td>41</td>
<td>6 5 0</td>
<td>2062.18</td>
<td>2062.18</td>
<td>16.60</td>
<td>0.000</td>
</tr>
<tr>
<td>42</td>
<td>5 2 2</td>
<td>2240.53</td>
<td>2240.53</td>
<td>20.59</td>
<td>0.000</td>
</tr>
<tr>
<td>43</td>
<td>11 2 3</td>
<td>2414.95</td>
<td>2414.95</td>
<td>26.09</td>
<td>0.000</td>
</tr>
<tr>
<td>44</td>
<td>9 3 3</td>
<td>2585.30</td>
<td>2585.30</td>
<td>35.15</td>
<td>0.000</td>
</tr>
<tr>
<td>45</td>
<td>18 3 2</td>
<td>2752.77</td>
<td>2752.77</td>
<td>44.10</td>
<td>0.000</td>
</tr>
<tr>
<td>46</td>
<td>11 6 2</td>
<td>2896.17</td>
<td>2896.17</td>
<td>57.34</td>
<td>0.000</td>
</tr>
<tr>
<td>47</td>
<td>21 5 7</td>
<td>3023.57</td>
<td>3023.57</td>
<td>70.39</td>
<td>0.000</td>
</tr>
<tr>
<td>48</td>
<td>15 4 5</td>
<td>3148.42</td>
<td>3148.42</td>
<td>81.15</td>
<td>0.000</td>
</tr>
<tr>
<td>49</td>
<td>24 16 5</td>
<td>3269.89</td>
<td>3269.89</td>
<td>94.00</td>
<td>0.000</td>
</tr>
<tr>
<td>50</td>
<td>29 10 6</td>
<td>3384.12</td>
<td>3384.12</td>
<td>115.27</td>
<td>0.000</td>
</tr>
<tr>
<td>51</td>
<td>27 10 7</td>
<td>3507.58</td>
<td>3507.58</td>
<td>137.88</td>
<td>0.000</td>
</tr>
<tr>
<td>52</td>
<td>41 24 9</td>
<td>3607.63</td>
<td>3607.63</td>
<td>167.00</td>
<td>0.000</td>
</tr>
<tr>
<td>53</td>
<td>60 16 16</td>
<td>3756.65</td>
<td>3756.65</td>
<td>237.69</td>
<td>0.000</td>
</tr>
<tr>
<td>54</td>
<td>61 20 20</td>
<td>3823.10</td>
<td>3823.10</td>
<td>275.36</td>
<td>0.000</td>
</tr>
<tr>
<td>55</td>
<td>56 15 24</td>
<td>3876.44</td>
<td>3876.44</td>
<td>313.90</td>
<td>0.000</td>
</tr>
<tr>
<td>56</td>
<td>67 19 17</td>
<td>3933.52</td>
<td>3933.52</td>
<td>352.96</td>
<td>0.000</td>
</tr>
<tr>
<td>57</td>
<td>78 16 24</td>
<td>3961.21</td>
<td>3961.21</td>
<td>401.97</td>
<td>0.000</td>
</tr>
<tr>
<td>58</td>
<td>79 19 24</td>
<td>3975.94</td>
<td>3975.94</td>
<td>456.13</td>
<td>0.000</td>
</tr>
<tr>
<td>59</td>
<td>73 24 29</td>
<td>3967.70</td>
<td>3967.70</td>
<td>505.61</td>
<td>0.000</td>
</tr>
<tr>
<td>60</td>
<td>63 23 29</td>
<td>3924.70</td>
<td>3924.70</td>
<td>549.61</td>
<td>0.000</td>
</tr>
<tr>
<td>61</td>
<td>62 30 40</td>
<td>3827.63</td>
<td>3827.63</td>
<td>609.89</td>
<td>0.000</td>
</tr>
<tr>
<td>62</td>
<td>89 19 37</td>
<td>3714.39</td>
<td>3714.39</td>
<td>660.29</td>
<td>0.000</td>
</tr>
<tr>
<td>63</td>
<td>95 21 40</td>
<td>3600.49</td>
<td>3600.49</td>
<td>722.31</td>
<td>0.000</td>
</tr>
<tr>
<td>64</td>
<td>84 32 44</td>
<td>3490.61</td>
<td>3490.61</td>
<td>758.35</td>
<td>0.000</td>
</tr>
<tr>
<td>65</td>
<td>87 25 49</td>
<td>3377.89</td>
<td>3377.89</td>
<td>796.96</td>
<td>0.000</td>
</tr>
<tr>
<td>66</td>
<td>74 21 46</td>
<td>3271.30</td>
<td>3271.30</td>
<td>834.78</td>
<td>0.000</td>
</tr>
<tr>
<td>67</td>
<td>82 33 56</td>
<td>3159.93</td>
<td>3159.93</td>
<td>863.11</td>
<td>0.000</td>
</tr>
<tr>
<td>68</td>
<td>90 29 54</td>
<td>3044.13</td>
<td>3044.13</td>
<td>897.30</td>
<td>0.000</td>
</tr>
<tr>
<td>69</td>
<td>71 39 55</td>
<td>2923.82</td>
<td>2923.82</td>
<td>922.57</td>
<td>0.000</td>
</tr>
<tr>
<td>70</td>
<td>97 36 66</td>
<td>2810.99</td>
<td>2810.99</td>
<td>954.57</td>
<td>0.000</td>
</tr>
<tr>
<td>71</td>
<td>87 37 73</td>
<td>2686.39</td>
<td>2686.39</td>
<td>961.63</td>
<td>0.000</td>
</tr>
<tr>
<td>72</td>
<td>76 34 96</td>
<td>2564.41</td>
<td>2564.41</td>
<td>960.79</td>
<td>0.000</td>
</tr>
<tr>
<td>73</td>
<td>84 38 82</td>
<td>2445.91</td>
<td>2445.91</td>
<td>952.18</td>
<td>0.000</td>
</tr>
<tr>
<td>74</td>
<td>75 44 70</td>
<td>2327.36</td>
<td>2327.36</td>
<td>952.09</td>
<td>0.000</td>
</tr>
<tr>
<td>75</td>
<td>85 52 80</td>
<td>2185.21</td>
<td>2185.21</td>
<td>956.33</td>
<td>0.000</td>
</tr>
<tr>
<td>76</td>
<td>82 49 92</td>
<td>2018.25</td>
<td>2018.25</td>
<td>938.46</td>
<td>0.000</td>
</tr>
<tr>
<td>77</td>
<td>78 54 100</td>
<td>1836.93</td>
<td>1836.93</td>
<td>891.92</td>
<td>0.000</td>
</tr>
<tr>
<td>78</td>
<td>66 41 106</td>
<td>1642.15</td>
<td>1642.15</td>
<td>822.16</td>
<td>0.000</td>
</tr>
<tr>
<td>79</td>
<td>68 53 83</td>
<td>1461.65</td>
<td>1461.65</td>
<td>753.08</td>
<td>0.000</td>
</tr>
<tr>
<td>80</td>
<td>59 47 71</td>
<td>1282.27</td>
<td>1282.27</td>
<td>704.43</td>
<td>0.000</td>
</tr>
<tr>
<td>81</td>
<td>59 52 104</td>
<td>1128.03</td>
<td>1128.03</td>
<td>622.38</td>
<td>0.000</td>
</tr>
<tr>
<td>82</td>
<td>54 43 85</td>
<td>972.59</td>
<td>972.59</td>
<td>562.80</td>
<td>0.000</td>
</tr>
<tr>
<td>83</td>
<td>51 46 73</td>
<td>830.53</td>
<td>830.53</td>
<td>498.11</td>
<td>0.000</td>
</tr>
<tr>
<td>84</td>
<td>40 42 67</td>
<td>695.43</td>
<td>695.43</td>
<td>445.52</td>
<td>0.000</td>
</tr>
<tr>
<td>85</td>
<td>42 29 72</td>
<td>583.73</td>
<td>583.73</td>
<td>403.22</td>
<td>0.000</td>
</tr>
<tr>
<td>86</td>
<td>22 40 57</td>
<td>482.47</td>
<td>482.47</td>
<td>343.36</td>
<td>0.000</td>
</tr>
<tr>
<td>87</td>
<td>22 36 53</td>
<td>392.95</td>
<td>392.95</td>
<td>289.60</td>
<td>0.000</td>
</tr>
<tr>
<td>88</td>
<td>24 26 58</td>
<td>319.50</td>
<td>319.50</td>
<td>234.49</td>
<td>0.000</td>
</tr>
<tr>
<td>89</td>
<td>18 23 42</td>
<td>249.29</td>
<td>249.29</td>
<td>187.72</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Occurrence-exposure rates:

- No-CVD to CVD: 0.003
- No-CVD to D: 0.001
- CVD to D: 0.085
Multistate life table of 3-state cardiovascular disease: males and females combined

<table>
<thead>
<tr>
<th>Age</th>
<th>Transition probability</th>
<th>Survival probability</th>
<th>Person years lived</th>
<th>Total person years lived</th>
<th>Life expectancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO-CVD</td>
<td>CVD to NO-CVD</td>
<td>CVD to death</td>
<td>CVD to Death</td>
<td>NO-CVD Death</td>
</tr>
<tr>
<td>40</td>
<td>0.003</td>
<td>0.001</td>
<td>0.081</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>41</td>
<td>0.003</td>
<td>0.002</td>
<td>0.000</td>
<td>0.996</td>
<td>0.003</td>
</tr>
<tr>
<td>42</td>
<td>0.002</td>
<td>0.001</td>
<td>0.093</td>
<td>0.991</td>
<td>0.006</td>
</tr>
<tr>
<td>43</td>
<td>0.005</td>
<td>0.001</td>
<td>0.109</td>
<td>0.988</td>
<td>0.007</td>
</tr>
<tr>
<td>44</td>
<td>0.003</td>
<td>0.001</td>
<td>0.082</td>
<td>0.983</td>
<td>0.011</td>
</tr>
<tr>
<td>45</td>
<td>0.007</td>
<td>0.001</td>
<td>0.044</td>
<td>0.978</td>
<td>0.013</td>
</tr>
<tr>
<td>46</td>
<td>0.004</td>
<td>0.002</td>
<td>0.034</td>
<td>0.971</td>
<td>0.019</td>
</tr>
<tr>
<td>47</td>
<td>0.007</td>
<td>0.002</td>
<td>0.095</td>
<td>0.965</td>
<td>0.022</td>
</tr>
<tr>
<td>48</td>
<td>0.005</td>
<td>0.001</td>
<td>0.060</td>
<td>0.957</td>
<td>0.027</td>
</tr>
<tr>
<td>49</td>
<td>0.007</td>
<td>0.005</td>
<td>0.052</td>
<td>0.951</td>
<td>0.030</td>
</tr>
<tr>
<td>50</td>
<td>0.009</td>
<td>0.003</td>
<td>0.051</td>
<td>0.939</td>
<td>0.035</td>
</tr>
<tr>
<td>51</td>
<td>0.008</td>
<td>0.003</td>
<td>0.050</td>
<td>0.929</td>
<td>0.041</td>
</tr>
<tr>
<td>52</td>
<td>0.011</td>
<td>0.007</td>
<td>0.052</td>
<td>0.919</td>
<td>0.046</td>
</tr>
<tr>
<td>53</td>
<td>0.016</td>
<td>0.004</td>
<td>0.076</td>
<td>0.902</td>
<td>0.054</td>
</tr>
<tr>
<td>54</td>
<td>0.012</td>
<td>0.004</td>
<td>0.057</td>
<td>0.884</td>
<td>0.065</td>
</tr>
<tr>
<td>55</td>
<td>0.016</td>
<td>0.005</td>
<td>0.070</td>
<td>0.870</td>
<td>0.072</td>
</tr>
<tr>
<td>56</td>
<td>0.014</td>
<td>0.004</td>
<td>0.074</td>
<td>0.852</td>
<td>0.080</td>
</tr>
<tr>
<td>57</td>
<td>0.017</td>
<td>0.005</td>
<td>0.047</td>
<td>0.836</td>
<td>0.087</td>
</tr>
<tr>
<td>58</td>
<td>0.019</td>
<td>0.004</td>
<td>0.058</td>
<td>0.818</td>
<td>0.097</td>
</tr>
<tr>
<td>59</td>
<td>0.019</td>
<td>0.005</td>
<td>0.051</td>
<td>0.799</td>
<td>0.107</td>
</tr>
<tr>
<td>60</td>
<td>0.018</td>
<td>0.006</td>
<td>0.056</td>
<td>0.780</td>
<td>0.117</td>
</tr>
<tr>
<td>61</td>
<td>0.021</td>
<td>0.006</td>
<td>0.051</td>
<td>0.761</td>
<td>0.124</td>
</tr>
<tr>
<td>62</td>
<td>0.024</td>
<td>0.008</td>
<td>0.063</td>
<td>0.741</td>
<td>0.134</td>
</tr>
<tr>
<td>63</td>
<td>0.024</td>
<td>0.005</td>
<td>0.054</td>
<td>0.717</td>
<td>0.143</td>
</tr>
<tr>
<td>64</td>
<td>0.026</td>
<td>0.006</td>
<td>0.054</td>
<td>0.697</td>
<td>0.152</td>
</tr>
<tr>
<td>65</td>
<td>0.024</td>
<td>0.009</td>
<td>0.056</td>
<td>0.675</td>
<td>0.162</td>
</tr>
<tr>
<td>66</td>
<td>0.025</td>
<td>0.007</td>
<td>0.060</td>
<td>0.653</td>
<td>0.169</td>
</tr>
<tr>
<td>67</td>
<td>0.022</td>
<td>0.006</td>
<td>0.054</td>
<td>0.631</td>
<td>0.175</td>
</tr>
<tr>
<td>68</td>
<td>0.025</td>
<td>0.010</td>
<td>0.063</td>
<td>0.613</td>
<td>0.180</td>
</tr>
<tr>
<td>69</td>
<td>0.029</td>
<td>0.009</td>
<td>0.058</td>
<td>0.591</td>
<td>0.184</td>
</tr>
<tr>
<td>70</td>
<td>0.024</td>
<td>0.013</td>
<td>0.058</td>
<td>0.569</td>
<td>0.191</td>
</tr>
<tr>
<td>71</td>
<td>0.034</td>
<td>0.013</td>
<td>0.067</td>
<td>0.548</td>
<td>0.193</td>
</tr>
<tr>
<td>72</td>
<td>0.032</td>
<td>0.013</td>
<td>0.073</td>
<td>0.522</td>
<td>0.199</td>
</tr>
<tr>
<td>73</td>
<td>0.029</td>
<td>0.013</td>
<td>0.095</td>
<td>0.499</td>
<td>0.201</td>
</tr>
<tr>
<td>74</td>
<td>0.034</td>
<td>0.015</td>
<td>0.083</td>
<td>0.478</td>
<td>0.196</td>
</tr>
<tr>
<td>75</td>
<td>0.031</td>
<td>0.018</td>
<td>0.071</td>
<td>0.455</td>
<td>0.196</td>
</tr>
<tr>
<td>76</td>
<td>0.038</td>
<td>0.023</td>
<td>0.080</td>
<td>0.432</td>
<td>0.196</td>
</tr>
<tr>
<td>77</td>
<td>0.039</td>
<td>0.024</td>
<td>0.093</td>
<td>0.406</td>
<td>0.197</td>
</tr>
<tr>
<td>78</td>
<td>0.041</td>
<td>0.028</td>
<td>0.106</td>
<td>0.380</td>
<td>0.194</td>
</tr>
<tr>
<td>79</td>
<td>0.039</td>
<td>0.024</td>
<td>0.121</td>
<td>0.354</td>
<td>0.189</td>
</tr>
<tr>
<td>80</td>
<td>0.045</td>
<td>0.035</td>
<td>0.104</td>
<td>0.331</td>
<td>0.180</td>
</tr>
<tr>
<td>81</td>
<td>0.045</td>
<td>0.035</td>
<td>0.106</td>
<td>0.305</td>
<td>0.176</td>
</tr>
<tr>
<td>82</td>
<td>0.050</td>
<td>0.044</td>
<td>0.154</td>
<td>0.281</td>
<td>0.173</td>
</tr>
<tr>
<td>83</td>
<td>0.053</td>
<td>0.042</td>
<td>0.140</td>
<td>0.255</td>
<td>0.160</td>
</tr>
<tr>
<td>84</td>
<td>0.058</td>
<td>0.052</td>
<td>0.136</td>
<td>0.230</td>
<td>0.151</td>
</tr>
<tr>
<td>85</td>
<td>0.054</td>
<td>0.057</td>
<td>0.140</td>
<td>0.205</td>
<td>0.144</td>
</tr>
<tr>
<td>86</td>
<td>0.068</td>
<td>0.047</td>
<td>0.164</td>
<td>0.182</td>
<td>0.135</td>
</tr>
<tr>
<td>87</td>
<td>0.043</td>
<td>0.078</td>
<td>0.153</td>
<td>0.161</td>
<td>0.125</td>
</tr>
<tr>
<td>88</td>
<td>0.052</td>
<td>0.085</td>
<td>0.167</td>
<td>0.142</td>
<td>0.113</td>
</tr>
<tr>
<td>89</td>
<td>0.070</td>
<td>0.075</td>
<td>0.219</td>
<td>0.122</td>
<td>0.101</td>
</tr>
<tr>
<td>90</td>
<td>0.067</td>
<td>0.085</td>
<td>0.200</td>
<td>0.105</td>
<td>0.088</td>
</tr>
</tbody>
</table>
Handling missing data in risk factors in repeated measurements

Abstract

Public health researchers involved in data analysis are invariably confronted with non-response and resultant missing values. This missing information is a recurring problem in the statistical literature. In longitudinal and in panel studies, the problem of missing data is very common. To construct the risk career (e.g. smoking career) or to elicit the impact of the risk career on life expectancy free of cardiovascular disease and life expectancy with cardiovascular disease in a population, imputation of these missing values is essential. Without imputation, we may not be able to capture all events that we would like to relate to the risk factors at a nearby point in time. This chapter aims to propose a potential method to impute missing values of risk factors in repeated measurement studies. We have considered two risk factors—smoking status and systolic blood pressure (BP). Smoking status was selected to illustrate the application of the imputation method to categorical risk factors. BP was selected to illustrate the imputation method for continuous risk factors. We used the original Framingham Heart Study cohort to illustrate the proposed method. Over 48 years of follow-up, the overall percentage of respondents from whom no observations were missing in the Framingham Heart Study was 41.5. To impute the missing values on smoking, we followed the history of smoking of the relevant individual and replaced the missing values by the smoking status closest in time. The method we propose for imputing blood pressure values is based on a regression model, which is fitted to individual data and imputes mean values for each individual from each individual observation. The algorithm of the proposed method is easy to manipulate and to understand. We have validated our method and found that the proposed method is justifiable for long-period longitudinal data. Applying this method to impute missing values of other risk factors is possible. The imputation may result in better estimates of the required parameters than when the missing cases are omitted.
4.1 Introduction

Incomplete or missing data is a recurring problem in data analysis. In particular, public health researchers involved in data analysis are invariably confronted with non-response and resultant missing values (Zhou et al., 2001). Missing information may exist in any empirical study. In longitudinal or panel studies, missing data are very common (Molenberghs, 2001). The main reason is that in longitudinal studies, subjects may drop out early or be unavailable during one or more data collection episodes. During follow up, respondents may fail to return in longitudinal studies because of death, illness or disability. Alternatively, subjects may be unwilling or may fail to complete a section of the questionnaire due to lack of time or interest. If such attrition is random with respect to the dependant variables, it can be ignored from the analysis. Otherwise, the probability of dropout depends upon covariates and/or previously observed response, or on current unobserved responses (Little, 1995). However, though the causes and problems with missing values in longitudinal research have been identified (Little, 1995; Zhou et al., 2001; Molenberghs, 2001; West and Dawson, 2002; Wei and Shih, 2001) the methods available for imputation of missing values are quite limited.

Repeated measurements are quite common in clinical trials. However, the number of longitudinal studies is increasing in public health research to measure the health status of the elderly and the life course of non-communicable diseases (e.g. cardiovascular disease). In developed countries, there are many longitudinal studies, where multiple waves of measurements (panel) are available. The multiple waves of measurement bring multiple opportunities for missing data. Dropout and intermittent non-response are found in most longitudinal studies (Little, 1995). Because of the missing information, we may not be able to analyze time-dependent covariates (e.g. risk factors). Most researchers use a part of the follow-up data or only the baseline information. For example, to calculate the ‘cardiovascular disease risk profile’, Anderson et al., (1991) used the FHS follow-up period from 1968 to 1975 with at least one observation of the listed risk factors. Stamler et al., (1999) used baseline information to measure the impact of several risk factors on the incidence of cardiovascular disease. Although the disease incidence and mortality is recorded in exact time, during the follow-up the likelihood of missing risk factors is quite common. Therefore, analysis of the risk career of a disease process or mortality is difficult without imputation of missing values.

To construct the risk career (e.g. smoking career) or to see the impact of the risk career on life expectancy free of cardiovascular disease and life expectancy with cardiovascular disease, imputation of missing values can be useful. Without imputation, we may not be able to capture all events that we would like to relate with the risk factors at a nearby point in time. Hence, the objective of this chapter
is to describe a potential method to impute missing values of risk factors in a repeated measurement study. To this end, we considered two risk factors—*smoking status* and *systolic blood pressure* (BP). Smoking status was selected to illustrate the imputation method for categorical risk factors. BP was selected to demonstrate the imputation method for continuous risk factors. We used the original Framingham Heart Study (FHS) cohort to apply the proposed method. In this study, the same cohort of 5,209 subjects has been followed for 48 consecutive years. Several risk factors of cardiovascular disease, cardiovascular disease incidence and mortality were recorded during the follow-up period. After 48 years of follow up, 77 percent of the respondents who had participated in the first exam during 1948-51 had died. In total 15,800 (22 percent) exams or responses are missing from the potential 87,578 responses.

In the FHS, all events (cardiovascular disease and death) are recorded at the exact time. Three types of missing information could be identified. First, there were the respondents who did not participate in a particular exam (*missing type I*). The second type of missing data concerned risk factor information that failed to be recorded for some of the respondents even though the respondents were present in that exam (*missing type II*). The third type referred to the exams where risk factor information had been omitted completely for all subjects although the respondents had been present at the exam in question (*missing type III*). For example, smoking status failed to be recorded at exam 6.

A common method for handling missing values is imputation. There are no universally applicable methods for handling missing data (Shih, 2002). Several methods of imputation are available, such as the hot deck approach, the regression method, multiple imputation and so on. Each method has its own advantages and disadvantages. The selection of the appropriate method depends on the assumptions made about the missing data mechanism. Many researchers use these methods for different data sets. Imputation methods for repeated measurement data are limited. This chapter describes the missing information in the FHS and subsequently illustrates the method proposed by us for imputing missing values.

We propose two methods for imputation: one for discrete variables and another one for continuous variables. In the case of a discrete variable, the history of the risk factor in question is followed after which the missing values are replaced by the risk factor status that is close in time. Continuous variables are imputed by fitting a regression model to each individual, which model is used to impute the missing value. That is, missing values are imputed based on the individually observed values. For an application, see Chapter 5.

Section 4.2 of this chapter describes the missing data mechanism. Section 4.3 provides an overview of missing data in the FHS. In this section, we have illustrated missing information on two risk factors: smoking status and systolic blood pressure. Section 4.4 describes the imputation methods that are proposed and applied to the
Framingham Heart Study. In subsection 4.4.1, we describe the imputation method applied for imputation of smoking status. The method we use for the imputation of BP is discussed in subsection 4.4.2. In Section 4.5, we validate our proposed methods. Section 4.6 concludes the chapter.

4.2 The missing data mechanism

The missing data mechanism is a concept often discussed when missing data occur. To understand the consequences of missing values and potential solutions for statistical analysis with missing values, some idea of why and how missing values occur is needed. The missing data mechanism is described in two parts: (i) substantive issues and (ii) problems. In the first part, we describe the types of missing data mechanisms. The second part deals with associated problems with missing data.

4.2.1 Substantive issues

In modern statistical exercise, the occurrence of missing values is usually viewed as a random phenomenon (Schafer, 2001). Little and Rubin (1987) define the occurrence of missing values by three unique types of missing data mechanisms. Missing values of a random variable $Y$ can be (i) Missing completely at random (MCAR), (ii) Missing at random (MAR) and (iii) Nonignorable. We assume that the $y_1$ portion of $Y$ (a realization of $Y$) is missing and the rest $y_2$ ($y_1 \cdot y_2$) is observed.

(i) Missing completely at random (MCAR)

Missing data is considered to be MCAR when, given two variables, say, $X$ and $Y$, the probability of response is independent of variables $X$ and $Y$. That is, cases with complete data are indistinguishable from cases with incomplete data. MCAR is the process in which the probability of data being missing is independent of both observed measurements (e.g. baseline covariates, observed responses) and unobserved measurements (those that would have been observed if the respondent had remained in the study). Under an MCAR process, whether or not a variable is observed does not affect its distribution (Little et al., 2000). This can be written as-

$$Pr(Y=y \mid y_1 \text{ missing}, y_2 \text{ observed}) = Pr(Y=y \mid y_2 \text{ observed}), \text{ where } y_1 \text{ and } y_2 \in Y$$

As an example, suppose serum cholesterol level (SCL) and age are variables of interest for our study. If the likelihood that a respondent would provide his or her serum cholesterol level is the same for all individuals regardless of their SCL or age, then the missing data is considered to be MCAR. Under MCAR, the observed
Handing Missing Data in Risk Factors in Repeated Measurements

Responses form a random sub-sample of the sampled responses (Rubin, 1976). Therefore, when data are MCAR there is no impact on bias and most standard approaches of analysis are valid.

(ii) Missing at random (MAR)
Missing data is considered to be MAR when, given two variables, $X$ and $Y$, the probability of response depends on $X$ but not on $Y$. That is, the cases with incomplete data differ from cases with complete data, but the pattern of data missingness is traceable or predictable from other variables in the database rather than being due to the specific variable on which the data is missing. Under MAR, the probability of missing depends on the observed data rather than on the unobserved data. The observed responses are a random sub-sample of the sampled values within a subclass defined by the observed data. Therefore, MAR is a more relaxed condition, assuming only that missing and observed distributions of $Y$ are identical, conditional on predictor $X$, i.e.,

$$Pr(Y=y|y_1\text{ missing }, y_2\text{ observed, } x\text{ observed}) = Pr(Y=y|y_2\text{ observed, } x\text{ observed})$$

In this situation, the missing data mechanism depends only on the covariates $X$ and is classified as covariate-dependent missing (William, 2000). For example, again using the example of serum cholesterol level and age, if the likelihood that an individual would provide his or her SCL varied according to an individual’s age, the missing data is considered to be MAR. Most of the missing data methods are designed under this assumption.

(iii) Nonignorable
When, given two variables, $X$ and $Y$, the probability of response depends on $X$ and possibly $Y$, missing data is considered to be nonignorable i.e.

$$Pr(Y=y|y_1\text{ missing, } y_2\text{ observed, } x\text{ observed}) = Pr(Y=y|y_1\text{ observed, } y_2\text{ missing, } x\text{ observed})$$

In other words, if the missing mechanism is neither MCAR nor MAR, then it is nonignorable. Under nonignorable, the pattern of data missingness is non-random and is not predictable from other variables in the database. An example of this would be if the likelihood of a respondent providing his or her SCL varied according to a person’s SCL (observed and missing) and age (covariate).

In repeated measurement data, there are ultimately two types of missing information: bounded missing and missing due to dropout. Bounded missing is defined as a missing value that has at least one observed value before, and at least one observed value after the period in which it is missing (SOLAS, 1999). Missing
due to dropout refers to the case where a participant is dropped from the follow-up and never comes back in that survey. The following table shows an example of bounded missing data and data missing due to dropout. The variables Exam1 to Exam5 are a set of longitudinal measures of SCL for 5 respondents. The SCL of the first respondent is missing at exam 2 and exam 3, which is bounded missing. Similarly, the second respondent’s SCL is bounded missing at exam 3. Respondent 4 has dropped out from exam 3 onward, which is missing due to dropout.

Table 4.1 An example of repeated measurement data (a hypothetical example of cholesterol level)

<table>
<thead>
<tr>
<th>Respondent Identification Number</th>
<th>Exam1</th>
<th>Exam2</th>
<th>Exam3</th>
<th>Exam4</th>
<th>Exam5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>212</td>
<td>*</td>
<td>*</td>
<td>223</td>
<td>240</td>
</tr>
<tr>
<td>2</td>
<td>190</td>
<td>185</td>
<td>*</td>
<td>232</td>
<td>215</td>
</tr>
<tr>
<td>3</td>
<td>240</td>
<td>212</td>
<td>227</td>
<td>218</td>
<td>235</td>
</tr>
<tr>
<td>4</td>
<td>222</td>
<td>231</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>185</td>
<td>198</td>
<td>175</td>
<td>192</td>
<td>200</td>
</tr>
</tbody>
</table>

*"* bounded missing

'-' missing due to dropout

4.2.2 Problems

The issue of missing data is the subject of increasing debate in contemporary statistics. Hence missing data is a problem in almost all areas of empirical research (Rubin, 1996). In any given study, missing data can have many causes. For instance, respondents may be unwilling to answer some questions (which is called item non-response) or may refuse to participate in a study (which is called unit non-responses) and so on. In a longitudinal study, there is always a higher chance of missing observations during follow-up than in a cross-sectional study. Thus, many longitudinal studies suffer from attrition; that is, from subjects dropping out prematurely (Laird, 1988; Hedeker and Gibbons, 1997; William, 2000).

Often, the question arises of why the cases with missing values should not simply be deleted rather than imputing values at all. Little and Rubin (1987), among other researchers, have demonstrated the dangers of simply deleting cases. Basically, case deletion strategies assume that the deleted cases are a relatively small proportion of the entire data set and are representative of it—that is, cases will then be missing completely at random. In most research settings, however, missing data are indicative of some pattern and cannot safely be assumed to be at random. In such circumstances, deletion can introduce substantial bias into the study (Laird, 1988). Moreover, the loss in sample size can appreciably diminish the statistical
power of the analysis. Missing values lead to less efficient estimates because of the reduced size of the database, while standard complete-data methods of analysis no longer apply. When cases are deleted from the data because of one or more variables with missing values, the number of remaining cases may be small even if the proportion of missing data is small for each variable.

It is obvious that if some observations are incomplete or missing, there is more uncertainty in inference than if the data had been complete. The main reasons are the smaller sample size, and the bias introduced when the pattern of missing data results from some process. Hence, analysis of data sets with incomplete information is more problematic than analysis of complete data sets. Therefore, in the presence of non-response that cannot plausibly be considered to be completely random, estimates of population parameters are subject to potential bias (U.S. Census Bureau, 2002; Shih, 2002).

4.3 Missing data in the FHS

We have used the original FHS cohort to illustrate the proposed method. The FHS consisted of 5,209 respondents (45% male) from a sample of adults aged 28 to 62 years residing in Framingham, Massachusetts between 1948 and 1951. The FHS is a long time longitudinal study. We have used the first 48 years of follow-up, from exam 1 in 1948 to exam 24 in 1998 of the FHS. For this type of study, missing information in any wave is possible because of (i) bounded missing, (ii) loss to follow-up if for other reasons than death or (iii) death. In the FHS, death is not a reason for loss of follow-up. We assumed that missing observations were only possible before death. Subsection 4.3.1 examines the total missing cases and 4.3.2 illustrates the missing values of two risk factors in the FHS.

4.3.1 Total missing cases in the FHS

Table 4.2 shows the distribution of absent or missing cases according to the total number of exams missing until end of follow-up or death. The first column of this table represents the number of exams missing per individual and column 2 is the frequency of missing cases. Out of a total of 5,209 respondents, 902 (17.32 percent) had missed only 1 exam. For 30 persons (0.58 %), 23 out of 24 exams were missing. A total of 2,160 (41.47 percent) persons participated in all exams before death or throughout 48 years of follow-up.
Table 4.2 Total number of exams missing

<table>
<thead>
<tr>
<th>Total number of absent exams</th>
<th>Frequency</th>
<th>Percent</th>
<th>Total number of absent exams</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2160</td>
<td>41.47</td>
<td>13</td>
<td>48</td>
<td>0.92</td>
</tr>
<tr>
<td>1</td>
<td>902</td>
<td>17.32</td>
<td>14</td>
<td>37</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>441</td>
<td>8.47</td>
<td>15</td>
<td>34</td>
<td>0.65</td>
</tr>
<tr>
<td>3</td>
<td>321</td>
<td>6.16</td>
<td>16</td>
<td>38</td>
<td>0.73</td>
</tr>
<tr>
<td>4</td>
<td>223</td>
<td>4.28</td>
<td>17</td>
<td>35</td>
<td>0.67</td>
</tr>
<tr>
<td>5</td>
<td>163</td>
<td>3.13</td>
<td>18</td>
<td>21</td>
<td>0.40</td>
</tr>
<tr>
<td>6</td>
<td>140</td>
<td>2.69</td>
<td>19</td>
<td>22</td>
<td>0.42</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>1.92</td>
<td>20</td>
<td>15</td>
<td>0.29</td>
</tr>
<tr>
<td>8</td>
<td>118</td>
<td>2.27</td>
<td>21</td>
<td>22</td>
<td>0.42</td>
</tr>
<tr>
<td>9</td>
<td>90</td>
<td>1.73</td>
<td>22</td>
<td>21</td>
<td>0.40</td>
</tr>
<tr>
<td>10</td>
<td>89</td>
<td>1.71</td>
<td>23</td>
<td>30</td>
<td>0.58</td>
</tr>
<tr>
<td>11</td>
<td>71</td>
<td>1.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>68</td>
<td>1.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Total</strong></td>
<td><strong>5209</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Table 4.3 shows the number of missing cases for each exam due to non-participation in the exam. The 1st column is the order of exam; the 2nd column shows the cumulative number of deaths; the 3rd column indicates the number of survivors; the 4th column is the number of persons who participated in the exam; the 5th column is the number of respondents who did not participate in the exam and the 6th column is the percentage of missing cases for reasons other than death.

A total of 5,209 respondents attended the first exam in the FHS. The death of a participant was recorded in the next most recent exam. If a subject died between $z^{th}$ and $(z+1)^{th}$ exam date, his or her death would be recorded at $(z+1)^{th}$ exam. For example, of the 5,209 persons who participated in the first exam, 33 died before the 2nd exam and 384 did not participate in the 2nd exam for reasons other than death. The remaining 4,792 (i.e. 5,209-33-384) individuals participated in the second exam (Table 4.3). Out of a total of 87,578 potential observations, some 71,778 observations were recorded. The non-response rate in the Framingham Heart Study was 22 percent. Table 4.3 and Figure 4.1 show that, as the years go by, the percentage of missing responses increases. Almost 7 percent of the respondents were missing at the 2nd round interview and about 34 percent were missing from the 24th exam. A sudden increase of the number of non-participants could be observed at exam 11. The NHLBI (National Heart, Lung, and Blood Institute) had directly staffed and funded the Framingham Study original cohort for 20 years. At exam 11, direct funding was running out, and the NHLBI was going to close the examinations. They were therefore unable to complete the full number of exams. The study ran on a variety of foundation type supports for a few years, and was
then picked up again and funded under a NHLBI contract with Boston University (Sorlie, personal communication). After 48 years of follow-up, 1250 respondents (almost 24 percent) of the initial members, are still under observation. A participant may be absent from one or several exams, and although the exact date of an event (if any occurred) will be recorded in any of the subsequent exams, the risk factor values (e.g. blood pressure) in the missing exam(s) will remain missing.

Table 4.3 Missing cases in given exam

<table>
<thead>
<tr>
<th>Exam</th>
<th>Cumulative number of deaths</th>
<th>Number of survivors</th>
<th>Number of cases participated</th>
<th>No participation</th>
<th>Missing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>5209</td>
<td>5209</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>5176</td>
<td>4792</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>87</td>
<td>5122</td>
<td>4416</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>151</td>
<td>5058</td>
<td>4541</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>230</td>
<td>4979</td>
<td>4421</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>332</td>
<td>4877</td>
<td>4259</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>429</td>
<td>4780</td>
<td>4191</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>548</td>
<td>4661</td>
<td>4030</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>688</td>
<td>4521</td>
<td>3833</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>833</td>
<td>4376</td>
<td>3595</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1017</td>
<td>4192</td>
<td>2955</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1186</td>
<td>4023</td>
<td>3261</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1395</td>
<td>3814</td>
<td>3133</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1583</td>
<td>3626</td>
<td>2871</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1776</td>
<td>3433</td>
<td>2632</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2024</td>
<td>3185</td>
<td>2351</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>2244</td>
<td>2965</td>
<td>2179</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2524</td>
<td>2685</td>
<td>1825</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2808</td>
<td>2401</td>
<td>1541</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3063</td>
<td>2146</td>
<td>1401</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>3283</td>
<td>1926</td>
<td>1319</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>3517</td>
<td>1692</td>
<td>1166</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3728</td>
<td>1481</td>
<td>1026</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3959</td>
<td>1250</td>
<td>831</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>71,778</td>
<td>15,800</td>
<td>22</td>
</tr>
</tbody>
</table>
4.3.2 Missing risk factors

In the FHS, the exact timing of events is known. The risk factors are missing in some exams. A risk factor variable can be binary (e.g. smoking status, yes/no) or continuous response (e.g. blood pressure). To illustrate the missing statistics in the FHS we considered two selected risk factors: smoking and blood pressure. It is possible to visualize the missing data in the FHS for other risk factors in the same manner.

In this section, the patterns of missing information on smoking and blood pressures in the FHS are described briefly. As we discussed earlier (Section 4.1), there are three types of missing cases in the FHS. Missing type I occurs when some respondents do not participate in that exam. Missing type II occurs when the risk factor information is not collected although the respondent participates in the exam. Missing type III occurs when the risk factor information for all respondents is not conducted in the relevant exam.

**Smoking**

Smoking status of the FHS respondents was recorded by asking two questions: currently smoking (yes or no) and number of cigarettes per day (24 hours). We have demonstrated current smoking status only. The smoking status recorded at each exam is given in Table 4.4. There is no information on smoking status at exam numbers 2, 3, 6 and 16. After 48 years of follow-up, some 71,778 recorded responses had been collected, 48.54 percent of which indicated that the respondent smoked,, 29.89 percent were recorded as non-smokers and the remaining 21.57 (type I 19.15 and type II 2.42 percent) percent were missing. Of the potential observations, 20.96 percent were missing type III i.e. smoking status failed to be...
recorded at exams 1, 2, 3 and 6. At the time of entry or at the first exam, 57.3 percent of the respondents were current smokers and only 0.6 percent was missing (missing type I). At the 24th exam, after 48 years of follow-up only 4.16 percent replied that they were current smokers and 62.24 percent indicated that they were non-smokers. Smoking status was missing for 33.52 percent respondents in last exam.

Table 4.4 Smoking status by exam, FHS

<table>
<thead>
<tr>
<th>Exam</th>
<th>Non-smoker (%)</th>
<th>Smoker (%)</th>
<th>Missing (%)</th>
<th>Total survivors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42.02</td>
<td>57.34</td>
<td>0.00 0.63</td>
<td>5209</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100.00 5176</td>
</tr>
<tr>
<td>3</td>
<td>41.89</td>
<td>46.58</td>
<td>10.22 1.30</td>
<td>5058</td>
</tr>
<tr>
<td>4</td>
<td>41.65</td>
<td>46.98</td>
<td>11.21 0.16</td>
<td>4979</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100.00 4877</td>
</tr>
<tr>
<td>6</td>
<td>43.66</td>
<td>43.89</td>
<td>12.32 0.13</td>
<td>4780</td>
</tr>
<tr>
<td>7</td>
<td>44.65</td>
<td>41.82</td>
<td>13.54 0.00</td>
<td>4661</td>
</tr>
<tr>
<td>8</td>
<td>46.07</td>
<td>38.71</td>
<td>15.22 0.00</td>
<td>4521</td>
</tr>
<tr>
<td>9</td>
<td>48.01</td>
<td>34.12</td>
<td>17.85 0.00</td>
<td>4376</td>
</tr>
<tr>
<td>10</td>
<td>26.00</td>
<td>13.17</td>
<td>29.51 31.32</td>
<td>4192</td>
</tr>
<tr>
<td>11</td>
<td>52.97</td>
<td>28.09</td>
<td>18.94 0.00</td>
<td>4023</td>
</tr>
<tr>
<td>12</td>
<td>55.45</td>
<td>26.66</td>
<td>17.86 0.03</td>
<td>3814</td>
</tr>
<tr>
<td>13</td>
<td>55.93</td>
<td>23.14</td>
<td>20.82 0.00</td>
<td>3626</td>
</tr>
<tr>
<td>14</td>
<td>56.25</td>
<td>20.33</td>
<td>23.33 0.09</td>
<td>3433</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100.00 3185</td>
</tr>
<tr>
<td>16</td>
<td>51.20</td>
<td>15.38</td>
<td>26.51 6.91</td>
<td>2965</td>
</tr>
<tr>
<td>17</td>
<td>56.46</td>
<td>11.21</td>
<td>32.03 0.30</td>
<td>2685</td>
</tr>
<tr>
<td>18</td>
<td>55.10</td>
<td>8.79</td>
<td>35.82 0.29</td>
<td>2401</td>
</tr>
<tr>
<td>19</td>
<td>56.34</td>
<td>8.29</td>
<td>34.72 0.65</td>
<td>2146</td>
</tr>
<tr>
<td>20</td>
<td>62.25</td>
<td>6.23</td>
<td>31.52 0.00</td>
<td>1926</td>
</tr>
<tr>
<td>21</td>
<td>63.30</td>
<td>5.50</td>
<td>31.09 0.12</td>
<td>1692</td>
</tr>
<tr>
<td>22</td>
<td>64.69</td>
<td>4.59</td>
<td>30.72 0.00</td>
<td>1481</td>
</tr>
<tr>
<td>23</td>
<td>62.24</td>
<td>4.16</td>
<td>33.52 0.08</td>
<td>1250</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48.54</td>
<td>29.89</td>
<td>19.15 2.42</td>
<td>20.96 87,578</td>
</tr>
</tbody>
</table>

Note: Missing type I: no participation in the exam
Missing type II: participated in the exam but missing in reporting smoking status
Missing type III: smoking status is not recorded
**Systolic blood pressure (BP)**

We analyzed the BP in mm Hg recorded by the second examiner. The systolic blood pressure status at each exam is given in Table 4.5. There are no type III missing responses in the BP recording of this study, only type I and type II. BP had been recorded for 92.2% respondents at the first exam; the remaining 7.83% were missing, i.e. no BP measurement was taken. Overall, 22.93 percent (type I 18.04 percent and type II 4.89 percent) values of BP were missing during the follow-up period, even though the respondents in question had been examined.

<table>
<thead>
<tr>
<th>Exam</th>
<th>SPS (%)</th>
<th>Missing (%)</th>
<th>Total survivors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Missing-type-I</td>
<td>Missing-type-II</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>92.17</td>
<td>0.00</td>
<td>7.83</td>
</tr>
<tr>
<td>2</td>
<td>87.60</td>
<td>7.42</td>
<td>4.98</td>
</tr>
<tr>
<td>3</td>
<td>80.38</td>
<td>13.78</td>
<td>5.84</td>
</tr>
<tr>
<td>4</td>
<td>78.65</td>
<td>10.22</td>
<td>11.13</td>
</tr>
<tr>
<td>5</td>
<td>79.43</td>
<td>11.21</td>
<td>9.36</td>
</tr>
<tr>
<td>6</td>
<td>83.92</td>
<td>12.67</td>
<td>3.40</td>
</tr>
<tr>
<td>7</td>
<td>85.98</td>
<td>12.32</td>
<td>1.69</td>
</tr>
<tr>
<td>8</td>
<td>84.47</td>
<td>13.54</td>
<td>2.00</td>
</tr>
<tr>
<td>9</td>
<td>83.43</td>
<td>15.22</td>
<td>1.35</td>
</tr>
<tr>
<td>10</td>
<td>81.86</td>
<td>17.85</td>
<td>0.30</td>
</tr>
<tr>
<td>11</td>
<td>38.50</td>
<td>29.51</td>
<td>0.32</td>
</tr>
<tr>
<td>12</td>
<td>80.86</td>
<td>18.94</td>
<td>0.20</td>
</tr>
<tr>
<td>13</td>
<td>81.65</td>
<td>18.78</td>
<td>0.50</td>
</tr>
<tr>
<td>14</td>
<td>78.19</td>
<td>20.82</td>
<td>0.99</td>
</tr>
<tr>
<td>15</td>
<td>75.12</td>
<td>23.33</td>
<td>1.54</td>
</tr>
<tr>
<td>16</td>
<td>72.72</td>
<td>26.19</td>
<td>1.10</td>
</tr>
<tr>
<td>17</td>
<td>70.46</td>
<td>26.51</td>
<td>3.04</td>
</tr>
<tr>
<td>18</td>
<td>67.37</td>
<td>32.03</td>
<td>0.60</td>
</tr>
<tr>
<td>19</td>
<td>63.64</td>
<td>35.82</td>
<td>0.54</td>
</tr>
<tr>
<td>20</td>
<td>65.14</td>
<td>34.72</td>
<td>0.14</td>
</tr>
<tr>
<td>21</td>
<td>68.28</td>
<td>31.52</td>
<td>0.21</td>
</tr>
<tr>
<td>22</td>
<td>69.00</td>
<td>31.09</td>
<td>0.00</td>
</tr>
<tr>
<td>23</td>
<td>69.07</td>
<td>30.72</td>
<td>0.20</td>
</tr>
<tr>
<td>24</td>
<td>46.24</td>
<td>33.52</td>
<td>20.24</td>
</tr>
<tr>
<td>Total</td>
<td>77.07</td>
<td>18.04</td>
<td>4.89</td>
</tr>
</tbody>
</table>
4.4 Missing value imputation

Imputation is the name given to any method whereby missing values in a data set are filled-in with reasonable estimates (SOLAS, 1999). The ultimate goal of any imputation method is to produce a complete data set. That complete data set can be analyzed using complete-data inferential methods. There are two common procedures for dealing with missing data. Either all data records in which any variable value is missing are excluded or ‘plausible’ values are substituted for the missing items (Goldstein and Woodhouse, 1996). Methods based upon the use of plausible or imputed values underlie most recommended procedures, and detailed discussions are given by Rubin (1987), Little and Rubin (1987) and Little (1993). To analyze the risk factor career in Framingham Heart Study, we need imputation. In the literature, there are several popular methods for handling missing data. Much of the literature involving missing data in public health research pertains to the various methods developed to handle the problem. For details of the available methods to impute missing values we refer to Rubin (1987); Hedeker and Gibbons (1997); Heitjan (1997); Little (1993); Little and Schenker (1995); Little and Rubin (1987). However, there are no efficient methods for imputing missing data in repeated measurement.

Handling missing values for repeated measurement, public health researchers frequently use the last value carried forward (LVCF) method. The last observed value is used to fill in missing values at a later point in the study. LVCF makes the assumption that the response remains constant at the last observed value. In any intervention study (e.g. clinical trial), this can be biased if the timing and rate of withdrawal is related to the treatment (William, 2000). For example, in the case of degenerative diseases, using the last observed value to impute missing data at a later point in the study can produce biased results.

Multiple imputation\(^1\) has good statistical properties, although this method has not yet been extensively used in longitudinal data analysis. Moreover, the method had not been used heretofore with repeated measurement data, especially follow-up data over a long period. The main reason for this is that missing values in repeated

---

\(^1\) Multiple imputation: MI is a Monte Carlo technique in which the missing values are replaced by \(m>1\) simulated versions, where \(m\) is typically small (e.g. 3-10). In Rubin's (1987) method for ‘repeated imputation’ inference, each of the simulated complete data sets is analyzed by standard methods, and the results are pooled to produce estimates and confidence intervals that incorporate missing-data uncertainty. Primarily, Rubin (1987) addresses potential uses of MI for large public-use data files from sample surveys and censuses. With the advent of new computational methods and software for creating MI's, however, the technique has become increasingly attractive for researchers in different backgrounds whose investigations are hindered by missing data. Schafer (1997) documents these methods in a recent book on incomplete multivariate data.
measurement data are mostly correlated with covariates that make the explanation of the missing mechanism critical. The lengthy procedure of multiple imputation and covariates dependency in repeated measurement data are an impediment to the application of multiple imputation. Little (1995) modeled the dropout mechanism in a repeated measurement study. He used a pattern mixture model for various missing data mechanisms. Wei and Shih (2001) proposed a partial imputation approach to analyze the repeated measurements with dependent dropouts. Their method is an extension of LVCF. Before choosing a missing data handling approach, we might keep in mind that one of the desired outcomes is maintaining (or approximating as closely as possible) the shape of the original distribution of responses. Some incomplete data handling methods do a better job of maintaining the distributional shape than others.

Because of the limited use and disadvantages of available methods in longitudinal data, we have proposed two approaches: one for categorical risk factors and another for continuous risk factors. For categorical risk factors, we propose an algorithm which is an extension of LVCF (say, ELVCF). For continuous risk factors, we propose a regression method which is an extension of the regression method (say, ERM). These methods are applicable for categorical and continuous missing data in repeated measurement study.

We assumed that the missing data in the FHS were MAR and that the pattern was monotone. We have illustrated the method with an example. To describe the ELVCF method, we treated smoking as a categorical variable: smokers and non-smokers. Based on several assumptions, we imputed values for missing smoking status. For the ERM method, we considered systolic blood pressure to be a continuous variable. We imputed the missing values of BP based on the observed distribution of BP. The imputation procedures are illustrated with an example.

4.4.1 Smoking status imputation

We first treated smoking as a categorical variable: smokers and non-smokers. People were viewed as having a ‘smoking career’ throughout the observation period. They could drop in and out of smoking status. There was no ‘ex-smoker’ category. Now, not everyone turned up at all rounds. We imputed likely values, and flagged these according to the following algorithm, assuming a missing value at exam or round i. An individual could be absent from one or from several consecutive exams or drop out. We therefore limited our method to imputing missing values for a maximum of two consecutive missing exams.

---

2 A monotone missing data pattern occurs when the variables can be ordered, from left to right, such that a variable to the left is at least as observed as all variables to the right.
Box  Smoking code after and before imputation

<table>
<thead>
<tr>
<th>Smoking code after and before imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1:</strong> smoking at that exam (from original data)</td>
</tr>
<tr>
<td><strong>0:</strong> not smoking at that exam (from original data)</td>
</tr>
<tr>
<td><strong>2=smoking, imputed</strong></td>
</tr>
<tr>
<td><strong>3=non-smoking, imputed</strong></td>
</tr>
<tr>
<td><strong>9=missing</strong></td>
</tr>
</tbody>
</table>

2: conditions are as follows:
- a. If $i^{th}$ round missing (9) and $(i-1)^{th}$ round smoked (1) and $(i+1)^{th}$ round smoked (1) then **code=2**
- b. If $i^{th}$ round missing (9) and $(i-1)^{th}$ round smoked (1) and $(i+1)^{th}$ round not smoked (0) then **code=2**
- c. If $i^{th}$ round missing (9) and $(i-1)^{th}$ round smoked (1) and $(i+1)^{th}$ round missing (9) then **code=2**

3: conditions are as follows:
- a. If $i^{th}$ round missing (9) and $(i-1)^{th}$ round not smoked (0) and $(i+1)^{th}$ round smoked (1) then **code=3**
- b. If $i^{th}$ round missing (9) and $(i-1)^{th}$ round not smoked (0) and $(i+1)^{th}$ round not smoked (0) then **code=3**
- c. If $i^{th}$ round missing (9) and $(i-1)^{th}$ round not smoked (0) and $(i+1)^{th}$ round missing (9) then **code=3**

9: conditions are as follows:
- a. If $i^{th}$ round missing (9) and $(i-1)^{th}$ round missing (9) and $(i+1)^{th}$ round missing (9) then **code=9**

Following the description and code scheme shown in the Box, the algorithm for imputation of missing smoking status can be described as follows:

i. If we have consistent observation at round $i-1$ and at round $i+1$, smoking status is imputed as observed status at both observed rounds.

ii. If we have inconsistent information at round $i-1$ or at round $i+1$, smoking status is imputed based on the assumption that the state changes at midpoint.

iii. If we have one missing round, the status of the previous round is imputed (i.e. carrying forward)

iv. If we have two consecutive missing rounds, the status of the round nearest in time is imputed (i.e. carrying forward for the first and carrying backward for the second)

v. If three or more consecutive rounds are missing, only the middle record is assigned the status 'missing'.

vi. If the information we have is all ‘right censored’ due to the fact that the subject failed to show up at all subsequent rounds, i.e. up to and including the 24th exam, the observations are recorded as missing.
The smoking history of an individual (case=16) participant in the FHS is shown in Table 4.6. This individual entered the FHS at age 33 as a non-smoker, was absent from the 2nd and 3rd exam (smoking status was not gathered at the 2nd and 3rd exam), and was a smoker at the 4th and 5th round. He was missing at the 6th round, a smoker at exam 7 to exam 10, a non-smoker at the 11th exam, but a smoker again at the 12th and 13th exam; from the 14th exam onward he was absent from the survey and his death (DTH) was recorded in 19th exam, i.e. he died between the 18th and 19th exam. According to our algorithm and coding scheme (Box), the missing values at the 2nd and 3rd exam were imputed as 3 and 2 respectively. He was absent from the 6th exam, which value is replaced by 2. As he was absent after the 13th
exam onward, all missing values remained missing after the 14th exam to the time of his death (i.e. before the 19th exam).

In the FHS, 87,578 smoking status values have been collected in all potential exams (from exam 1 to exam 24). Following the algorithm described above, we imputed the missing values of the smoking variables in the FHS. An explanation of the coding used for smoking status before and after imputation is given in the Box.

Table 4.7 shows the smoking status distribution of males and females of the FHS cohort before and after imputation. A total of 33,289 (38.0 percent) observations of smoking status were missing in the FHS. Of these 11,905 (13.6 percent) responses of non-smokers and 10,438 (11.9 percent) responses as smokers were able to be imputed. After imputation, 10,946 (12.5 percent) observations were still missing. Following the exclusion of the missing observations, before imputation 51.7 percent of males and 28.4 percent of females were shown to be smokers. After imputation, 55.2 percent of males and 30.1 percent of females were found to be smokers.

Table 4.7  Smoking status before and after imputation, FHS

<table>
<thead>
<tr>
<th>Smoking status</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% of total values (N)</td>
<td>% of total values (N)</td>
<td>% of total values (N)</td>
</tr>
<tr>
<td>Before imputation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-smoker</td>
<td>30.2 (10949)</td>
<td>48.3</td>
<td>44.1 (22653)</td>
</tr>
<tr>
<td>Smoker</td>
<td>32.3 (11705)</td>
<td>51.7</td>
<td>17.5 (8982)</td>
</tr>
<tr>
<td>Missing</td>
<td>37.4 (13556)</td>
<td>-</td>
<td>38.4 (19733)</td>
</tr>
<tr>
<td>After imputation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded</td>
<td>30.2 (10949)</td>
<td>34.1</td>
<td>44.1 (22653)</td>
</tr>
<tr>
<td>Imputed</td>
<td>9.5 (3438)</td>
<td>10.7</td>
<td>16.5 (8467)</td>
</tr>
<tr>
<td>Total</td>
<td>39.7 (14387)</td>
<td>44.8</td>
<td>60.6 (31120)</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded</td>
<td>32.3 (11705)</td>
<td>36.5</td>
<td>17.5 (8982)</td>
</tr>
<tr>
<td>Imputed</td>
<td>16.6 (6003)</td>
<td>18.7</td>
<td>8.6 (4435)</td>
</tr>
<tr>
<td>Total</td>
<td>48.9 (17708)</td>
<td>55.2</td>
<td>26.1 (13417)</td>
</tr>
<tr>
<td>Missing</td>
<td>11.4 (4115)</td>
<td>-</td>
<td>13.3 (6831)</td>
</tr>
</tbody>
</table>

Note: N is the number of responses

### 4.4.2 Systolic blood pressure imputation

Blood pressure is measured as a continuous variable in the FHS. Before imputation, we plotted the mean value of BP by age of respondents, because BP is age dependent. We were interested to see whether individual prediction equations could
be derived from which missing values could be imputed. We explored the shape of the observed mean of BP against age. We found that the shape fit a quadratic relationship of BP as age changed, i.e. non-linear regression, where BP was a response variable and age and age*age were independent variables. The observed means of BP by age are shown in Figure 4.2. For the observed plot, the X-axis represents respondent age at the time of exam and Y-axis represents BP mean by age at that corresponding exam. After plotting the observed BP mean against age, the trend line drawn followed a quadratic equation

\[ y = 115.82 + 0.9812x + (-0.0081)x^2 \text{ with } r^2 = 0.925 \]  

which is a polynomial of second order. This polynomial shows the relationship of BP with age at the population level. We then drew the same trend line for each individual and found that it fit properly, i.e. the observed and prediction line were close to each other. This prediction demonstrated that quadratic regression fit BP quite well. Therefore, we used a simple quadratic regression model for each individual to obtain a better prediction and imputation.

Figure 4.2 Observed mean of blood pressure by age, FHS

Before describing the basic steps in the imputation of BP, we imposed some limits on the prediction rules and on the number of consecutive imputations. To calculate the prediction equation, we excluded all values of BP up to 4 years before death and imputed missing BP where age was present (i.e. exam date is present but BP is missing). Imputation was only used for exams at which the subject turned up (i.e. age not missing). Theoretically, this method could be used to calculate the BP
value for any age of an individual regardless of whether they attended an exam or not.

We excluded the values just before death (in 4 years). The reason is that, at the end of life, the relationship between high blood pressure and mortality becomes less consistent, at which point BP is no longer predicted well by age. Thus trying to include these values would have made our prediction less accurate. These are therefore left out of our prediction derivation to avoid the problem.

**Basic steps**

**Step 1 Population level**

Plot the observed mean of BP against age and find out the shape of the observed curve for overall population.

**Step 2 Individual level**

(i) **Regression coefficient:** estimate regression coefficients for each individual. Let ‘a’ denotes regression constant, b and c regression coefficients. Using the value of a, b and c for each individual we made our prediction equation.

\[
y_i = a_i + b_i \cdot \text{age}_i + c_i \cdot \text{age}_i \cdot \text{age}_i,
\]

where \(i\) is individual. Now for given value of age and parameters a, b and c, predict BP.

(ii) **Comparison:** plot the predicted BP with the observed one (Figures 4.2).

(ii) **Imputation:** for given value of age and parameters a, b and c, impute the missing values of BP.

The whole procedure is described with the help of an example. An individual entered the FHS at age 33, at which time his BP was 138 mm Hg. Except for the 2\textsuperscript{nd} and 3\textsuperscript{rd} exam, his BP was recorded up to 13\textsuperscript{th} exam, after which he was absent from subsequent exams (i.e. his age was not recorded) and died between the 18\textsuperscript{th} and 19\textsuperscript{th} exam, which is recorded in 19\textsuperscript{th} exam. We imputed the missing values from the 2\textsuperscript{nd} and 3\textsuperscript{rd} exam according to our algorithm. The variable BP was observed blood pressure and BP_4 was observed, but BP observations 4 years before death were excluded. The variables, BP\_a, BP\_b and BP\_c represent regression (non-linear) parameter estimates of that individual. From the variable BPCODE, we can identify how many responses were missing and imputed. Code 1 indicates no missing and no imputation and code 2 is missing and imputed. The variable BPIMP shows imputed values using our approach.
Table 4.8 Blood pressure level of an individual before and after imputation, FHS

<table>
<thead>
<tr>
<th>Exam ID</th>
<th>Age</th>
<th>BP</th>
<th>Death</th>
<th>BP_4</th>
<th>BP_a</th>
<th>BP_b</th>
<th>BP_c</th>
<th>BPCODE</th>
<th>BPIMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33</td>
<td>138</td>
<td>19</td>
<td>138</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>138</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
<td>missing</td>
<td>19</td>
<td>missing</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>2</td>
<td>133.7</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>missing</td>
<td>19</td>
<td>missing</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>2</td>
<td>133.9</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>140</td>
<td>19</td>
<td>140</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>140</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>132</td>
<td>19</td>
<td>132</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>132</td>
</tr>
<tr>
<td>6</td>
<td>43</td>
<td>122</td>
<td>19</td>
<td>122</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>122</td>
</tr>
<tr>
<td>7</td>
<td>45</td>
<td>124</td>
<td>19</td>
<td>124</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>124</td>
</tr>
<tr>
<td>8</td>
<td>47</td>
<td>134</td>
<td>19</td>
<td>134</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>134</td>
</tr>
<tr>
<td>9</td>
<td>49</td>
<td>134</td>
<td>19</td>
<td>134</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>134</td>
</tr>
<tr>
<td>10</td>
<td>51</td>
<td>150</td>
<td>19</td>
<td>150</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>150</td>
</tr>
<tr>
<td>11</td>
<td>53</td>
<td>138</td>
<td>19</td>
<td>138</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>138</td>
</tr>
<tr>
<td>12</td>
<td>56</td>
<td>170</td>
<td>19</td>
<td>170</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>170</td>
</tr>
<tr>
<td>13</td>
<td>59</td>
<td>114</td>
<td>19</td>
<td>114</td>
<td>141.33</td>
<td>-0.50</td>
<td>0.01</td>
<td>1</td>
<td>114</td>
</tr>
</tbody>
</table>

4.5 Validation of the methods

To control for the adverse effects of missing data on a particular analysis, we made use of imputation. To assess the implications of the imputation method and associated assumptions, we needed a sensitivity analysis. Sensitivity analysis is used to ascertain how a given model output depends upon the input parameters. Two different algorithms (one for categorical variable and another one for continuous variable) are discussed and applied to impute the missing values in repeated measurement data. We assumed that the missing values in the FHS occurred at random. Thus, to validate our method and assumption, we performed a sensitivity analysis for the missing value imputation under different scenarios.

Overall, 26 percent of the missing responses in smoking status and 6 percent of the missing information on the BP risk factor were imputed. These imputed observations could influence the final outcome. Therefore, a sensitivity analysis of the imputation method was carried out. In the literature, most of the imputation methods deal with the mean value at population level. Using our method, missing values are imputed from the individual follow-up or observations that may not be comparable to other available methods. We carried out the sensitivity analysis based on the simple random selection procedures.

As illustrated, we imputed the missing values of two risk factors: smoking status and blood pressure, and created a scenario using simple random sampling technique for both the categorical and the continuous risk factors. We validated our imputation method based on this scenario. First, we imputed the missing values using the above method. Excluding the remaining missing observations after
imputation, we assumed a complete data set without any missing observations. Second, we deleted various observations randomly from this complete data set. In the scenarios, we deleted 10 percent, 20 percent and 30 percent of the observations based on simple random selection procedures. Third, we imputed the randomly deleted observations applying the same methods that we applied to impute the missing values for the original data set. Fourth, we compared the scenarios. We assumed that after imputation of the randomly deleted observations we would achieve the same output that was obtained after imputation for the original data set. If this assumption held, our methods were valid.

The scenarios for smoking status after imputing the randomly deleted responses are shown in Table 4.9. There were 61.9 percent non-smoking responses and 38.1 percent smoking responses in all recorded smoking statuses (excluding missing responses). After imputation, this distribution was 59.4 percent and 40.6 percent respectively. The change in the percentage distribution of imputed values after random deletion with different scenarios remains almost same. We found that even after deleting 30 percent of responses, the change in percent distribution of the number of responses was minimal. This nominal change was mainly due to the restriction that we imposed on the imputation of missing values, namely that the values for a maximum of two consecutive missing exams were permitted to be imputed.

Table 4.9 Scenarios for smoking status - responses are imputed after random deletion

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Percent of non-smoking responses (number of responses)</th>
<th>Percent of smoking responses (number of responses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed population</td>
<td>61.9 (33602)</td>
<td>38.1 (20687)</td>
</tr>
<tr>
<td>After imputation</td>
<td>59.4 (45507)</td>
<td>40.6 (31125)</td>
</tr>
<tr>
<td>Imputed after random deletion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>59.2 (45303)</td>
<td>40.8 (31169)</td>
</tr>
<tr>
<td>20%</td>
<td>59.2 (44758)</td>
<td>40.8 (30891)</td>
</tr>
<tr>
<td>30%</td>
<td>59.3 (43779)</td>
<td>40.7 (30083)</td>
</tr>
</tbody>
</table>

The scenarios for blood pressure after imputing the randomly deleted responses are shown in Table 4.10 and Figure 4.3. The mean value of BP in the observed population was 136.87 mm Hg (SE 0.087). After imputation, this was 136.76 mm Hg (SE 0.0844). The change in the mean values of BP imputed after random deletion with different scenarios was almost negligible... Even after 30 percent random deletion, no substantial changes occurred in the BP mean and SE. After imputation with different scenarios, the total number of responses differed mainly because of two restrictions. (i) Imputation is only possible if we have at least
three responses. (ii) There is a problem of contiguity, i.e. values that are imputed too far away in time from the nearest time (i.e. age).

Table 4.10  Scenarios for blood pressure- responses are imputed after random deletion

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>BP mean</th>
<th>Standard error of mean</th>
<th>Total number of responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed population</td>
<td>136.9</td>
<td>0.087</td>
<td>67496</td>
</tr>
<tr>
<td>After imputation</td>
<td>136.8</td>
<td>0.0844</td>
<td>71637</td>
</tr>
<tr>
<td>Imputed after random deletion</td>
<td>136.8</td>
<td>0.085</td>
<td>70079</td>
</tr>
<tr>
<td>10%</td>
<td>136.8</td>
<td>0.085</td>
<td>70079</td>
</tr>
<tr>
<td>20%</td>
<td>136.9</td>
<td>0.085</td>
<td>69872</td>
</tr>
<tr>
<td>30%</td>
<td>137.1</td>
<td>0.088</td>
<td>67189</td>
</tr>
</tbody>
</table>

The age-specific mean values of BP with different scenarios are presented in Figure 4.3. This figure shows that imputed age-specific BP means with different scenario remains almost unchanged. Estimated values with different scenarios differ slightly at older ages (after age 80), since at older ages, the relationship of BP with age is more complex, the number of responses is smaller and the number of missing observations is higher.
4.6 Discussion

This chapter proposed to demonstrate a potential method to impute missing values of risk factors in a repeated measurement study. We illustrated the method using the well-known Framingham Heart Study. It was found that in over 48 years of follow-up, the percentage overall of non-missing observations was 41.47 percent in the Framingham Heart Study. We illustrated the missing values and the method of imputation used for two risk factors- smoking status as categorical variable and blood pressure as continuous variable, both of which were used in order to measure the risk career. The idea behind the proposed methodology is relevant to any study design subject to incomplete responses and having at least three repeated measurements.

We used a different method to impute missing records for a categorical risk factor than the method adopted for continuous risk factors. To impute smoking status, we followed the history of smoking status of the individual and imputed the value, which is closest in time to the true value, i.e. the value closest in time to the missing value. The method proposed for the imputation of BP is different. This method is an extension of regression procedure. Using the ordinary regression method, the expected population means are used to impute missing values. Using our regression method, we imputed the missing value for each individual from the information on each individual. The proposed method has several advantages over current approaches. Our method is easy to understand. The calculation procedure is very simple and not time consuming. Of paramount importance is that the curve has a good fit. These methods are easy to manipulate and applicable both for categorical and continuous risk factors in any long-time follow-up data. It also allows investigators to easily assess the sensitivity of the results obtained applying this method.

Using the proposed methods, we can easily reconstruct the risk factor career and measure the impact of this risk factor career from the life course perspective. For example, with the help of this method we were able to reconstruct the smoking career and to measure the life history of smokers and non-smokers, in order to construct the multistate smoking status life table in Chapter 5 of this study.

To validate our methods, we have presented the prevalence of smoking and blood pressure in the form of a scenario analysis. If we should wish to estimate the effect of imputation on some outcomes, e.g. CVD incidence and mortality, only the variability of the estimates will decrease because of the increasing sample observations. These observations are not independent of the observed values. Therefore, they should not be used. To predict the risk factor status at a specific age, we needed to impute the missing values. Using this approach, De Laet et al., (2003) predicted blood pressure at age 40, 50, and 60 and estimated the effect of
blood pressure on the life table outcomes. Based on the predicted blood pressure, the life table outcomes are more reliable than the observed information.

The major limitation of our method is that a certain number of rounds or waves are necessary, i.e. application of this method is possible only if at least three rounds of measurement are available. In developed countries, the number of longitudinal surveys has been on the rise for some time and the numbers of waves are increasing over time. Our method will be applicable to these data sets in the near future. Another limitation of this study is that if the individual returns to the study after a long period of time or the time span between one observation and the next is too long, the prediction may not fit with the observed trend.

Regarding the smoking career, we imputed missing values for two consecutive exams if bounded on both sides by non-missing values. Missing values have been imputed for BP only for the exam if the respondent’s age was reported. Using the same approach, we may impute missing values for all exams and construct the full-length non-missing data set. We may conclude that by applying this method, it is possible to increase the utility of missing values in repeated measurement, and increase the statistical precision of final results.

References


Smoking career

Abstract

In spite of the overwhelming evidence of the harmful effects of cigarette smoking, little is known about how smoking evolves over the life course. In this chapter, we have taken a novel approach to analyze the “career or pattern over time” of smoking in a well-documented population. The objective of this chapter was to analyze smoking status throughout life and translate it into life years lived in a smoking and non-smoking state. Multistate life tables were developed to capture the life history of smoking. The data were obtained from the Framingham Heart Study (FHS), which has a 48-year follow-up of 5,209 respondents residing in Framingham, Massachusetts, ranging in age from 28 to 62 years at the baseline survey between 1948-1951. Three states {smoking, non-smoking, dead} were distinguished. We defined smokers as individuals who were recorded as smoking at any given examination and non-smokers as those who were recorded as non-smoking at that examination. The smoking status recorded at each examination throughout the 48 years of follow-up of the FHS formed the basis for age-specific transition rates between smoking states.

At age 10, the total life expectancy of the FHS cohort was 66 years for males and 72 years for females. Males were shown to spend two-thirds of their expected lifetime (68%) smoking, females less than one third (28%). The difference was less at higher ages, by which time males with a history of smoking had either quit or died. The smoking careers of males started 2 years earlier than that of females. The average age (median age) at which males started smoking was 15, for females, 17. Fifty-eight percent of the males who started smoking did quit at some point, while for females this proportion was considerably lower, namely 37%. Compared to men, women smokers were more persistent. If a person quit smoking, they were most likely to do so between the ages of 50 and 70. After quitting smoking, a certain percentage will resume the habit over time. The probability of a relapse was 26% for both males and females. The novel approach described here demonstrated the usefulness of multistate life tables for describing how smoking evolves over the life course. While we cannot assume that similar patterns apply to other
populations, the method can be used for other risk factors to indicate current risk factors in life courses.

5.1 Introduction

Cigarette smoking was once so common that to imagine a world without it was difficult. Nevertheless, for thousands of years, consistent tobacco use was not the norm: in fact, such high prevalence and regular consumption are historically unusual. Today, the adverse health consequences of smoking is widely recognized (Peto et al., 1992; 1994; 1996; Doll et al., 1954; 1956; 1964; 1976; 1994; 1998; Thun and Heath, 1997; Surgeon General, 1964; 1979; 1989; 1990; Royal College of Physicians, 1962; 1971; Fielding, 1985). However, smoking remains the number one cause of preventable death in developed countries and it continues to be the largest single preventable cause of death and disability in the United States (Center for Disease Control and Prevention (CDC), 2001). The millions of lives that tobacco claims each year makes tobacco one of the most important public health issues of the present time (Bartecchi et al., 1994). Globally, an estimated 3 million people die from smoking-related disease each year (Peto et al., 1996). Tobacco causes 1 in 4 deaths in developed countries, 1 in 8 deaths in developing countries and 1 in 6 deaths worldwide (Barnum, 1994). Another study estimated that smoking may account for 25 percent of all deaths (Shopland, et al., 1990; Crimmins, 1981), 30 percent of all coronary heart disease deaths and for about 30 percent of all cancer death, counting 85 percent of all lung cancer deaths (American Cancer Society, 1986). Within 30 years, the number of global tobacco-related deaths will rise to 10 million per year, and tobacco will become the single largest cause of death worldwide (World Bank, 1999). However, very little research has targeted the smoking career on an empirical basis. The objective of this chapter was to analyze smoking status throughout life and to translate this into life years lived in the smoking and the non-smoking state.

Epidemiology and medical demography study the impact of risk factors on chronic diseases. Many risk factors are time varying. Smoking is a very important risk factor and a major preventable cause of morbidity and premature death. The hazards of smoking have been documented by many studies (Doll et al., 1954; 1956; 1964; 1976; 1994), which have been summarized by the Royal College of Physicians, (Royal College of Physicians 1962; 1971), the United States Surgeon General, (Surgeon General, 1964; 1979; 1989) and the International Agency for Research on Cancer (IARC monographs, 1986). To determine the risk and hazard of smoking associated with morbidity and mortality, researchers have in the past applied survival analysis to calculate the relative risk and the smoking attributable risk of mortality (cause specific or all cause combined) (Rogers et al., 2002),
Generally based on smoking status at a single point in time. In this chapter, we have taken a novel approach towards the analysis of the “smoking career or pattern over time” in a well-documented population. Less is known about the career of smokers. The well-known Framingham Heart Study afforded us with a unique opportunity to calculate the smoking career of a cohort of 5,209 people that has been followed since the middle of 1948. Some 48 years of follow-up of the smoking history of that cohort were available to us.

Despite various studies over the past 40 years, only a few studies have focused on the life expectancies of smokers and non-smokers (Miller and Gerstein, 1983; Rogers and Powell-Griner, 1991; Doll et al., 1994). However, to our knowledge there has heretofore been no approach that studied smoking status throughout life. There are several reasons for this: in the first place, most of the researchers used cross-sectional information, which meant that only the current smoking status was recorded; second, individuals have started smoking at increasingly early ages, and the effect is cumulative, largely between age 35-70. Gathering information on the age at which an individual starts smoking, quits or switches their smoking habits and on smoking behavior (e.g. style of smoking) over the life course is difficult. In the third place, their small sample size, local population base and emphasis on one smoking status rather than a range of statuses have limited many studies.

The studies calculating life expectancies of smokers and non-smokers by sex, used separate single decrement life tables (alive and dead state) for smokers and non-smokers (Miller and Gerstein, 1983; Rogers and Powell-Griner, 1991; Doll et al., 1994). The shortcomings of this approach included the use of smoking prevalence i.e. smoking status at a specific point of time. Chiefly, these studies considered the smoking status before death or any other event (e.g. lung cancer). Using this method, the current smoking status of a real population, adjusted for mortality levels, was reflected. It may therefore not represent the lifetime smoking status at the level of the individual. These studies used prevalence or stock data instead of incidence or flow data, and avoided reverse transition, i.e. non-smoking to smoking or smoking to non-smoking. The resultant estimation was mainly for the overall population. In our approach, we minimized all these restrictions. We used smoking incidence, which is measured over a long period of time. We focused on flows instead of stocks. Back flow, too, was captured. When, analyzing the life histories of smokers and non-smokers, we were therefore able to translate smoking status into life years lived in a smoking and a non-smoking state within a synthetic cohort.

We constructed a multistate life table to investigate the smoking career. Some of the life course questions that the multistate smoking status life table was able to answer were: what proportion of lifetime is spent by individuals as smokers? What is the lifetime probability of quitting smoking? If people quit smoking, when do they quit? Male-female mortality differs among smokers (Waldron, 1986). Does the
smoking career of a male differ from that of a female? We answered these questions, applying multistate life table techniques to the life histories of smokers and non-smokers in the FHS original cohort. The FHS study offers a unique historical documentation of fifty years of health damage due to smoking.

Section 5.2, in which the proposed method is illustrated, consists of a number of: data source, smoking status definition, imputation of missing smoking status, model specification, the input data and the life table construction. The results are given in Section 5.3. Section 5.4 concludes this chapter following a brief discussion.

5.2 Methods

We constructed smoking status life tables focusing on transitions between smokers and non-smokers in the first 48 years of follow-up from the Framingham Heart Study original cohort. Smoking prevalence at different times of follow-up, smoking episodes, survival probability, life expectancies in smoking and non-smoking state and lifetime probability of quitting smoking were considered separately for males and females.

The following subsections describe the data source, the definition of smoking status, imputation of smoking status, the input data and the construction of the multistate smoking status life table.

5.2.1 Data source

The data used for this study came from the Framingham Heart Study original cohort. The Framingham Heart Study is a well-known, epidemiological ongoing longitudinal study focusing on cardiovascular disease. The Framingham Study was designed to find out how those who develop cardiovascular diseases differ from those who remain free of the diseases over the life course in order to identify risk factors for cardiovascular disease. The original study cohort consisted of 5,209 respondents (45 percent male) from a sample of adults aged 28 through 62 years residing in Framingham, Massachusetts between 1948 and 1951. The participants were tracked by standardised biennial cardiovascular examination, daily surveillance of hospital admissions, death information and information from physicians and other sources outside the clinic, ensuring highly accurate follow-up of death and clinically presenting cardiovascular disease. Among other characteristics, the smoking status of the participants was recorded at each examination. Details of the FHS are documented elsewhere (Dawber and Moore, 1952; Shurtleff, 1971). For the current study, we used the data regarding smoking status over 48 years of follow-up (exam rounds 1 to 24) of the FHS original cohort.
5.2.2 Smoking status

The smoking status of the FHS respondents was recorded on the basis of the responses to two questions. Participants were asked whether they currently smoked (yes or no) and about the number of cigarettes smoked within 24 hours. The first is a binary variable, the second a count variable. To characterize the pattern of smoking over time and the effect of smoking on mortality, our definition of smoker versus non-smoker at any point in time took into account the smoking history recorded in each exam. “Smokers” were individuals recorded as smoking at a given examination and “non-smokers” were those who were recorded as non-smoking in that examination. The smoking status is therefore time varying and up-dated.

At the first exam, all current smokers were asked ‘at what age’ they had started smoking. At that exam, all ever-smokers were also asked when they had relapsed back into smoking. Some ever-smokers were not able to indicate when they had started. As we were interested in constructing the smoking career, the respondents whose starting time of smoking was unknown were excluded (277 respondents). Smoking has the tendency to start at early ages. We assumed all participants were non-smokers at age 10, i.e. smoking initiation was possible from age 10 onwards. In the FHS, a few individuals had started smoking even before the age of 10 (45 respondents). For the purpose of this study, we assumed that these individuals had started at age 10. We moreover excluded respondents who had cardiovascular disease at the first exam (139 respondents), while smoking status was missing for 27 respondents at the first exam and 72 respondents were lost to follow-up. Finally, we were left with 4694 (43 percent males) participants whose smoking careers we proposed to investigate.

5.2.3 Missing value imputation

The aim of this chapter was to construct the smoking career of the Framingham Heart Study cohort. Smoking status data was collected biennially on the FHS participants. Over the long period of follow-up, there were several exams at which smoking status was not recorded. For instance, no data on the smoking status of the respondents was collected at exams 2, 3, 6 and 16. Sometimes respondents were present at an exam without their smoking status being recorded. Without imputation of the missing values, construction of the smoking career was not plausible, nor would we be able to follow the smoking career of various individuals. We therefore imputed the missing values of the smoking status.

We imputed smoking status essentially assuming that changes in smoking status occurred midway through an unobserved period and that imputation was only allowable for values 2 years or less (i.e. 1 round) from a recorded value. Thus if smoking status was known at an exam prior to a missing value, this value was
imputed forward. If a value was not available from the exam prior to the missing value the value from the next exam was imputed backward. If a missing value was surrounded by two other missing exams, it was not imputed and remained missing. For details of the imputation method and the values after and before imputation, we refer to Chapter 4, Section 4.4.

5.2.4 Model specification and input data

We constructed a 3-state multistate smoking status life table. The state space was \{smoking, non-smoking, dead\}. Smoking and non-smoking were communicable states and dead was an absorbing state. For instance, at age 30 a male could be a smoker; the same man could be non-smoker at age 50 and a smoker again at age 65, to become a non-smoker once again before death. Here, we have used the updated smoking status.

The age-specific occurrences (event counts) and exposures were the basic input data for the multistate smoking status life table. Since smoking status was recorded at each biannual exam, we assumed that a transition from non-smoking to smoking or smoking to non-smoking would occur midway between two exam dates. Transition to death was recorded at the exact date. Both the occurrences and exposures were calculated using the approach described in Section 3.3.3 of Chapter 3. We estimated occurrences and exposures separately for the total population, for males and for females.
5.2.5 Multistate life table construction

The multistate smoking status life tables were constructed based on the age-cohort observational plan (Willekens, 1987). The calculation of occurrence-exposure rates, the conversion into transition probability and other life table statistics were described briefly as follows.

To construct the life tables, it was necessary to convert the transition rates $M_{ij}[x,x+1]$ (see equation 2.10 in Chapter 2) into transition probabilities. The transition rates were converted to probabilities for use in the life table by assuming that within an age interval (one-year), the transitions (events) were uniformly distributed. Following the standard approach suggested by Rogers (1975; 1995) and Willekens et al. (1982; 1987), the transition probability matrix $P_{[x,x+1]}$ was calculated from the matrix of occurrence-exposure rates, $M_{[x,x+1]}$. The transition probability matrix $P_{[x,x+1]}$ can be calculated from the transition rate matrix $M_{[x,x+1]}$ using:

\[
P_{[x,x+1]} = \left(I + \frac{1}{2}M_{[x,x+1]}\right)^{-1}\left(I - \frac{1}{2}M_{[x,x+1]}\right)
\]

where $I$ is an identity matrix of order as of $M_{[x,x+1]}$ matrix. Using the 3-state model (Figure 5.1), the multistate model of the 3x3, a matrix of transition rates was specified as:

\[
M_{[x,x+1]} = \begin{pmatrix}
M_{11}[x,x+1] & -M_{21}[x,x+1] & 0 \\
-M_{12}[x,x+1] & M_{22}[x,x+1] & 0 \\
-M_{13}[x,x+1] & -M_{23}[x,x+1] & 1
\end{pmatrix}
\]

where the suffix 1 stands for current smoker; 2 currently non-smoking and 3 is dead. The diagonal elements $M_{ii}[x,x+1]$ were calculated from the non-diagonal entries on the condition that the sum of entries over the destination status was zero.

The probability that a person was not a smoker at exact age $x$ was:

\[
l_{\text{nsm}}(x+1) = l_{\text{nsm}}(x) \times \left[1 - q_{\text{sm,nsm}}[x,x+1] - q_{\text{nsm,sm}}[x,x+1]\right] + l_{\text{nsm}}(x) \times q_{\text{sm,nsm}}[x,x+1]
\]

Similarly, the probability that he or she was a smoker was:

\[
l_{\text{sm}}(x+1) = l_{\text{sm}}(x) \times \left[1 - q_{\text{sm,nsm}}[x,x+1] - q_{\text{sm,sm}}[x,x+1]\right] + l_{\text{sm}}(x) \times q_{\text{sm,nsm}}[x,x+1].
\]
Assuming the transitions occurred in the middle of the age interval, the life table exposure was calculated using the formula,

\[ L_{i}[x, x+1) = 0.5 \times [L_i(x) + L_i(x+1)] \]

All life tables were constructed from age 10 on, with the last at age 95 being open-ended. For the oldest open age group (in this case age 95 and above), the following formula could be used:

\[ M_i(95+) = \frac{1}{e(95+)} \text{, for } i = 1, 2 \]

Both for males and females, we assumed that mortality rates after age 95 were constants irrespective of smoking status. We used the Massachusetts States Life Tables (Centers for Disease Control and Prevention, 1988-91). For males, \( e(95) = 2.92 \) years, for females \( e(95) = 3.40 \) years, and 3.29 years for the total population. Using the life table formula in stationary population, we found that \( M(95+) = 1/2.92 = 0.34 \) for males, and \( M(95+) = 1/3.30 = 0.29 \) for females. We also assumed no mortality prior to age 30.

The total number of person-years lived was calculated using the formula,

\[ T_{i}[x, x+1) = \sum_{i=10}^{95} L_i(t) \]

Life expectancy at age \( x \) was, \( e(x) = T_i[x, x+1) / l(x) \), where \( l(x) = l_{sm}(x) + l_{ns}(x) \).

### 5.3 Results

The aim of this chapter was to analyze the smoking career. For this purpose, we constructed a multistate smoking status life table. Before describing the results of the multistate smoking status life table, we described the different episodes of smoking and changes in smoking status over 48 years of follow-up of the FHS. We obtained the life expectancy, survival of the cohort, lifetime probability of quitting smoking and differences in the number of years lived by smoking status from the multistate smoking status life table. As smoking initiation, quitting, switching and health effects are different between males and females (Prescott et al., 1998; Marang-van de Mheen et al., 2001), we presented the results separately for males and females.
5.3.1 Changes in smoking status

The changes in smoking status of the FHS cohort at different points in time (after each 10 years of follow-up of those people who survived and reported their smoking status) are presented in Table 5.1. At study entry (1948-51), 86 percent of males and 43 percent of females were smokers. This prevalence was slightly higher compared to the smoking prevalence at the first exam reported in Chapter 4, as we excluded the respondents who reported that they had quit smoking before entry into FHS. As expected, the prevalence of smoking declined for both males and females over time. After 40 years of follow up, the prevalence of smoking fell from 86 percent to 9 percent for males and from 43 percent to 10 percent for females. Smoking prevalence declined more sharply among male smokers compared to females. Smoking prevalence was shown to decrease an average of 20 percent for males and nearly 10 percent for females per ten years of follow-up. During 48 years of follow-up, 83 percent of male participants and 72 percent of female participants died.

<table>
<thead>
<tr>
<th>Time of Follow-up</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>At first exam</td>
<td>86</td>
<td>43</td>
<td>61</td>
</tr>
<tr>
<td>Non-smoking</td>
<td>14</td>
<td>57</td>
<td>39</td>
</tr>
<tr>
<td>After 10 years of follow-up</td>
<td>72</td>
<td>39</td>
<td>52</td>
</tr>
<tr>
<td>Smoking</td>
<td>14</td>
<td>57</td>
<td>39</td>
</tr>
<tr>
<td>Non-smoking</td>
<td>28</td>
<td>61</td>
<td>48</td>
</tr>
<tr>
<td>After 20 years of follow up</td>
<td>49</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>Smoking</td>
<td>51</td>
<td>72</td>
<td>64</td>
</tr>
<tr>
<td>Non-smoking</td>
<td>28</td>
<td>61</td>
<td>48</td>
</tr>
<tr>
<td>After 30 years of follow up</td>
<td>30</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>Smoking</td>
<td>70</td>
<td>80</td>
<td>76</td>
</tr>
<tr>
<td>Non-smoking</td>
<td>70</td>
<td>80</td>
<td>76</td>
</tr>
<tr>
<td>After 40 years of follow up</td>
<td>9</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Smoking</td>
<td>91</td>
<td>91</td>
<td>90</td>
</tr>
<tr>
<td>Non-smoking</td>
<td>91</td>
<td>91</td>
<td>90</td>
</tr>
</tbody>
</table>

5.3.2 Smoking episodes

We assumed that at age 10, all participants (4694) were non-smokers. However, of the 2013 non-smoking males at age 10, 1759 (87%) had experienced at least one episode of smoking. Of the 2694 females, 48% had experienced at least one
episodes of smoking during follow-up. Among males, 13% (females 8%) had experienced more than two episodes of smoking. At a certain age, they initiated smoking, then quit only to take up the habit again. Among male smokers, at least 48 percent quit smoking. At least 31 percent of female smokers quit. Of all first-time quitters, both male and female, almost 26 percent restarted smoking.

Based on biannual reporting of smoking status over 48 years, the FHS cohort was found to have experienced total 10,485 episodes, of 6880 episodes (66 percent) of non-smoking and 3605 (34 percent) episodes of smoking. Males experienced 39 percent of smoking episodes and females experienced 30 percent of smoking episodes.

Table 5.2  Total number of smoking and non-smoking episodes based on biannual reporting of smoking status over 48 years of follow-up of FHS

<table>
<thead>
<tr>
<th></th>
<th>Non-smoking episodes (%)</th>
<th>Smoking episodes (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>6880 (66)</td>
<td>3605 (34)</td>
</tr>
<tr>
<td>Male</td>
<td>3203 (61)</td>
<td>2152 (39)</td>
</tr>
<tr>
<td>Female</td>
<td>3677 (70)</td>
<td>1548 (30)</td>
</tr>
</tbody>
</table>

5.3.3 Transition rates

The empirical age-specific transition rates by smoking status are shown in Figure 5.2. In the FHS, the mean (median) age of initiation of smoking was 16 (15) years for males and 18 (17) years for females. Overall, from age 30 to 70 the rate of transition from non-smoking to smoking remained stable. Most restarters were in this age band. Smokers started quitting at age 30; the number of quitters was shown to increase after age 50 and to remain high until age 74. Cessation of smoking subsequently leveled off after age 75.

At age 15 or 16, the rate of initiation of smoking was nearly 0.7 for males and 0.2 for females. Males tended to start smoking at early ages. Females started smoking between the ages of 12-20. The rates of restarters among males were twice that of females. The pattern of quitting smoking was nearly the same both for males and females. Starting at age 30, cessation levels increased from age 50 and continued until age 74 after which stabilization occurred.

---

1Age-specific transition rates are plotted for 5-year age band instead on one year age band to avoid the irritated behavior of transition rates
As expected, the age-specific death rates for male and female smokers were higher than for non-smokers. Compared to males, the differences in mortality rates by smoking status were higher among females. This could be why smoking has been shown to be more harmful to females than to male smokers (Marang-van de Mheen et al., 2001; Surgeon General, 2001; Dresler, 1998). The zero mortality rate prior to age 30 was a model input, as mortality follow-up did not begin until age at entry into observation (minimum age at entry into observation was 28).

Figure 5.2 Transition rates by smoking status

Male and female

![Transition rates by smoking status](image-url)
5.3.4 Survival probability

The survival probabilities or state occupancies of male and female smokers and non-smokers are presented in Figure 5.3. This survival curve demonstrates the survival proportion of a synthetic cohort with a particular smoking status. It pictures how smoking evolves over the life course. The lower portion represents the probability of being alive at that age and being a non-smoker. The upper portion represents probability of being alive and a smoker.

The survival probabilities may be used to determine interesting conditional probabilities. For instance, the probability of a 10-year old non-smoker being a non-smoker at age 25 was nearly 50 percent (25 percent for males and 70 percent for females). The survival probability in non-smoking state was the lowest at age 45 (43 percent). The probability of survival in a non-smoking state increased until age 75, after which it leveled off. Non-smokers were seen to survive longer and smokers to die prematurely.
Figure 5.3  Survival probability in a non-smoking state and survival probability in a smoking state of a cohort of 10-year-old males and females of FHS

Total population

Male
5.3.5 Life expectancy

We used the transition probabilities estimated from the FHS to calculate the expected number of years in each smoking status. The multistate life table translates the age-specific transition probabilities in dwelling times or number of years spent as a smoker and as a non-smoker. We estimated the expected dwelling times in each smoking status for reference ages 10, 30, 50 and 70. The FSH recorded the age at which respondents started smoking. Although 45 persons started smoking before or at age 10, they were assumed to have started at age 10, implying that everyone was a non-smoker at exact age 10. Table 5.3 shows the dwelling times.

The life expectancy at age 10 was 69.13 years (65.71 for males and 71.66 for females). Males smoked for considerably more years than females. A male of age 10 could expect to spend 41.13 years as a smoker, which was 63% of the expected lifetime. Females of the same age could expect 19.97 years as smokers, which was 28% of the remaining lifetime. At age 50, the picture was quite different. Males of that age could expect to spend 53% of the remaining lifetime as smokers (13.91 years of the expected total of 26.46 years). Fifty-year old females smoked 24% of the remaining years. This was an interesting observation. Although when younger, the share of remaining life in which an average male smoked was considerably larger than that of an average female, the difference was relatively small at age 50 and was further reduced at higher ages. The change may be attributed to two factors. The first is quitting behavior. Males were more likely to quit smoking than females and they quit at an earlier age (see Section 5.3.6). Second, and more importantly, is the selection factor. Males with a history of smoking died at a younger age than males who had never smoked. As a result, their share in the older male population was smaller.
### Table 5.3  Number of years that a person of a given age and sex may expect to spend in each smoking status, FHS

<table>
<thead>
<tr>
<th>Life expectancy at age</th>
<th>10</th>
<th>30</th>
<th>50</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total life expectancy</td>
<td>69.14</td>
<td>49.14</td>
<td>29.76</td>
<td>14.38</td>
</tr>
<tr>
<td>Years spent non-smoking</td>
<td>40.11</td>
<td>27.64</td>
<td>19.35</td>
<td>11.78</td>
</tr>
<tr>
<td>Years spent smoking</td>
<td>29.03</td>
<td>21.50</td>
<td>10.41</td>
<td>2.60</td>
</tr>
<tr>
<td>Difference</td>
<td>11.07</td>
<td>6.14</td>
<td>8.93</td>
<td>9.17</td>
</tr>
</tbody>
</table>

### 5.3.6 Lifetime probability of quitting smoking

One of the useful features of the life table technique is that it can be used to calculate the lifetime probability of an event. For example, we are aware that many smokers quit smoking. What is the probability that an individual, who was smoking at age 20, will quit smoking before age 40, 60 or death? The probabilities of quitting smoking at different ages are presented in Table 5.4. Smoking cessation has immediate and substantial health benefits, both symptomatically and pathophysiologically, and dramatically reduces the risk of most smoking-related diseases (Office of the U.S. Surgeon General, 1989).

For the synthetic cohort derived from transitions within the Framingham cohort, the probability of quitting smoking before age 50 was 10 percent. Before age 70 it was 37 percent and before death it was 46 percent. More than 50 percent of the smokers did not quit smoking during their lifetime. Those who quit were more likely to do so between the ages of 50 and 70 than at any other age. The lifetime probability of quitting smoking for males and females of the FHS participants was 58 percent and 37 percent, respectively. Before age 50, it was 12 percent and 8 percent, respectively. By the age of 70, nearly half of the men had quit smoking. By age 70, only 30 percent of the females had quit smoking. The percentage of females quitting smoking over the life course was 21 percent lower than males.
Table 5.4 Probability of quitting smoking

<table>
<thead>
<tr>
<th>Probability (%) of quitting smoking before age</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population</td>
<td>2</td>
<td>10</td>
<td>23</td>
<td>37</td>
<td>46</td>
</tr>
<tr>
<td>Males</td>
<td>3</td>
<td>12</td>
<td>29</td>
<td>47</td>
<td>58</td>
</tr>
<tr>
<td>Females</td>
<td>2</td>
<td>8</td>
<td>19</td>
<td>30</td>
<td>37</td>
</tr>
</tbody>
</table>

5.3.7 Person-years lived by smoking status

This study measured the life history of smoking and non-smoking in terms of the differences in the life-table person years lived in smoking and non-smoking states. The age-specific person years lived in the non-smoking status were subtracted from the age-specific person years lived in the smoking status (Figure 5.4). This approach allowed us to gauge the differences in the age-specific person-years lived by smoking status (for details, see Section 2.4.7). A negative difference (non-smoker-smoker) meant that an individual spent more time in the smoking state at that age interval. Overall, the synthetic FHS cohort spent more time in the smoking state in the age band 22 to 55 compared to the non-smoking state. Males between the ages of 15 and 65 spent more time in the smoking state than as non-smokers. However, throughout life, female participants spent more time in the non-smoking state than in the smoking state.

Figure 5.4 Differences of the person-years lived: non-smokers–smokers
5.4 Discussion

We have presented an analysis of the smoking career of a white American population, namely the original Framingham Heart Study cohort. We observed their smoking experiences throughout life and translated the smoking status into life years lived as a smoker and non-smoker within a synthetic cohort. Overall, the FHS cohort experienced longer life in a non-smoking state than in a smoking state. Male participants spent more time as smokers between the ages of 15 and 65 than as non-smokers. Females spent more time in a non-smoking state than smoking state. Using the FHS cohort, we found that less than half of smokers quit smoking during their lifetime.

The smoking career of the male participants was different from that of a female participant. Men usually initiated smoking at an early age (2 years earlier), which was consistent with the present generation. Among male smokers, at least 48 percent quit smoking. At least 31 percent of female smokers quit smoking. Of male non-smokers at age 10, only 1 in 5 remained a non-smoker and 4 in 5 initiated smoking before age 25. Similarly, among female non-smokers at age 10, about 2 in 3 remained non-smokers and 1 in 3 initiated smoking before age 25. Compared to men, women smokers were more persistent. The restarting rate was twice as high among males compared to females. The probability of relapse was 26 percent for both males and females.

The important findings of this chapter are the life course indicators: life expectancy in a non-smoking state and life expectancy in a smoking state, the lifetime probability of quitting, and the differences in person-years lived by smoking status. At age 10, the total life expectancy of the FHS cohort was 66 years for males and 72 years for females. Males spent two-thirds of their expected lifetime (68%) smoking. This was less than one third for females (28%). The difference declined at higher ages when males with a history or smoking either quit or die. A female expected to survive more years as a non-smoker than in the smoking state.

The lifetime probability of quitting smoking for males and females of the FHS participants was 58 percent and 37 percent, respectively. Over the life course, the percentage of females that quit smoking was 21 percent less compared to males. More than 50% of smokers did not quit smoking in lifetime; individuals who quit were more likely to be between the ages of 50 and 70. Females were more persistent smokers, which could explain why smoking is more harmful to female smokers. There could be other possible causes, such as smoking intensity in terms of number of cigarettes smoked or hormonal changes. However, evidence shows that female smokers are at much greater risk of lung cancer than men (Risch et al., 1993; Dresler, 1998).

The strength of the present chapter lies in its focus on a well-documented epidemiologically defined community-based population and a 48-year prospective
consistent follow-up of the same cohort. Such a long-term monitoring of a large cohort enabled us to investigate the initiations, patterns of quitting, restarting and to characterize the individual smoking career. However, these findings are highly period and place specific. They are limited to the American population that started smoking in first half of the last century. Another strength of this chapter was the novel approach to analyzing the smoking career. Our smoking status definition included the up-dated information on smoking status rather than the status at a specific point in time i.e. we used incidence of smoking. We defined smoking status explicitly.

The importance of the described method is at least threefold.

Firstly, a knowledge of the historical smoking and quitting patterns is important for the interest taken by current studies in the cumulative smoking histories of current populations. The smoking career of today’s generation will differ from that of the generation of the mid-1950s. For instance, prevalence of smoking was highest after World War 2, which might reflect the quitting effect. Female prevalence and consumption of smoking increased after the 1980s and women started smoking more like men (Doll et al., 1994). A recent report by the Surgeon General shows that prevalence of smoking is increasing among teenage girls compared to boys (Surgeon General, 2001). Today’s persistent smokers may well have smoked a substantial number of cigarettes throughout adult life, whereas few of the smokers during the period 1950s can have done so (Peto et al., 2000). Therefore the smoking career of the present generation will be different than the past generation. However, lifetime smoking exposure of older generations now is based on these historical measures, so it also has current relevance.

Secondly, several life course indicators, such as expected years lived in a non-smoking and a smoking state, lifetime probability of quitting smoking and the differences in the time spent by smoking status are the new indicators among the public health community.

Thirdly, this method can be used for other risk factors to indicate current risk factor life courses.

Our novel approach demonstrates one way of explicitly taking into account the changes in risk factors throughout life. However, our approach is data demanding. For the life course analysis of other risk factors, for instance to construct the cholesterol career, hypertension career, obesity career, diabetes career and so on, epidemiologists and medical demographers can use this method in more innovative ways than any other methods. Since smoking is the major preventable cause of premature death, we need further insights into the smoking career of the present generation. To do so, our approach could be a standard one.
References


International Agency for Research on Cancer monographs on the evaluation of the carcinogenic risk of chemicals to humans Lyons, IARC, 1986 (NO. 38)


Life history of cardiovascular disease
A comparison of smokers and non-smokers

Abstract

Smokers combine an increased risk of a number of diseases with an increased risk of death. The duration of morbidity depends on that balance. This study compares the burden of cardiovascular disease in terms of lifetime risk and life years lived with disease between smokers and non-smokers. We constructed multistate life tables describing transitions through various cardiovascular diseases for 4723 smokers and non-smokers observed during 20 biannual observations in the Framingham Heart Study. The risk of developing any cardiovascular disease before age 70 is higher among smokers. Associated with their longer life expectancy, male non-smokers have higher lifetime risks of coronary heart disease, myocardial infarction, stroke and congestive heart failure, while female non-smokers have higher lifetime risks of coronary heart disease and congestive heart failure. Non-smokers live 8.66 years (95% CI 7.61- 9.63) (men) and 7.59 (95% CI 6.33- 8.92) (women) years longer than smokers but also spend more years with cardiovascular disease: 2.43 (95% CI 1.72-3.16) years for males and 2.66 (95% CI 1.87-3.38) years for females. Non-smokers live more years free of cardiovascular disease: 6.22 (95% CI 5.09-7.30) years for males and 4.93 (95% CI 3.54-6.29) for females. Not smoking will not eliminate cardiovascular disease, but it will postpone it to older ages. Smoking, by shortening life, decreases the years lived with cardiovascular disease throughout the life course. Paradoxically, in a non-smoking society, more people will live with cardiovascular disease but this will be concentrated at the end of life.
6.1 Introduction

In this chapter, we investigated cardiovascular disease in the life course of smokers and non-smokers and addressed the compression of morbidity hypotheses. The thesis of compression of morbidity, as put forward by Fries (Fries, 1980), suggested that lifestyle modification might decrease morbidity. The main propositions of Fries were based on the observations that the length of life is limited and that chronic disease can be postponed. He pointed to the decline of tobacco consumption as an example. If incidence (inflow) decreases and mortality (outflow) remains constant, the prevalence (stock) will decrease. However, smoking has been identified as the major cause of death among smokers. The age adjusted disease incidence of non-smokers is lower, but because of a lower mortality they live a longer period at risk. Recent papers have suggested that smokers experience both more (Nusselder et al., 2000; Bronnum-Hansen and Juel, 2001) and fewer (Ferrucci et al., 1999; Martel et al., 2000) years lived with disability than non-smokers, and that they cost more to health care (Barendregt et al, 1997) over the life course. While the reason(s) for these differences is not yet clear, probable significant factors include differences in smoking status definitions, start ages, assumptions regarding the relationship between smoking relative risks and age, model types and disability definitions. The potential age dependence of the relative risks associated with smoking are probably the most crucial.

Cardiovascular diseases are common both among smokers and non-smokers, and increase sharply with age. The interplay between different incidence rates, mortality rates and age structures between smokers and non-smokers cannot be gauged intuitively (Lloyd-Jones, et al., 1999). The question addressed in this chapter is whether non-smoking will shorten the time lived with cardiovascular disease, given the competing forces of an increased risk of cardiovascular disease and increased mortality in smokers.

Section 6.2 discusses data source, smoking status definition, estimation of hazard ratios and the construction of multistate life table. Results are described in Section 6.3. First, we described the hazard ratios of disease occurrences or deaths in sub-section 6.3.1. Second, the life table outcomes are illustrated in sub-section 6.3.2. The chapter is concluded with a discussion in Section 6.4.

6.2 Data and Methods

We took advantage of the data available from the Framingham Heart Study, which gave us the unique historical documentation of fifty years of cardiovascular health damage due to smoking. We compared the cardiovascular life histories of smokers and non-smokers by constructing multistate life tables from the first 40 years of follow-up from the original Framingham Heart Study cohort. This enabled a
comparison to be made of the life years lived with and without cardiovascular disease and lifetime risk of cardiovascular disease for smokers and non-smokers.

Data source
The original Framingham Heart Study cohort consisted of 5209 respondents (45% male) from a random sample of adults aged 28 through 62 years residing in Framingham, Massachusetts between 1948 and 1951. The participants were tracked by standardised biennial cardiovascular examinations, daily surveillance of hospital admissions, death information and information from physicians and other sources outside the clinic, ensuring highly accurate follow-up of death and clinically presenting cardiovascular disease. In the Framingham Heart Study current smoking status (yes or no) was recorded at all but 4 of the 20 biennial exams. Missing values for this variable were recorded at each exam.

For this chapter, we used the data on age at the onset of cardiovascular disease or death over forty years of follow-up (exam rounds 1 to 21) for the 4723 participants without cardiovascular disease at study entry. Some 139 respondents had cardiovascular disease at study entry. The smoking status of 13 respondents was unknown during the follow-up; for another 334 respondents, only one observation of smoking status was available. These (139+13+334=486) were excluded. Smoking status information was missing on 13% of the 4723 respondents for more than half of the rounds (excluding the four rounds with no information at all).

Definition of smoking status
Smoking status for each participant was allocated based on the current smoking status recorded at each available exam between study entry and death or study exit. We classified never smokers as those with all available smoking records coded as a non-smoker and always smokers as those with all available smoking records coded as a smoker. Ever smokers were the rest of the participants, characterized by a mixture of smoking and non-smoking throughout the study.

Of the 4723 participants included, 381 (18%) men were classified as never smokers and 674 (33%) as always smokers. 1384 (52%) women were found to be never smokers and 448 (17%) were classified as always smokers. The remaining 1018 (49%) men and 818 (31%) women were classified in the residual group of ever smokers.

Estimation of hazard ratios
The risk of cardiovascular disease or death for smokers or ever smokers relative to never smokers was calculated using Cox proportional hazards analysis. The events considered were death, and onset of: all cardiovascular disease (CVD); all coronary heart disease; acute myocardial infarction; stroke, and congestive heart failure
(Shurtleff, 1971). Coronary heart disease included angina pectoris, coronary insufficiency, myocardial infarction and sudden death. Cardiovascular disease included all coronary heart diseases, all cerebrovascular diseases (including stroke), intermittent claudication and congestive heart failure.

Analyses were performed separately for each sex with an adjustment for the mean values of educational levels. Educational status has an influence on smoking status and on cardiovascular disease and mortality. Educational status is the main socioeconomic indicator of the FHS cohorts. We treated it as a confounder. The information on educational status was collected at the first exam. The educational level was grouped into six categories: 1 8th grade or less; 2 high school but not a graduate; 3 high school graduate; 4 college but not a graduate; 5 college graduate; and 6 post graduate or business college or nursing school. Age at transition was the time variable in this model. We have presented mean hazard ratios.

**Multistate life tables**
Separate multistate life tables were created to analyze annual transitions through each of the cardiovascular disease types described above. The basic multistate life table structure had the state space \{NO-CVD, history of CVD, dead\} (Section 3.3.2, Figure 3.1(b)), CVD represents one of the specific CVD states (cardiovascular disease, coronary heart disease, acute myocardial infarction, stroke, or congestive heart failure). For example, in the life table for all CVD, the possible transitions were, “NO-CVD” to “death”, “NO-CVD” to “history of CVD”, and “history of CVD” to “death”.

The empirical transition rate was calculated for each single year of age by dividing the number of events between exact ages by the corresponding risk period of exposure in each state. Each set of rates was calculated separately for male never smokers, male always smokers, female never smokers and female always smokers. The method used to estimate the observed occurrence-exposure was the same as those described in Section 3.3.3 in Chapter 3. For smoothing empirical age-specific transition rates, we applied Gompertz regression models (Gompertz, 1825). In this model the transition rates depend on age alone. For life table construction, the rates were converted to probabilities, by assuming that within each single year age interval the hazard remained constant and taking into account the competition between risks (Schoen, 1988). All life tables were constructed from age 50 and closed at age 90 using the Massachusetts life expectancy at age 90 for 1989-91 (males 3.93 years, females 4.76 years, total population 4.55 years) (Centers for Disease Control and Prevention, 1988-91). The procedure used to construct the life tables was the same as the procedure described and applied in Chapter 3 (Section 3.3.4).

The measures directly available from the multistate life tables are life expectancy with and without disease and the lifetime risk of an event over a certain period (inclusive of lifetime). Confidence intervals were derived using a non-
parametric bootstrap procedure, based on 2,000 replicates, in S-plus 2000 (MathSoft Inc., Washington, USA). We have reported the bootstrap bias-corrected, adjusted 95% confidence intervals (Efron and Tibshirani, 1993). The basic steps to calculate the confidence intervals of the outcomes of multistate life table were discussed in Section 2.4.5 of Chapter 2 of this study. An example of an S-plus script is presented in Technical Appendix 6.1

6.3 Results

The results are presented in two parts. First, we have described the risk of cardiovascular disease or death for smokers or ever smokers relative to never smokers calculated using Cox proportional hazards analysis. Second, we have described the life table outcomes: survival probability, lifetime probability of disease, differences in the number of years spent with disease and life expectancy by smoking status.

6.3.1 Disease incidence and death

The risk of death or cardiovascular disease incidence was analysed for ever and always smokers relative to never smokers using proportional hazards regression. As expected, always smokers were found to have an increased risk of all cardiovascular disease sub-types examined, ranging from a hazard ratio of 1.29 [95% CI 1.04-1.60] for coronary heart disease (females) to 2.00 [95% CI 1.38-2.91] for stroke (males) (Table 6.1). Always smokers also had a significantly higher risk of dying once they had cardiovascular disease compared to never smokers (Table 6.2), with relative risks ranging from 1.28 [95% CI 0.99-1.66] from myocardial infarction (males) to 2.23 [95% CI 1.85-2.68] from cardiovascular disease (females). The risk of death among smokers without cardiovascular disease was significantly higher compared to never smokers (Table 6.2). Note that people without cardiovascular disease constitute a mixed group, including healthy people and people with cancer, and other diseases.
Table 6.1  Risk of cardiovascular disease (including sudden death) by smoking status, relative to never smokers (95% confidence intervals in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>CHF</th>
<th>Stroke</th>
<th>Total number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.15</td>
<td>1.04</td>
<td>0.98</td>
<td>1.05</td>
<td>1.45</td>
<td>381</td>
</tr>
<tr>
<td></td>
<td>(0.98-1.34)</td>
<td>(0.87-1.24)</td>
<td>(0.79-1.22)</td>
<td>(0.77-1.43)</td>
<td>(1.03-2.08)</td>
<td></td>
</tr>
<tr>
<td>Always smoker</td>
<td>1.65</td>
<td>1.41</td>
<td>1.38</td>
<td>1.64</td>
<td>2.00</td>
<td>674</td>
</tr>
<tr>
<td></td>
<td>(1.39-1.94)</td>
<td>(1.17-1.71)</td>
<td>(1.09-1.74)</td>
<td>(1.17-2.28)</td>
<td>(1.37-2.91)</td>
<td></td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.04</td>
<td>0.97</td>
<td>1.24</td>
<td>1.32</td>
<td>1.05</td>
<td>1384</td>
</tr>
<tr>
<td></td>
<td>(0.91-1.19)</td>
<td>(0.81-1.15)</td>
<td>(0.97-1.59)</td>
<td>(1.03-1.72)</td>
<td>(0.83-1.35)</td>
<td></td>
</tr>
<tr>
<td>Always smoker</td>
<td>1.42</td>
<td>1.29</td>
<td>2.06</td>
<td>1.62</td>
<td>1.75</td>
<td>448</td>
</tr>
<tr>
<td></td>
<td>(1.21-1.67)</td>
<td>(1.05-1.59)</td>
<td>(1.55-2.74)</td>
<td>(1.62-2.24)</td>
<td>(1.31-2.34)</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2  Risk of death (from given states) by smoking status, relative to never smokers (95% confidence intervals in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Non-CVD</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>CHF</th>
<th>Stroke</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever smoker</td>
<td>0.92</td>
<td>0.81</td>
<td>0.78</td>
<td>0.71</td>
<td>1.14</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>(0.71-1.21)</td>
<td>(0.68-0.98)</td>
<td>(0.63-0.96)</td>
<td>(0.55-0.91)</td>
<td>(0.81-1.60)</td>
<td>(0.72-1.58)</td>
</tr>
<tr>
<td>Always smoker</td>
<td>2.51</td>
<td>1.62</td>
<td>1.55</td>
<td>1.28</td>
<td>1.64</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>(1.93-3.27)</td>
<td>(1.34-1.95)</td>
<td>(1.26-1.91)</td>
<td>(0.99-1.66)</td>
<td>(1.15-2.34)</td>
<td>(1.32-3.02)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.05</td>
<td>0.94</td>
<td>1.05</td>
<td>1.20</td>
<td>1.01</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>(0.86-1.29)</td>
<td>(0.79-1.12)</td>
<td>(0.84-1.31)</td>
<td>(0.88-1.65)</td>
<td>(0.76-1.34)</td>
<td>(0.68-1.25)</td>
</tr>
<tr>
<td>Always smoker</td>
<td>2.31</td>
<td>2.23</td>
<td>2.13</td>
<td>1.96</td>
<td>1.48</td>
<td>2.08</td>
</tr>
<tr>
<td></td>
<td>(1.86-2.87)</td>
<td>(1.85-2.68)</td>
<td>(1.67-2.72)</td>
<td>(1.40-2.75)</td>
<td>(1.05-2.10)</td>
<td>(1.49-2.90)</td>
</tr>
</tbody>
</table>

Ever smoking was not significantly associated with an increased risk of mortality or of any of the cardiovascular disease sub-types examined. This is most likely due to the heterogeneity in smoking intensities (duration and time since quitting) in this group. To make comparisons between smokers and non-smokers we used the more homogeneous groups of never and always smokers for all further analyses.

6.3.2  Life table outcomes

Survival probability

The dynamics of differential incidence of cardiovascular disease and death over increasing age cannot be determined intuitively. Therefore, the age-specific rates of
transition were combined using multistate life tables. Figure 6.1 shows the survival free of cardiovascular disease, coronary heart disease, acute myocardial infarction, stroke or death of fifty-year-old smokers and never-smokers. While never smokers have a much larger gap between total survival and disease-free survival, overall they lead a longer life and survive longer free of cardiovascular disease than always smokers. The mortality incidence differences between the different smoking groups lead to large differences in early mortality and morbidity. Among males, 46% of always smokers and 15% of never smokers die before the age of 70. Among females, these figures are 30% of always smokers but only 10% of never smokers. While an average six in ten male never smokers who were free of cardiovascular disease at age 50 will still be alive and free of cardiovascular disease twenty years later, only four in ten always smokers will remain in this state (Figure 6.1). Of females free of cardiovascular disease at age 50, only 6 in 10 always smokers compared to 7 in 10 never smokers will be alive and free of cardiovascular disease at age 70.

Figure 6.1 Survival curves illustrating the probability of surviving and surviving free of cardiovascular disease (CVD), coronary heart disease (CHD), myocardial infarction (MI), or stroke

Male never smokers

![Graph showing survival curves for different conditions]
Male smokers

Female never smokers
Lifetime probability of disease
The probability of developing cardiovascular disease before age 70 or deaths among person free of CVD at age 50 is presented in Table 6.3. As can be seen in this table, always smokers experience a significantly greater risk than never smokers for developing every cardiovascular sub-type examined before the age of 70. However, over a lifetime never-smokers have approximately the same risk of cardiovascular disease as always smokers, simply because they live longer (Table 6.3). For males, while never smokers tend to have a higher lifetime risk of any coronary heart disease, myocardial infarction, stroke and congestive heart failure than always smokers, this is only statistically significant at the 5% level for myocardial infarction.
Life expectancy

Concordant with the often-higher lifetime risk of disease and the lower mortality rates post disease, the duration of disease is generally longer among never smokers (Table 6.4). Male never smokers live 2.43 [95% CI 1.72-3.16] years longer with a history of cardiovascular disease, 2.01 [95% CI 1.43-2.66] years longer with a history of coronary heart disease, 1.13 [95% CI 0.60-1.68] years longer with a history of myocardial infarction, 0.68 [95% CI 0.33-0.96] years longer with a history of stroke and 0.39 [95% CI 0.13-0.65] years longer with heart failure. Female neversmokers live 2.66 [95% CI 1.87-3.38] years longer with a history of cardiovascular disease, 1.82 [95% CI 1.26-2.47] years longer with a history of coronary heart disease, and 0.49 [95% CI 0.11-0.83] years longer with a history of stroke, but show no significant difference in time spent with a history of myocardial infarction or heart failure.
Of course, most impressive is the change in life expectancy free of any cardiovascular disease: 50 year-old always smokers live five to six years less free of cardiovascular disease than never smokers (men 6.22 [95% CI 5.09-7.30] years and women 4.93 [95% CI 3.54-6.29] years). Both male and female always smokers live significantly fewer years free of myocardial infarction, stroke and congestive heart failure than never smokers. The combination of the extra years lived with and without cardiovascular disease leads to a difference in total life expectancy at age 50 between never smokers and always smokers of around 8 years (men 8.66 [95% CI 7.61-9.63], women 7.59 [95% CI 6.32-8.92], Table 6.4).

Table 6.4 The burden of cardiovascular disease in always smokers versus never smokers free of any cardiovascular disease at age 50. Life years lived with a history of cardiovascular disease. 95% confidence intervals are presented in parentheses

<table>
<thead>
<tr>
<th>Males</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
<th>Total LE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never smokers</td>
<td>7.25</td>
<td>5.36</td>
<td>3.20</td>
<td>1.33</td>
<td>1.00</td>
<td>30.42</td>
</tr>
<tr>
<td>smokers</td>
<td>(6.77-7.81)</td>
<td>(4.92-5.75)</td>
<td>(2.85-3.58)</td>
<td>(1.10-1.57)</td>
<td>(0.82-1.19)</td>
<td>(29.79-30.92)</td>
</tr>
<tr>
<td>Always smokers</td>
<td>4.81</td>
<td>3.34</td>
<td>2.07</td>
<td>0.64</td>
<td>0.61</td>
<td>21.77</td>
</tr>
<tr>
<td>smokers</td>
<td>(4.34-5.35)</td>
<td>(2.87-3.75)</td>
<td>(1.73-2.51)</td>
<td>(0.49-0.90)</td>
<td>(0.46-0.78)</td>
<td>(20.96-22.66)</td>
</tr>
<tr>
<td>Difference (never-always)</td>
<td>2.43</td>
<td>2.01</td>
<td>1.13</td>
<td>0.68</td>
<td>0.39</td>
<td>8.66</td>
</tr>
<tr>
<td>(never-always)</td>
<td>(1.72-3.16)</td>
<td>(1.43-2.66)</td>
<td>(0.60-1.68)</td>
<td>(0.33-0.96)</td>
<td>(0.13-0.65)</td>
<td>(7.61-9.63)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Females</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
<th>Total LE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never smokers</td>
<td>6.23</td>
<td>4.05</td>
<td>1.30</td>
<td>1.36</td>
<td>1.06</td>
<td>34.14</td>
</tr>
<tr>
<td>smokers</td>
<td>(5.83-6.68)</td>
<td>(3.71-4.41)</td>
<td>(1.12-1.51)</td>
<td>(1.18-1.60)</td>
<td>(0.90-1.23)</td>
<td>(33.68-34.63)</td>
</tr>
<tr>
<td>Always smokers</td>
<td>3.57</td>
<td>2.23</td>
<td>1.07</td>
<td>0.87</td>
<td>0.71</td>
<td>26.55</td>
</tr>
<tr>
<td>smokers</td>
<td>(3.04-4.34)</td>
<td>(1.74-2.76)</td>
<td>(0.78-1.49)</td>
<td>(0.59-1.20)</td>
<td>(0.48-1.27)</td>
<td>(25.35-27.69)</td>
</tr>
<tr>
<td>Difference (never-always)</td>
<td>2.66</td>
<td>1.82</td>
<td>0.23</td>
<td>0.49</td>
<td>0.35</td>
<td>7.59</td>
</tr>
<tr>
<td>(never-always)</td>
<td>(1.87-3.38)</td>
<td>(1.26-2.47)</td>
<td>(0.17-0.61)</td>
<td>(0.11-0.83)</td>
<td>(0.19-0.63)</td>
<td>(6.33-8.92)</td>
</tr>
</tbody>
</table>

**Difference of the years spent with disease**

The difference in life years lived with cardiovascular disease and its various subtypes by age is shown in Figure 6.2. The estimation procedure is described in Section 2.4.7 of Chapter 2. Importantly, the life years lost to cardiovascular disease for never smokers fall predominantly late in life, with always smokers living more years with cardiovascular disease throughout middle age (Figure 6.2). Non-smokers live longer with cardiovascular disease, but predominantly at older ages.
Figure 6.2 Difference in life years lived with cardiovascular disease, coronary heart disease, myocardial infarction, stroke, or congestive heart failure, by age (smokers – never smokers)

Males

Females
6.4 Discussion

This chapter shows the cardiovascular life course of the Framingham Heart Study cohort by smoking status. It was seen that non-smoking expanded both the duration of life free of cardiovascular disease and the duration of cardiovascular disease. Over a life course, non-smokers lived longer with cardiovascular disease and its manifestations such as cerebrovascular and coronary disease. We found that male and female never smokers lived 2.43 and 2.66 years, respectively, more with a history of cardiovascular disease than always smokers. Importantly, never smokers also lived more years free of cardiovascular disease: 6.22 years for males and 4.93 years for females. We demonstrated that this was the result of a complex combination of (i) the increased rate of cardiovascular disease, (ii) the increased rate of mortality from the cardiovascular disease state and (iii) the higher rate of mortality from the non-cardiovascular disease state to death associated with smoking.

While the increased incidence and mortality rates together led to fewer years lived with each form of cardiovascular disease for always smokers throughout the life course, this was seen to result from a combination of more years lived with disease at younger ages and less at older ages. Similarly, the risk of cardiovascular disease before age 70 was consistently greater for always smokers, in both males and females. Six out of ten never smoking males and seven out of ten never smoking females could expect to reach age 70 alive and free of any cardiovascular disease. This was strikingly less for smokers: four out of ten males and six out of ten females.

Previous analysis suggested that smokers lived longer with coronary heart disease than non-smokers (Barendregt, 1998). However, our empirical study took into account the increased risk of mortality after disease incidence due to smoking and the age dependence of the relative risks associated with smoking.

The main power of this study is in the Framingham Heart Study, which gave us 40 years of follow-up of a single cohort. All transition rates were estimated from this study, and treated with the time-honored multistate life-table method (Schoen, 1988; Rogers, 1995). Such life-tables are intuitively simple and appealing because they transparently describe the disease epidemiology as a life course. The transition rates at every age are the result of a mixture of both the broad age range at entry and the long follow-up, and they may come from very different periods. However, as long as the dynamics between “always smokers” and “never smokers” are not very different in these periods, this should not bias the results presented here.

Correct estimation of the transition rates is crucial. Testing the modifying effect of potential confounders, such as obesity, blood pressure and cholesterol level, increased the numbers of missing values, and selected a healthier subgroup
with more complete observations. As this may introduce selection bias, we have presented the life table outcomes without further adjustment.

Here, we have chosen our case definitions to maximize transparency, power and homogeneity. If, instead, smoking is used as a time varying covariate, the same individual is able to participate in both cohorts, non-smoking and smoking. However, the timing of incidence and mortality after starting and quitting smoking is different, which makes interpretation of the synthetic cohorts so constructed difficult. Smoking status at baseline yields cohorts that are too heterogeneous, including future quitters as smokers and future starters as non-smokers. We therefore described the life course for “always smokers” and “never smokers”, defined as smoking or non-smoking for 100% of the recorded exams in the Framingham cohort.

Life course analysis translates transition rates into dwelling times, and gives more transparent information about the consequences of risk. The risks of smoking translated into the loss of 8.66 (men) and 7.59 (women) life years. This is in the same order of magnitude as found in many other studies (Bronnum-Hansen et al., 2001; Martel et al., 2000; Rogers and Powell-Griner, 1991), but this study adds the consequences of smoking for cardiovascular disease in terms of incidence and duration throughout the life course. Male never smokers appear to have a greater lifetime risk of all forms of cardiovascular disease, including those considered most severe such as congestive heart failure and stroke. However, female never smokers show no difference in the lifetime risk of myocardial infarction, congestive heart failure or stroke compared to always smokers.

It appears that in smokers the higher outflow through mortality more than compensates the higher inflow through cardiovascular disease incidence: smoking leads to a shorter life expectancy with cardiovascular disease by increasing mortality from all causes. However, it would be possible to estimate health-adjusted life expectancy (combining life years free of CVD and life years with CVD) using “weigh” years lived with CVD against years lived without CVD. From this calculation it may be possible to conclude that smoking results in less health-adjusted life expectancy than non-smoking, as one might expect. In public health terms, the effects of smoking on cardiovascular disease are an expression of the folk wisdom "those who don't want to die old and disabled, are advised to die young".

The increased mortality among smokers will remain one of the primary challenges to public health as long as smoking is not abolished. But this study shows that public policy should not conclude simplistically that morbidity would decrease if people stop smoking. Incidence of cardiovascular morbidity will be postponed, years lived free of cardiovascular disease will be gained, but incidence as well as the life years lived with cardiovascular disease will be increased in the older population of never smokers. The central paradox of health and health care is
caused by ageing: the more successful we are in postponing the onset of, and the mortality associated with, age-related diseases with high case fatality, such as cardiovascular disease, the longer we live, the more we suffer from the disease at older age, and the more we need care. If, through that care, we successfully palliate morbidity, it is money well spent.

References


Technical Appendix 6.1*

S-Plus code to construct multistate life table along with confidence intervals of the life table outcomes using non-parametric bootstrapping.

lowage<-25 # lowest age (LA)
hiage<-105 # highest age (HA)
hiind<-hiage-lowage+1 # difference between 105-25 and plus 1
Ages<-lowage:hiage #25, 26,27, 105

ltlowage<-50 # life table (LT) lowest age
lthiage<-90 # LT highest age
lthiind<-lthiage-ltlowage+1 # difference between LT lowage & LT hiage
ltAges<-ltlowage:lthiage # LT ages: 50, 51,52,53, ....90

lowoffset<-ltlowage-lowage+1 #lowest offset, needed to estimate smoothed rates
hioffset<-lthiage-lowage+1 #highest offset

uitvoer<-matrix(data=0,nrow=lthiind,ncol=6) # define a matrix of lthiind by 6
lx<-vector(mode="numeric",length=lthiind) # define lx as a vector & length=lthiind
Lx<-vector(mode="numeric",length=lthiind) # define dx as a vector & length=lthiind

pxV<-vector(mode="numeric",length=lthiind) # define pX as a vector & length=lthiind
LxV<-vector(mode="numeric",length=lthiind) # define LxV as a vector & length=lthiind

H2Drate<-vector(mode="numeric",length=lthiind) # define H2Drate
as a vector & length=lthiind
H2Crate<-vector(mode="numeric",length=lthiind) # define H2Crate as a vector & length=lthiind
C2Drate<-vector(mode="numeric",length=lthiind) # define C2Drate
as a vector & length=lthiind
MaleBool<-logical(length=1) # define length 1 for last age group
Maleex90<-3.93 # life expectancy for last age group (male)
Femaleex90<-4.76 # life expectancy for last age group (female)

PY<-function(x) # a function to calculate person years
{
  for(i in 1:(lthiind-1))
  if ((x[i] > 0) & (x[i+1]>0) & (x[i]!=x[i+1])) # if non-negative and not equal to the previous value
    x[i] <- ((x[i]-x[i+1])/(-log(x[i+1]/x[i]))) # exponential estimate
  else x[i]<-0.5*(x[i]+x[i+1])+0.0000001 # linear estimate (an approximation)
  }

DoPY<-function(tmat,tmp) # a function to estimate exact age specific exposure
{
  lowb<-trunc(as.double(tmat[1]))
  hibb<-trunc(as.double(tmat[2]))
  lowb<-min(lowb,hiage)
  hibb<-min(hibb,hiage)
  tmp[lowb-lowage+1]<-tmp[lowb-lowage+1]+

---

* This S-plus code was written by Jan Barendregt, Public Health Department, Erasmus University, Rotterdam. The short description was noted by Abdullah Al Mamun, Population Research Centre, University of Groningen.
max((1.0-as.double(tmat[1])-lowb)*min(1,hihb-lowb),
min(as.double(tmat[2])-as.double(tmat[1]),lowb+1-as.double(tmat[1])))
tmp[(lowb-lowage+2):(hihb-lowage)]<-
  tmp[(lowb-lowage+2):(hihb-lowage)]+min(1,max(0,hihb-lowb))
tmp[hihb-lowage+1]<-tmp[hihb-lowage+1]+
    (as.double(tmat[2])-hihb)*min(1,hihb-lowb)
tmp
}

PYExAges<-function(tmat)
{
  tmp<-vector(mode="numeric",length=hiind)
  tmp<-apply(tmat,1,DoPY,tmp)
  tmp<-apply(tmp,1,sum)
  for (i in 1:hiind) if (tmp[i]==0) tmp[i]<-0.0001
  tmp
}

LifeTabsub<-function(Deaths,EntryCVD,H2Drate,H2Crate,C2Drate,MaleBool)
#
# a function to construct LT
#
{  
  lx[1]<-100000 # initial value of lx (e.g. at age 50 everybody is healthy)
  lxCVD[1]<-0 # initial value of lxCVD (e.g. nobody with CVD at age 50)
  pH<-1-exp(-H2Drate-H2Crate) # transition probability to leave healthy state
  pCVD<-1-exp(-C2Drate) # transition probability to leave CVD state
  for(i in 2:lthiind)
  {
    dx[i-1]<-lx[i-1]*pH[i-1] # of transitions from healthy states
    lx[i]<-lx[i-1]-dx[i-1]
    pH2CVD<-pH[i-1]*(H2Crate[i-1])/(H2Crate[i-1]+H2Drate[i-1]) # transition probability healthy to CVD
    dxCVD[i-1]<-lxCVD[i-1]*pCVD[i-1] # of transitions from CVD to death
    lxCVD[i]<-lxCVD[i-1]-dxCVD[i-1]+pH2CVD*lx[i-1]
  }
  Lx<-PY(lx)
  LxCVD<-PY(lxCVD)
  if (MaleBool)
  {
    tmp2<-(lx[lthiind]-dx[lthiind])*(Maleex90-1)
    tmp3<-(lxCVD[lthiind]-dxCVD[lthiind])*(Maleex90-1)
  }
  else
  {
    tmp2<-(lx[lthiind]-dx[lthiind])*(Femaleex90-1)
    tmp3<-(lxCVD[lthiind]-dxCVD[lthiind])*(Femaleex90-1)
  }
  i<-lthiind
  while(i>0)
  {
    tmp2<-tmp2+Lx[i] # calculate Tx for no-CVD state
    uitvoer[i,1]<-tmp2/(lx[i]+lxCVD[i]) # calculate ex for no-CVD state
    tmp3<-tmp3+LxCVD[i] # calculate TxCVD
    uitvoer[i,2]<-tmp3/(lx[i]+lxCVD[i]) # calculate exCVD
    uitvoer[i,3]<-uitvoer[i,1]+uitvoer[i,2] # Total life expectancy
    i<-i-1
  }
  uitvoer[,4]<-(Lx*H2Crate)/lx[1]
  uitvoer[,5]<-lx
}
uitvoer[6]<-LxCVD
# uitvoer     # if the survival curves are needed
LTrisk<-sum(uitvoer[1:lthiind,4])     # lifetime risk
risk70<-sum(uitvoer[1:20,4])         # lifetime risk until age 70

ttr<-c(HE=uitvoer[1,1],DE=uitvoer[1,2],LE=uitvoer[1,3],LTrisk=LTrisk,Riskb70=risk70)
}

LifeTab<-function(lrauw)
# LT function by sex and smoking status
{
  MaleBool<-(lrauw[1,2]==1)
  Deaths<-table(factor(trunc(lrauw$YDTHE-lrauw$YBIRTH),levels=Ages),
                 na.include(factor(cut(lrauw$YCVDI-lrauw$YBIRTH,breaks=c(45,100)))))
  # Numbers of deaths by age, Deaths[,1] are the CVD deaths, Deaths[,2] are the NonCVD deaths
  # YDTHE is year of death; YBIRTH is year of birth; YCVDI is year of CVD

  EntryCVD<-table(factor(trunc(lrauw$YCVDI-lrauw$YBIRTH),levels=Ages))
  # Numbers of incident cases by age

  tmat<-matrix(nrow=nrow(lrauw),ncol=2)  # A dataframe with entries by age in col 1 and exits by age in col 2

  tmat[,1]<-lrauw$YENTRY-lrauw$YBIRTH    # YENTRY is year of 1st entry into the follow-up
  tmat[,2]<-ifelse(is.na(lrauw$YCVDI-lrauw$YBIRTH),
                   ifelse(is.na(lrauw$YDTHE-lrauw$YBIRTH),lrauw$YENTRY-lrauw$YBIRTH+40,
                           lrauw$YDTHE-lrauw$YBIRTH),lrauw$YCVDI-lrauw$YBIRTH)
  PYfuNonCVD<-PYExAges(tmat)

  tmat[,1]<-lrauw$YCVDI-lrauw$YBIRTH
  tmat[,2]<-ifelse(is.na(lrauw$YDTHE-lrauw$YBIRTH),lrauw$YENTRY-lrauw$YBIRTH+40,
                   lrauw$YDTHE-lrauw$YBIRTH)
  tmat<-na.exclude(tmat)  # Remove the NonCVD cases
  PYfuCVD<-PYExAges(tmat)

  nlsmat<-data.frame(Ag=ltAges,Occ=Deaths[lowoffset:hioffset,2],PY=PYfuNonCVD[lowoffset:hioffset])
  # A dataframe with age in col 1, occurrences by age in col 2, and PY in col 3

  glmpars<-glm(Occ~offset(log(PY))+Ag,family=poisson(link=log),data=nlsmat)  # smoothing death rates
  H2Drate<-exp(coef(glmpars)[1]+coef(glmpars)[2]*ltAges)

  nlsmat[,2]<-EntryCVD[lowoffset:hioffset]
  glmpars<-glm(Occ~offset(log(PY))+Ag,family=poisson(link=log),data=nlsmat)
  H2Crate<-exp(coef(glmpars)[1]+coef(glmpars)[2]*ltAges)

  nlsmat[,2]<-Deaths[lowoffset:hioffset,1]
  nlsmat[,3]<-PYfuCVD[lowoffset:hioffset]
  glmpars<-glm(Occ~offset(log(PY))+Ag,family=poisson(link=log),data=nlsmat)
  C2Drate<-exp(coef(glmpars)[1]+coef(glmpars)[2]*ltAges)

  tt<-LifeTabsub(Deaths,EntryCVD,H2Drate,H2Crate,C2Drate,MaleBool)
  tt
}
ByLifeTab <- function(lrawu)
{
  tttt <- by(lrawu, list(Male = (lrawu$SEX == 1), Smoker = (lrawu$SMSTATUS == 'Always')), LifeTab)
  uitpars <- matrix(nrow = 4, ncol = 5)  # Is a by-object, bootstrapping requires matrix output
  uitpars[1,] <- tttt[1][[1]]
  uitpars[2,] <- tttt[2][[1]]
  uitpars[3,] <- tttt[3][[1]]
  uitpars[4,] <- tttt[4][[1]]
  uitpars
  #tttt
}

options(contrasts = c(factor = "contr.treatment", ordered = "contr.poly"))  # Factor contrast has to be
# treatment, not helmert (the default)

#temp <- ByLifeTab(SMNeverCVD)

temp <- bootstrap(SMNeverCVD, ByLifeTab, B = 1000)  # hist(temp$replicates[,1],nclass=40) gives a
# histogram, for 1-40 outputs
Cox and Gompertz regression models
An assessment with empirical estimates

Abstract

To construct a life table (with or without covariates), age-specific transition rates are necessary. Transition rates can be derived directly from the data set or can be estimated by fitting regression models to the data. We used the Cox model, which is semi-parametric, and the Gompertz model, which is parametric. The age-specific transition rates obtained from the Cox and Gompertz models in the presence or absence of covariates and the consequences for the life table estimates have not been compared to the empirical rates. In this chapter, an assessment of these two widely used models is made by means of a comparison with age-specific occurrence-exposure rates. We used 48 years of follow-up of cardiovascular morbidity and mortality in the original Framingham Heart Study (FHS) cohort to illustrate and compare model estimates. We estimated the age-specific transition rates for both the null model (without covariates) and the model with covariates. Finally, the different estimates of the rates were transferred into life table outcomes. As expected, both the Cox and the observed rates were found to be the same in the null model. On adding covariates to the Cox model, the estimated age-specific rates did not overlap with the observed rates because of the proportionality assumption. The Gompertz estimate behaved in the same way in both the null model and in the model with covariates. The Cox null model could be applied instead of observed rates if the number of occurrences at each age group was sufficiently large, in which case the Cox null model and observed rates were seen to yield the same transition rates. The Gompertz model (with or without covariates) fit well with estimates of cardiovascular disease and mortality transition rates in the FHS population. Since the variability of the estimated transition rates in the Gompertz model was less than in the Cox model or empirical rates, life table estimates using Gompertz in the presence or absence of covariates is recommended.
7.1 Introduction

Survival analysis examines and models the duration of events of interest. It focuses on the distribution of survival times. While there are well-known methods for estimating survival distributions (e.g., Kaplan-Meier estimate), the most interesting method to model survival examines the relationship between survival and one or more predictors, usually termed covariates (Fox, 1997). In this chapter, we have focused on two survival models: the semi-parametric Cox regression model (Cox, 1972; 75) and the parametric Gompertz regression model (Gompertz, 1825; 1827).

The Cox regression model is the model on which modern survival analysis is founded and is widely used in numerous fields. The parametric Gompertz regression model has dominated mortality analysis for over 100 years and has been applied in a range of disciplines from botany to sociology. To construct a life table (with or without covariates), we needed age-specific transition rates. Transition rates can be derived directly from the data set (e.g., occurrence-exposure rates) or can be estimated by fitting regression models to the data. However, the age-specific transition rates obtained from the Cox and Gompertz models in the presence or absence of covariates and the consequences for the life table estimates have not been assessed. In this chapter, an assessment of these two widely used models is made by comparing the estimated rates with empirical age-specific occurrence-exposure rates.

The Cox proportional hazard model is the most general of the regression models, because it is not based on any assumptions concerning the nature or shape of the underlying distribution. The basic assumption of this model is that the underlying transition rate (rather than the survival time) is a function of the independent variables (covariates). No assumptions are made about the shape of the hazard function (Kleinbaum, 1996; Blossfeld and Rohwer, 2002; Therneau and Grambsch, 2000). Proportional hazard models consider the transition rates at each time among those subjects who have not failed. The predicted transition rates using the Cox model gives an efficient estimate, if proportionality exists, as compared to a parametric proportional hazards model, such as the Weibull, even when the data actually come from the parametric model (Kleinbaum, 1996). To know the effect of covariates on the occurrences of event of interest, researchers mostly estimate the hazard ratios. However, the estimated age-specific rates, and accordingly the life table outcomes using Cox model with age as time scale, in the presence or absence of covariates, has not been compared with empirical age-specific rates.

While the Cox model does not have any pre-specified form of the baseline hazard, the Gompertz function follows an exponential form (increasing or decreasing). The Gompertz model has fewer parameters and some other facilities. Both the Cox and Gompertz models have advantages and disadvantages (described in Section 7.2). One advantage of the Gompertz model is that it produces smooth
age-specific transition rates. To estimate the age-specific transition rates for use as input in a life table, researchers prefer to use smoothed transition rates (e.g. a log-rate model) to avoid the stochastic variability in the transition rates across ages (Nusselder, 1998). In this thesis, we applied the Gompertz model (Chapter 9) as a regression model and as a smoothing technique (Chapter 6).

In this chapter, first the null-model is compared with the empirical rates. The transition rates obtained from the Cox null model will be the same as the empirical rates. The transition rates estimated applying the Cox model with covariates should be proportional to different values of the covariates. The Gompertz model smoothes the observed rates. Second, the age-specific transition rates are then used to construct life tables. The multistate life table is used to derive several life table indicators: life expectancy, life expectancy free of CVD and lifetime probability of CVD. The life table outcomes could demonstrate the differences in the Cox, Gompertz and observed rates. We used the first 48 years of follow-up of the original Framingham Heart Study (FHS) cohort for estimation of the transition rates and model comparisons.

We have described the Cox model and the Gompertz model with their limitations in Section 7.2. Sub-section 7.2.3 describes the data source used to illustrate the transition models. Transition rates are compared in Section 7.3. Life table estimates are compared in Section 7.4. Section 7.5 concludes this chapter.

### 7.2 Models and data

The objective of the transition rate models in lifetime analysis, survival analysis, failure time analysis, or in event history analysis is to study (describe and explain) the time, $T$, until an individual has changed state. In our study, the time scale was age, say $X$, instead of time $T$. Let $X$ represent age. We regard $X$ as a random variable with a cumulative distribution function $F(x) = \Pr(X \leq x)$, probability density function $f(x) = dF(x)/dx$, the survival function $S(x) = \Pr(X > x) = 1 - F(x)$ and the hazard rate at age $x$, conditional on survival to that age:

$$\mu(x) = \lim_{\Delta x \to 0} \frac{\Pr[(x \leq X < x + \Delta x) | X \geq x]}{\Delta x} = \frac{f(x)}{S(x)}$$

Here $\mu(x)$ represents the transition intensities estimated from the model. The empirical occurrence-exposure rates ($M(x)$) were calculated from the data using equation 2.7 (Chapter 2). We estimated the age-specific transition rates using Cox, Gompertz and observed occurrence-exposure rates. The Cox and Gompertz models, the estimations and their limitations are described in the below.
7.2.1 The Cox regression model

The impact of various explanatory variables that might affect the transitions or event occurrences is examined using the Cox (1972; 1975) hazards model. Under this model, it is assumed that the hazard function is proportional for different values of an explanatory variable. More specifically, the rate of event occurrences for an individual with explanatory variable value given by the row vector \( z^{(ij)} \) is assumed to be

\[
\mu_i(x) = \mu_{0ij}(x) \exp(z^{(ij)}\beta^{(ij)})
\]  

(7.1)

where \( i \) is the state of origin, \( j \) is state of destination, \( \mu_{0ij}(x) \) is the “baseline” rate of transition from \( i \) to \( j \) to which all force of events functions are proportional (i.e. shape is unspecified), \( z^{(ij)} \) is a row vector of \( p \) measured covariates and \( \beta^{(ij)} \) is a column vector of \( p \) parameters. The baseline hazard function \( \mu_{0ij}(x) \) has been defined as the hazard for an individual with values of all explanatory variable equals to zero. The ability of this method to detect the differences of event rates associated with the different values of covariates is influenced by the extent to which the forces of event occurrences are proportional. In that case, the row vector \( Z^{(ij)} \) is independent of time \( x \).

The partial likelihood method gives an estimate of the parameters of \( \beta^{(ij)} \) but no direct estimate of the underlying baseline hazard rate, i.e. without knowing the baseline hazard, we estimate \( \beta^{(ij)} \) parameters. The partial likelihood method for the Cox model was developed under the assumption of continuous data, but real data sets often contain tied event times. Among the exposed people, tied events are those events that occur at the same time or age. Ties may occur because continuous event times are grouped into intervals or because the event time scale is discrete. Recently, Therneau and Grambsch (2000) summarized four variants of the computing algorithm that are commonly used to address the tie issues. (i) Breslow approximation - which is the simplest in terms of programming, and consequently was the only method available in the earliest Cox model routines, is still the default method in almost all packages. However, the solution is the least accurate. It counts failed individuals more than once in the denominator, which produces both a conservative bias and estimate regression coefficients too close to 0 in absolute value (Cox and Oakes, 1984); however, the method is fast. TDA’s and SPSS’s approach is based on a proposal made by Breslow. (ii) Efron approximation - is quite accurate unless the proportion of ties or the number of tied events relative to the size of the risk set is extremely large, and is as fast as the Breslow method. This approximation is the default in S-Plus. (iii) Exact partial likelihood - this approach engages an exhaustive account of the possible risk sets at each tied event time, and
can require an excessive amount of computation time if any of the individual death
times has a large number of events (>10 say) (Thernau and Grambsch, 2000). S-
Plus uses this as the exact option and SAS as the discrete option. (iv) Average likelihood-
this is often very close to the Efron approximation. It also engages an exhaustive
account, but a substitution allows this to be replaced by the numerical evaluation of
an integral. This estimation is nearly untraceable computationally when the number
of tied events at any time is even moderately large. In SAS, it is called the exact
option.

**Advantages and disadvantages**
The application of the Cox model is very simple, software is available and no
assumptions are made about the nature or shape of the hazard function. However,
there are several disadvantages to this model. The relative lack of precision for the
partial likelihood estimates of parameters compared with maximum likelihood
estimates is expected to decrease with an increase in the sample size for most
empirical situations (Efron, 1977; Oakes, 1977; Wong 1986). Another disadvantage
of this model is the presence of tied events, although several approaches have been
developed to approximate the tied events. The partial likelihood method does not
permit us to analyze the form of time dependence directly. This method is relatively
disadvantageous when the form of time dependence itself is of substantive interest
(Yamaguchi, 1991). Finally, the partial likelihood method is based on weaker
theoretical foundations than is the maximum likelihood method (Kalbfleisch and
Prentice, 1980; Yamaguchi, 1991). In our application of the Cox model, the
estimation of age-specific transition rates will be influenced by sample size
(especially at older ages) and tied events as well.

### 7.2.2 The Gompertz regression model

The “Gompertz law” (Gompertz, 1825; 1827) of mortality states that the force of
mortality (e.g. instantaneous death rate) increases exponentially with age because
the ‘resistance to death’ declines with age. The change in resistance to death is a
latent causal process that we cannot measure directly, only the effect (e.g death) can
be measured (Willekens, 2001). The basic assumption of the Gompertz law is that
the resistance to death declines exponentially with age. The basic notion is that at
each time period or point, a person loses a constant fraction of his or her remaining
‘vital force’ or vitality. Gompertz summarized the effect of unobserved processes as
a parametric form of age dependence. The Gompertz model dominated mortality
forecasting for more than 100 years (Olshansky and Carnes, 1997), and it is also
used in duration analysis in general (Blossfeld and Rohwer, 2002). In this study, we
assumed that any disease or mortality transition (following the basic model of
cardiocvascular disease, Figure 3.1(b), Chapter 3) followed Gompertz’s law. In
addition, all age-specific transition rates were assumed to depend on the incorporated covariates in the model. The transition rate at age $x$ is given by the expression

$$\mu(x) = b \exp(cx)$$  \hspace{1cm} (7.2)$$

Where, $b$ and $c$ are parameters to be estimated from the data. When $c=0$, the model trims down to the simple exponential model. The Gompertz model has two parameters and both of them can be used to include covariates. The model formulation for the transition rate from the origin state $i$ to destination state $j$ is

$$\mu_{ij}(x) = b^{(ij)} \exp(c^{(ij)}x)$$  \hspace{1cm} (7.3)$$

where, $b^{(ij)} = \exp\{\beta^{(ij)}\}$. Following Blossfeld and Rohwer (2002), the associated coefficient vectors $\beta^{(ij)}$, linked exponentially and $c^{(ij)}$, linked linearly, are model parameters to be estimated. The $\beta^{(ij)}$ vector is linked exponentially to make sure that the estimated transition rate will not become negative. $c^{(ij)}$ is the shape parameter.

Following Rohwer and Potter (1999, TDA user’s Manual), we estimated the parameters from the Gompertz regression model using the maximum likelihood method of estimation. Details of the derivation of the maximum likelihood estimates of the Gompertz model with and without covariates are given in Rohwer and Potter (1999).

**Gompertz in the context**

In this study, we considered the age at transition to an event or censoring as a time variable. The transition rates at each age interval depended on the age of an individual at that time and the associated covariates incorporated in the Gompertz model. Gompertz is one of the best known models, applied in a range of disciplines from botany to sociology. For mortality forecasting, Gompertz is still one of the best models in demographic literature. Since we estimated age-specific death rates either from non-cardiovascular disease or cardiovascular disease states, we assumed that the Gompertz mortality law is plausible in this case. As age increases (especially at older ages) the risk of cardiovascular disease increases irrespective of risk factor status at that age (Stamler et al., 1999). Therefore, the assumption that the shape of the transition rate from non-cardiovascular disease to cardiovascular disease follows Gompertz’s law is also plausible. Some of the advantages and limitations of the Gompertz model are discussed here.
Advantages and disadvantages
The transition rates follow a Gompertz law, i.e. they increase exponentially with age, but the level and shape of the age-dependency is dependent on several covariates. Since our interest is to incorporate covariates, there will be an impact of these covariates. Therefore, we may assume that as more knowledge becomes available in the Gompertz model, the age dependency of the rate not only depends on the Gompertz law but also on individual characteristics (e.g. covariates). Manton et al. (1997) also argued that as more knowledge becomes available on disease and risk factors, the contribution of the Gompertz component is reduced. The simple Gompertz model without any covariates assumes that there is no important heterogeneity across individuals and each transition rate, as we are able to incorporate covariates with each parameter of this model that reduce the individual heterogeneity. The Gompertz regression model provides smooth transition rates and reduces the variability of measurement. The Gompertz function is monotonically increasing or decreasing. The Gompertz model can be used as a multidimensional context (Manton, et al., 1994).

7.2.3 Data source
We used the first 48 years of follow-up of the original Framingham Heart Study cohort consisting of 5209 respondents (45% male) from a sample of adults aged 28 through 62 years, residing in Framingham, Massachusetts between 1948 and 1951. The participants were tracked by standardised biennial cardiovascular examination, daily surveillance of hospital admissions, death information and information from physicians and other sources outside the clinic, ensuring highly accurate follow-up of death and clinically presenting cardiovascular disease. For this chapter, we used the data regarding age at onset of cardiovascular disease or death over forty-eight years of follow-up (exam rounds 1 to 24) for the 3481 participants who were free of cardiovascular disease at age 50.

For the purpose of illustration, we considered educational status as a covariate. We considered educational levels as a constant categorical covariate: low education (maximally high school without having graduated) and high education (at least graduation from high school). The educational status is described in Chapter 6.

7.3 Transition rates comparison
Age-specific transition rates were estimated using the Cox and the Gompertz regression model. The estimated rates were compared to the age-specific observed occurrence-exposure rates. The observed age-specific occurrence-exposure rates were estimated using the method described in Section 3.4.4 of Chapter 3. To illustrate, age-specific transition rates were estimated for the basic model type 3(a)
and 3(b) in Chapter 3, where, 3(a) was a two state (alive, dead) model and 3(b) a 3-state \{NO-CVD, CVD, Dead\} model. First, we compared different transition rates without any covariate i.e. the null model. Second, rates were compared with the presence of education, i.e. the model with covariate. Rates were compared from age 50 to 90.

### 7.3.1 Model without covariates (null model)

**Transition rates alive to death**

We used Cox and Gompertz models to estimate the age-specific death rates. The model estimates were compared to the observed rates in Figure 7.1. In the null model, we only estimated the baseline hazard. As expected, the estimated rates from the Cox model without any covariate was the same as the empirical rates, which showed that if the sample population was sufficiently large (before age 80), the Cox model and the observed rates were equal. Because of small sample size and estimation procedure, the Cox estimated transition rates at older ages differed slightly from the empirical rates. The estimated transition rates by the Gompertz model were smooth and increased exponentially. The Gompertz estimates fit quite well with both the empirical rates and the Cox model for males and females (Figure 7.1).

![Figure 7.1](image)

**Transition rates NO-CVD to CVD**

The age-specific transition rates NO-CVD to CVD of the model estimates and observed rates by sex are presented in Figure 7.2. Like mortality rates, cardiovascular disease incidence rates also increase as age increases. Both the
observed and the Cox rates show a similar pattern. The disease rates estimated applying the Gompertz model are shown to fit well compared to the Cox and the empirical rates.

Figure 7.2  Comparison of age-specific rates of transition from NO-CVD to CVD: Cox, Gompertz and observed occurrence-exposure rates, by sex

Transition rates \textit{NO-CVD to death}

The transition rates from NO-CVD to death are compared in Figure 7.3. The Cox null model fits with the empirical rates. At older ages, small differences were seen due to the limited number of transitions and the approximation of the Cox model (see Section 7.2.2, the limitation of Cox model). Like other transitions, Gompertz fits quite well for the NO-CVD to death transition.

Figure 7.3  Comparison of age-specific rates of transition NO-CVD to death: Cox, Gompertz and observed occurrence-exposure rates, by sex
7.3.2 Model with covariates

The transition rates, described above were also compared by educational status. For the sake of simplicity, the educational levels are divided into the two categories low education and high education.

Transition rates alive to death

Incorporating educational status as a covariate, a comparison was then made between the model estimates and the empirical rates (see Figure 7.4). This figure clearly demonstrates that the age-specific transition rates estimated by the Cox model differ from the observed rates although the trend is same. The Cox estimates are less extreme than the observed rates because of the proportionality assumption. Given the exponential nature, Gompertz increased exponentially, despite the irregular changes of the observed rates and the Cox estimates. Mortality rates for the male low education group appear to level off after age 80, but the Gompertz estimates increase exponentially.

Figure 7.4 Comparison of age-specific death rates: Cox, Gompertz and observed occurrence-exposure rates, by sex and education
Transition rate *NO-CVD to CVD*
Similarly, the cardiovascular disease transition rates are compared by educational status in Figure 7.5. We observed the same behavior of the Cox model because of proportionality assumption. The Gompertz rates increased exponentially, as usual.

Figure 7.5 Comparison of age-specific rates of transition NO-CVD to CVD: Cox, Gompertz and observed occurrence-exposure rates, by sex and education

**Transition NO-CVD to death**
The transition rates from NO-CVD to death by educational status are shown in Figure 7.6. Both the observed and Cox estimates show irregular patterns. The Gompertz estimates increase exponentially by nature, whether the observed rates increase, decrease or remain at the same level (Figure 7.5, Female: high education). This irregular behavior of the observed rates was probably due to the small number of transitions or sample selection.
Similarly, we can also compare the age-specific transition rates for the transition from CVD to death. We did not estimate the post-disease transition rates for the Cox model, since our applied software (SPSS or TDA) was unable to handle delay entry or left censored cases, although estimating the baseline hazard for the Cox model was very simple. On the other hand, using a number of advanced statistical packages (STATA, S-plus), gave us the option to deal with delay entry or left censoring, but made it difficult to estimate the baseline hazard for the Cox model. To construct the basic CVD life table using the Cox model, CVD to death transition rates were replaced by the empirical rates.
7.4 Life table outcomes

By comparing different transition rates, we were able to see how one curve fits better compared to another or how the outliers behaved over age. As was seen, both the Cox and the empirical rates had nearly the same outliers, while Gompertz gave smooth and exponentially increasing estimates of transitions rates. Here, we have constructed several life tables, using the age-specific transition rates obtained from the observed occurrence-exposure rates, Cox and Gompertz regression models. Some of the basic life table estimates such as survival curves, lifetime risk of developing cardiovascular disease and life expectancies are compared. Life tables are constructed following the procedure described in Section 3.3.4 of Chapter 3. Since CVD to death transition rates in the Cox model are replaced by the empirical rates, this might influence the life expectancy with CVD in the Cox model. We therefore used the MSLT to compare the life expectancy free of CVD, survival probability in NO-CVD state and lifetime probability of disease. Total life expectancies are presented from a single decrement (SDLT) life table.

Survival curves

The age profiles of survival free of cardiovascular disease for the FHS male and female participants in the MSLT population are shown in Figure 7.7. This figure shows the survival free of cardiovascular disease in fifty-year-old men and women by educational status.

As expected, the more highly educated men and women led longer lives and survived longer free of any cardiovascular disease than less well educated individuals. Overall, the fitted survival curves using the observed rates and model estimated rates were very close to each other. For instance, a marginal difference between the fitted survival curves is displayed in Figure 7.7 (Female low education), as they did not overlap at every age. However, this small difference was to be expected as the Cox and empirical transaction rates are neither smooth nor equal. At older ages, the survival curves estimated by the Gompertz regression model turn slightly downwards compared to the Cox and observed rates.
Figure 7.7 Survival curves illustrating the probability of surviving free of cardiovascular disease by age and educational status.
**Lifetime probability**

We compared the probability of cardiovascular disease-free people at age 50 developing cardiovascular disease by sex and education (Table 7.1 and Figure 7.8). The procedure to estimate the lifetime risk is the same as the procedure applied in Section 3.4.3 of Chapter 3.

The age-specific probability of developing cardiovascular disease is estimated by the Cox model to be the same as the observed estimate in the null model. Adding a covariate to the Cox model, causes the age-specific probability to jump less than the observed rates, because of the proportionality assumption of Cox model. The Gompertz estimate gives smooth age-specific probabilities of developing cardiovascular disease.

Figure 7.8  Age-specific probability of cardiovascular disease free people at age 50 developing cardiovascular disease, by sex and education
While Figure 7.8 only demonstrates the overall fit of the models compared to observed estimates, the probabilities of developing cardiovascular disease before age 70 and 90 are shown in Table 7.1. The probability of developing disease before age 90 according to both model estimates and the observed one (except high education females) was the same. Before age 70, the model estimate varied by one to two percent. The result also demonstrated that the probability of developing cardiovascular disease by educational status varied from one to three percent for males and one to seven percent for females.

<table>
<thead>
<tr>
<th></th>
<th>Before age 70</th>
<th>Before age 90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Cox</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>Low education</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>High education</td>
<td>40</td>
<td>39</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Low education</td>
<td>28</td>
<td>26</td>
</tr>
<tr>
<td>High education</td>
<td>21</td>
<td>22</td>
</tr>
</tbody>
</table>
Life expectancy
In addition to survival probability and the risk of developing cardiovascular disease, the total life expectancy (from SDLT) and life expectancy free of cardiovascular disease (MSLT) at age 50 and 70 is presented in Table 7.2. Using the observed rates, a male at age 50 can expect to survive 26.90 additional years, while this was 27.01 years according to the model-derived rates. Similarly for females, the observed Cox and Gompertz estimates of life expectancy at age 50 were 32.03, 32.16 and 32.13 years respectively. As can be seen from this table, the model estimate of the total life expectancies and life expectancies free of cardiovascular disease at age 50 and 70 do not differ much. The results are comparable in the absence or presence of a covariate. A slight difference (0 to 0.5 years) continues mainly because of different assumptions of the models and sample size.

Table 7.2 Total life expectancy (SDLT) and life expectancy free of cardiovascular disease (MSLT)

<table>
<thead>
<tr>
<th></th>
<th>Age 50</th>
<th></th>
<th>Age 70</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Low education</td>
<td>High education</td>
<td>Total</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (SDLT) Observed</td>
<td>26.90</td>
<td>26.33</td>
<td>27.29</td>
<td>12.55</td>
</tr>
<tr>
<td>Cox</td>
<td>27.01</td>
<td>26.40</td>
<td>27.43</td>
<td>12.69</td>
</tr>
<tr>
<td>Gompertz</td>
<td>27.01</td>
<td>26.41</td>
<td>27.43</td>
<td>12.41</td>
</tr>
<tr>
<td>Free of CVD Observed (MSLT) Cox</td>
<td>20.83</td>
<td>20.12</td>
<td>21.32</td>
<td>7.40</td>
</tr>
<tr>
<td>Gompertz</td>
<td>20.99</td>
<td>20.37</td>
<td>21.52</td>
<td>7.45</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (SDLT) Observed</td>
<td>32.03</td>
<td>30.96</td>
<td>32.69</td>
<td>15.86</td>
</tr>
<tr>
<td>Cox</td>
<td>32.16</td>
<td>31.28</td>
<td>32.47</td>
<td>16.02</td>
</tr>
<tr>
<td>Gompertz</td>
<td>32.13</td>
<td>31.08</td>
<td>32.44</td>
<td>15.71</td>
</tr>
<tr>
<td>Free of CVD Observed (MSLT) Cox</td>
<td>26.60</td>
<td>25.13</td>
<td>27.51</td>
<td>11.16</td>
</tr>
<tr>
<td>Gompertz</td>
<td>26.82</td>
<td>25.49</td>
<td>27.63</td>
<td>11.10</td>
</tr>
</tbody>
</table>
7.5 Discussion

The aim of this chapter was to estimate the age-specific transition rates applying Cox and Gompertz regression models, and to assess the model estimates by comparing them to the observed occurrence-exposure rates. The different estimates of the rates were subsequently transferred into life table outcomes. In the null model, both the Cox and the observed rates were similar. Adding covariates to the Cox model did not cause the estimated age-specific rates to overlap with the observed rates, because of the proportionality assumption. The Gompertz estimates behaved the same way in both the null model and the model with covariates.

When the available data on the exact timing of events and occurrences of events followed a specific shape, a parametric model could be used instead of a semi-parametric Cox model. The Cox null model could be applied instead of observed rates when the number of occurrences at each age group was sufficiently large. In that situation, the Cox null model and the observed rates were exactly the same. The estimation of the Cox null model was simple, easy and software is available.

The Gompertz model smoothed the rates exponentially. When applying the Cox model and estimating observed rates, we could only predict the rates within the follow-up time. For the Cox estimate, we needed a baseline hazard to delineate the predicted rates. Gompertz could be used both for macro data and micro data. Cox could be used mainly for the micro data. Gompertz had fewer parameters. Cox provided partial likelihood estimates, which is theoretically weaker. The precision of the partial likelihood estimates of parameters can be much less than that for the maximum likelihood estimates when the sample size is small (Coleman, 1981). When we estimated the age-specific transition rates at older ages the sample population was small. Because of the small sample size and partial likelihood method, the transition rates at older ages were not same as the observed rates.

The Cox proportional hazard model depends on the proportionally assumption. It does not have any untenable distributional assumptions. The main advantages of the Gompertz estimate are the smooth estimates that are more interpretable and easy to apply for the life table construction. Since Gompertz provides smooth estimates, the variability of the results will be less compared to the Cox method and the observed rates. However, to estimate the transition rates and accordingly construct a life table for the cardiovascular disease history in the FHS, a Gompertz model with or without covariate fits well. Using Gompertz regression, the rates could be predicted for any ages. Since the variability of the estimated transition rates in the Gompertz model is less than the Cox or empirical rates, life table estimates using Gompertz in the presence or absence of covariate is recommended.
References


A multivariate analysis of the risk factors of cardiovascular disease and its subtypes

Abstract

The objective of this chapter was to investigate the effect of risk factor status at middle age (age 30 to 50) on the incidence of cardiovascular disease and its subtypes and post-disease mortality. The analysis is based on a 48-year follow-up of the original Framingham Heart Study cohort. The following risk factors were considered: smoking, body mass index, serum cholesterol level and blood pressure. Three risk levels - optimal, elevated and high - were distinguished for each risk factor. The effects of the major cardiovascular risk factors are expressed in terms of relative risk (RR) of disease incidence and death. We performed both univariate and multivariate regression analysis using the Cox proportional hazard model. The univariate analysis indicates the effect of the level of a single risk factor irrespective of the levels of the other risk factors. Since the effect of a risk factor may differ by level from another risk factor, multivariate analysis was used to control for the confounding effects of the presence of other risk factors. Interaction effects were not included in the multivariate analysis. Although there were interactions between the risk factors, the sample size did not permit the study of the interaction effects.

Univariate analysis indicated that each risk factor had a substantial effect on the risk of CVD and mortality. For instance, males with a life history of smoking (always smoker) were found to have a 34% excess risk (ER) of CVD and a 63% excess mortality compared to never smokers. Obese males (BMI \( \geq 30 \)) were shown to have an 81% excess risk of CVD and a 47% excess mortality compared to males of normal weight. The effect of smoking was aggravated by the presence of other risk factors. For instance, a male always smoker of normal weight, with an optimal blood pressure and an optimal cholesterol level had a 44% excess risk of CVD and a 72% excess mortality as weighed against a comparable male who never smoked. The effect of obesity, hypertension, or high cholesterol level on CVD was reduced by the presence of other risk factors. For instance, a downward change in excess
risk of 41% for CVD and 42% for mortality was found for a never-smoking obese male with an optimal blood pressure, and an optimal cholesterol level. The effect of obesity, hypertension or high cholesterol was less strong when the confounding effect of other risk factors was removed (in the absence of interaction effects). A significant reduction in long-term risk may be brought about by appropriate prevention or intervention of major cardiovascular risk factors, ultimately leading to a reduction of the incidence of cardiovascular disease.

8.1 Introduction

To obtain precise estimates of the net or partial effect of a specific cardiovascular risk factor on cardiovascular disease incidence and mortality, we needed to control for other risk factors. The best strategy for the prevention of cardiovascular disease is to take into consideration various risk factors simultaneously (Anderson et al., 1991; Lowe et al., 1998). The multifactoral risk factor impact on the incidence of cardiovascular disease and mortality has been measured (Stamler et al., 1999; Anderson et al., 1991). Most of the studies measured multifactoral risk factor impact relying on one measurement and a short follow-up. Any major risk factor left untreated for many years has the potential to produce cardiovascular disease. Even though the use of risk-reducing drugs can significantly lower the risk when begun in later years, there is no evidence that it can return a person to the optimal risk status of a younger person (Grundy et al., 1999). In public health research, an important aim of primary prevention is to reduce cardiovascular disease over the long term and not just over the short term. Therefore, the primary objective of this chapter was to investigate the effect of the risk factor status at middle age (age 30 to 50) on the incidence of cardiovascular disease, its subtypes and post-disease mortality considering a long follow-up. Moreover, in this chapter the groundwork is laid for the next chapter, in which an MSLT is created for a single risk factor and for a combination of several risk factors.

Risk factors in early life, like tobacco use and obesity, not only affect an individual's own later health but also the health of the next generation (WHO, 2002). People with low risk profiles in middle age survive longer (Stamler et al., 1999, Lowe et al., 1998) and consume lower average annual costs for medical care in older age (Daviglus et al., 1998). In this chapter, we have focused on an association of the multifactoral risk factor status at middle age and the disease incidence and mortality at older ages. Although there are many established risk factors for CVD, we focused on the following major CVD risk factors: smoking, systolic blood pressure (SBP), diastolic blood pressure (DBP), blood pressure (BP), serum cholesterol level (SCL), and body mass index (BMI). These are the standard cardiovascular risk factors (Pooling Project Research Group, 1978; Berenson et al.,
A multivariate analysis of cardiovascular risk factor

1998; Stamler et al., 1999; Lowe et al., 1998). They are likely to have a beneficial impact on all-cause mortality and incidence of cardiovascular disease (Norrish et al., 1995).

The major cardiovascular risk factors are additive in predictive power (Wilson et al., 1998). Thus, the total risk of a person can be estimated by summing the risk conveyed by each of the major risk factors (Grundy et al., 1999). We performed both univariate and multivariate analysis of the major cardiovascular risk factors in middle age to assess their effects on CVD and mortality in the later ages of life. We estimated the effect of each risk factor status separately in a univariate analysis. The multivariate analysis was carried out to study the relative risk of cardiovascular disease or death associated with a risk factor, adjusted for the presence of other risk factors. These associations between risk factor status in middle age and disease incidence and mortality in later ages were investigated using the first 48 years of follow-up data in the Framingham Heart Study. We measured the effects in terms of relative risk of disease incidence and mortality.

In Section 8.2, we have illustrated the data and methods. In this section, the data source, risk factor definition, state space, transitions, relative risk estimation and model specification are described. Results are discussed in Section 8.3. In subsection 8.3.1, we have briefly described the disease transitions. Deaths are discussed in subsection 8.3.2. The chapter is concluded with a discussion in Section 8.4.

8.2 Data and methods

8.2.1 Data source

We used the original Framingham Heart Study cohort, which consisted of 5209 respondents (45% male) from a sample of adults aged 28 through 62 years residing in Framingham, Massachusetts between 1948 and 1951. The selection criteria and study design have been described elsewhere (Dawber et al., 1951). In the Framingham Heart Study, current smoking status (yes or no, number of cigarettes smoked in a day), serum cholesterol level, systolic blood pressure, diastolic blood pressure, height and weight were measured and recorded at most biennial examinations.

For this chapter, we have used the data regarding the age at onset of cardiovascular disease and its subtypes, and age at death over 48 years of follow-up (exam rounds 1 to 24) of the participants who were free of cardiovascular disease at age 50. We excluded the participants whose risk factor status was below optimal levels (Table 8.1). People with risk levels that are below the optimal are often at a high risk of experiencing disease and mortality. For instance, low blood cholesterol
increases mortality (Jacobs et al., 1992). We found a total of 3045 participants for whom all risk factors had been recorded at least two exams during the age interval from 30 to 50\(^1\). The same number of participants was included in both the univariate and the multivariate regression analysis.

### 8.2.2 Risk factor definition

An individual was included in our analysis if he or she had appeared between the ages of 30 and 50 at minimally two exams at which data on all risk factors were recorded. Smoking status was allocated for each participant based on the current smoking status recorded at each available exam from age 30 to 50. We classified *never smokers* as those with all available smoking records coded as a non-smoker and *always smokers* as those with all available smoking records coded as a smoker. *Ever smokers* constituted the remaining group of participants, a group that is characterized by a mixture of smoking and non-smoking throughout the period from entry at the survey to age 50. Systolic blood pressure or diastolic blood pressure is defined at each exam based on the mean value of recorded SBP or DBP from two different examiners. We took the average of the recorded mean SBP or DBP between age 30 and 50 years. Using World Health Organization (WHO, 1999) guidelines, we categorized blood pressure combining SBP and DBP as follows: [optimal BP- SBP<120 and DBP<80, high BP- SBP>140 or DBP>90, otherwise high normal or elevated BP. Likewise, we took the average of recorded serum cholesterol and body mass index (BMI)\(^2\) between age 30 and 50 as predictors. Body mass index was calculated as weight in kg/height in m\(^2\) (m is meter). We defined three BMI categories based on the World Health Organization guidelines (1998): [normal weight- BMI of 18.5 to 24.9 kg/m\(^2\), overweight- BMI of 25 to 29.9 kg/m\(^2\) and obese- BMI greater than or equal to 30 kg/m\(^2\)]. Similarly, we defined three categories of total cholesterol levels: [optimal SCL- SCL<200 and SCL>160, high normal SCL- SCL>200 and SCL<240, high SCL- SCL>240].

\(^1\) Note that at onset of the FHS follow-up, respondents were at least 28.

\(^2\) **BMI**: BMI is derived by dividing weight (in kg) by height square (m\(^2\)). Respondent’s weights are measured at all exams but heights are missing for several exams (9 exams out of 24 exams). Since adults’ height usually does not change during a short time interval, we assume that the height for the missing exams is same as of the nearest recorded exam (before age 50).
The risk factor exposures are defined in Table 8.1 as follows:

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Exclude</th>
<th>Optimal</th>
<th>Moderately elevated/high normal</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking (yes/no)</td>
<td>&lt;100</td>
<td>Never smoker</td>
<td>Ever smoker</td>
<td>140+</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>100-119.9</td>
<td>120-139</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg)</td>
<td>&lt;60</td>
<td>80-89.9</td>
<td></td>
<td>90+</td>
</tr>
<tr>
<td>Serum cholesterol (mg per deciliter)</td>
<td>&lt;160</td>
<td>200-239.9</td>
<td></td>
<td>240+</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>&lt;18.5</td>
<td>25-29.9</td>
<td></td>
<td>30+</td>
</tr>
</tbody>
</table>

For males, the mean values in the optimal risk category were a BMI of 23 w/h², SBP 114 mm Hg, DBP 74 mm Hg and SCL 184 mg per deciliter, and in the high risk category, a BMI of 32 kg/m², SBP 152 mm Hg, DBP 97 mm Hg and SCL 270 mg per deciliter (Technical Appendix 8.1). Overall, the differences in mean values of the optimal risk profiles and high-risk profiles by sex were comparable. Respectively 67 percent and 18 percent of the men were defined as smokers and non-smokers, compared to 39 percent and 48 percent of the women. Nearly 56 percent of women (36 percent men) had a normal BMI. Thirty-six percent of women (26 percent men) were classified into the low SBP group. More than 21 percent of the sampled population had high SBP. Twenty-four percent of males had high DBP (16 percent females). Thirty percent of males and 24 percent of females had hypertension. Thirty-nine percent of men and 36 percent of women had high cholesterol levels.

The basic multi-state model has the state space \{NO-CVD, history of CVD, dead\} where each ((Figure 3.1(b), Chapter 3)) of the model has CVD represented by one of the specific CVD states: all cardiovascular disease, all coronary heart disease, acute myocardial infarction, stroke and congestive heart failure. For example, for the CVD model, the possible transitions are NO-CVD to death, NO-CVD to CVD, and CVD to death. Following the risk factor categories as defined in Table 8.1, the number of possible transitions throughout 48 years of follow-up of the Framingham original cohort are presented in Technical Appendix 8.1. Since we have analyzed the data on men and women separately, the number of transitions by each category of risk
factor is given in Technical Appendix 8.1 by sex. There were always fewer people in the ever-smoking categories both for men and women.

### 8.2.3 Estimation of relative risk and model specification

**Relative risk**

To measure the effect of risk factors on disease incidence and mortality, we estimated the relative risk. The relative risk (RR) is the ratio of the absolute risk of a given group to that of a reference group, i.e. it is the ratio of two absolute risks. Literally, relative risk represents the ratio of the incidence in the exposed population divided by the incidence in the unexposed population. The denominator of the ratio can be either the average risk of the entire population or the risk of a group free of the risk factors. For instance, the relative risk of smokers developing cardiovascular disease can be defined as the ratio of the risk of a smoker developing CVD to the risk of non-smoker developing CVD. We have estimated the relative risks of each transition (e.g. from NO-CVD to CVD) separately for males and females. Relative risks were estimated to determine the likely effect of risk factors.

To estimate the relative risk for an event occurrence throughout 48 years of follow-up of the FHS for each risk factor, the optimal risk category (e.g. never smoker in Table 8.1) of each risk factor was taken as the reference category. The events considered were death and the onset of all CVD, CHD, MI, CHF and stroke. The relative risk of each risk factor was calculated using Cox regression analysis, taking age at transition as the time scale.

**Model specification**

Univariate analysis was performed to estimate the effect of each single risk factor separately on disease incidence and post-disease mortality. Multivariate analysis was carried out to estimate the impact of several risk factors, as well as to investigate the potential confounding effects. If the relative risks were statistically significant, the upward or downward change in excess risk (CIER) in the multivariate model is given in the parenthesis with bold phase. This risk in the multivariate model is estimated using the following formula:

\[
\text{CIER} = \frac{(\text{RR}_{\text{multivariate}} - \text{RR}_{\text{univariate}})}{(\text{RR}_{\text{univariate}} - 1)} \times 100.
\]

Take, for example, smoking for which the RR changes from 1.34 in univariate analysis to 1.44 in multivariate analysis, which is a change of 29%. An upward CIER is indicated by a “+”, a downward CIER by a “-”. The excess risk was estimated using the simple formula: RR-1.
We used Cox proportional hazard models to estimate the relative risk. A basic feature of the Cox proportional hazard model is that hazard curves or transition rates for different explanatory variables must be proportional to \( \mu(x) \), the baseline hazard which is left unspecified. For each transition \( i \) to \( j \) we fit the Cox regression model with covariates and tested the assumption of proportionality. Although there are many procedures to test the proportionality, we applied two simple ones: (1) a graphical method, where we plotted the log-log of survival function, and (2) the time-dependence of covariates. We opted for the graphical method because we had categorical covariates with few categories. Sometimes the mere visual inspection of the plotted curves is inadequate for determining whether the differences in the log-log plot by different categories are sufficiently large as to violate the proportionality assumption. For instance, we tested the proportionality of listed risk factors on the process time of the occurrences of cardiovascular disease. If visual inspection was inadequate to inspect the proportionality, we tested the time-dependence of covariates. This test is based on the idea that if the proportionality assumption is correct, there should be no interaction effect between the covariates and process time. Details of this procedure are described elsewhere (Blossfeld and Rohwer, 2002).

We found a few cases where post-disease mortality rates were non-proportional for some of the risk factor status. For males, we found the CVD to death rates to be non-proportional by BMI categories; MI to death was non-proportional for BMI categories; stroke to death was non-proportional for smoking status. For females, we found CHD to death by BMI category to be non-proportional in the univariate analysis; MI to death was non-proportional in both the univariate and multivariate analysis for BMI category. Post-disease mortality was influenced by other factors status and disease treatment at that time. The association with risk factor status at midlife was complex as well. We estimated the relative risk of post-disease death, assuming that the proportionally assumption remains valid for all risk factor statuses.
8.3 Results

8.3.1 Disease incidence

Univariate analysis

The effect of each risk factor on disease incidence and mortality was estimated separately for men and women (Table 8.2). For men, smoking was a highly significant predictor which increased the occurrences of CVD (RR= 1.34, CI: 1.11-1.61) and stroke (RR=1.82, CI: 1.17-2.82) compared to non-smoking. The higher the BMI in midlife, the higher the relative risk of CVD and its subtypes at older ages. This ranged from 1.23 (CI: 1.03-1.47) for CHD in the moderate BMI group to 2.37 (CI: 1.50-3.75) for stroke in the obese group. The relative risk of experiencing CHF was 2.20 (1.50-3.22) for obesity, which was consistent with the estimate by Kenchaiah et al. (2002). Both systolic and diastolic blood pressure had significantly positive effects on the occurrences of CVD and its subtypes. For moderately elevated SBP, the RR varied from 1.36 (CI: 1.14-1.62) for CVD to 1.68 (CI: 1.18-2.41) for stroke compared to optimal SBP. For high SBP, the RR ranged from 1.98 (CI: 1.47-2.66) for MI to 3.76 (CI: 2.56-5.52) for CHF. For high DBP, it fluctuated from 1.85 (CI: 1.51-2.26) for MI to 3.41 (2.33-5.01) for stroke. A similar effect was observed for the blood pressure categories. A high or moderately high cholesterol level had a significant effect on the occurrence of CVD, CHD, MI and CHF.

For women smokers, the risk of experiencing MI (RR=1.49, CI: 1.15-1.94), CHF (RR= 1.34, CI: 1.03-1.75) and stroke (RR=1.39, CI: 1.07-1.81) was higher and statistically significant compared to never smoking women. Females with moderate or high BMI in midlife had a significantly higher relative risk of experiencing CVD and its subtypes compared to the optimal BMI group. The higher the SBP or DBP, the higher the relative risks for all cardiovascular disease subtypes. For high SBP, this ranged from 2.12 (CI: 1.52-2.96) for stroke to 3.19 (CI: 2.30-4.40) for CHF. For DBP it ranged from 2.25 for CVD (CI: 1.88-2.70) or MI (CI: 1.65-3.06) to 2.75 (CI: 2.03-3.72) for CHF. Like in men, high cholesterol levels in women had a significantly higher effect on the occurrence of CVD (RR= 1.51, CI: 1.23-1.85), CHD (RR=2.12, CI: 1.58-2.83) and MI (RR=1.79, CI: 1.23-2.60).
Table 8.2  Relative risk of cardiovascular disease and its subtypes (including sudden death) by single risk factor (95% confidence intervals in parentheses) during age 30-50

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.03</td>
<td>0.85</td>
<td>0.79</td>
<td>1.86</td>
<td>0.92</td>
</tr>
<tr>
<td>(0.80-1.31)</td>
<td>(0.64-1.13)</td>
<td>(0.55-1.13)</td>
<td>(1.10-1.176)</td>
<td>(0.57-1.48)</td>
<td></td>
</tr>
<tr>
<td>Always smoker</td>
<td>1.34</td>
<td>1.14</td>
<td>1.16</td>
<td>1.82</td>
<td>1.30</td>
</tr>
<tr>
<td>(1.11-1.61)</td>
<td>(0.93-1.40)</td>
<td>(0.90-1.50)</td>
<td>(1.17-2.82)</td>
<td>(0.93-1.81)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-5292.65**</td>
<td>-3992.*</td>
<td>-2598.64*</td>
<td>-1212.12*</td>
<td>-1489.67</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.24</td>
<td>1.23</td>
<td>1.04</td>
<td>1.71</td>
<td>1.18</td>
</tr>
<tr>
<td>High normal</td>
<td>(1.07-1.45)</td>
<td>(1.03-1.47)</td>
<td>(0.84-1.30)</td>
<td>(1.21-2.40)</td>
<td>(0.88-1.57)</td>
</tr>
<tr>
<td>High</td>
<td>1.81</td>
<td>1.84</td>
<td>1.83</td>
<td>2.37</td>
<td>2.20</td>
</tr>
<tr>
<td>(1.46-2.25)</td>
<td>(1.44-2.36)</td>
<td>(1.37-2.46)</td>
<td>(1.50-3.75)</td>
<td>(1.50-3.22)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-5285.96***</td>
<td>-3984.34***</td>
<td>-2490.93***</td>
<td>-1208.40***</td>
<td>-1484.51***</td>
</tr>
<tr>
<td><strong>SBP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.36</td>
<td>1.47</td>
<td>1.46</td>
<td>1.41</td>
<td>1.68</td>
</tr>
<tr>
<td>High normal</td>
<td>(1.14-1.62)</td>
<td>(1.20-1.80)</td>
<td>(1.13-1.88)</td>
<td>(0.95-2.09)</td>
<td>(1.18-2.41)</td>
</tr>
<tr>
<td>High</td>
<td>2.30</td>
<td>1.98</td>
<td>1.97</td>
<td>3.27</td>
<td>3.76</td>
</tr>
<tr>
<td>(1.88-2.81)</td>
<td>(1.56-2.52)</td>
<td>(1.47-2.66)</td>
<td>(2.15-4.97)</td>
<td>(2.56-5.52)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-5266.31***</td>
<td>-3979.26***</td>
<td>-2591.92***</td>
<td>-1199.14***</td>
<td>-1467.19***</td>
</tr>
<tr>
<td><strong>DBP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.18</td>
<td>1.07</td>
<td>1.15</td>
<td>1.75</td>
<td>1.37</td>
</tr>
<tr>
<td>High normal</td>
<td>(1.00-1.39)</td>
<td>(0.88-1.30)</td>
<td>(0.91-1.45)</td>
<td>(1.20-2.55)</td>
<td>(1.00-1.89)</td>
</tr>
<tr>
<td>High</td>
<td>2.01</td>
<td>1.85</td>
<td>1.69</td>
<td>3.41</td>
<td>2.69</td>
</tr>
<tr>
<td>(1.68-2.39)</td>
<td>(1.51-2.26)</td>
<td>(1.31-2.18)</td>
<td>(2.33-5.01)</td>
<td>(1.94-3.74)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-5261.44***</td>
<td>-3975.32***</td>
<td>-2593.80**</td>
<td>-1196.04***</td>
<td>-1467.16***</td>
</tr>
<tr>
<td><strong>BP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.34</td>
<td>1.42</td>
<td>1.42</td>
<td>1.46</td>
<td>1.48</td>
</tr>
<tr>
<td>High normal</td>
<td>(1.11-1.62)</td>
<td>(1.14-1.78)</td>
<td>(1.08-1.87)</td>
<td>(0.94-2.27)</td>
<td>(1.01-2.17)</td>
</tr>
<tr>
<td>High</td>
<td>2.19</td>
<td>2.09</td>
<td>1.96</td>
<td>3.17</td>
<td>3.22</td>
</tr>
<tr>
<td>(1.80-2.68)</td>
<td>(1.65-2.65)</td>
<td>(1.47-2.63)</td>
<td>(2.04-4.91)</td>
<td>(2.19-4.74)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-5265.46***</td>
<td>-3974.87***</td>
<td>-2591.32***</td>
<td>-1198.46***</td>
<td>-1468.76***</td>
</tr>
<tr>
<td><strong>SCL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.53</td>
<td>1.56</td>
<td>1.43</td>
<td>1.34</td>
<td>0.84</td>
</tr>
<tr>
<td>High normal</td>
<td>(1.22-1.93)</td>
<td>(1.18-2.07)</td>
<td>(1.02-2.01)</td>
<td>(0.84-2.12)</td>
<td>(0.56-1.25)</td>
</tr>
<tr>
<td>High</td>
<td>2.22</td>
<td>2.48</td>
<td>2.08</td>
<td>1.45</td>
<td>1.50</td>
</tr>
<tr>
<td>(1.76-2.79)</td>
<td>(1.88-3.27)</td>
<td>(1.49-2.91)</td>
<td>(0.91-2.31)</td>
<td>(1.02-2.19)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-5270.71***</td>
<td>-3966.47**</td>
<td>-2589.77***</td>
<td>-1215.17</td>
<td>-1483.43**</td>
</tr>
</tbody>
</table>

*p-value: *<0.05; **<0.01; ***<0.001
## Continuation of Table 8.2...

### Female

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.00</td>
<td>0.84</td>
<td>1.14</td>
<td>1.29</td>
<td>1.10</td>
</tr>
<tr>
<td>(0.80-1.25)</td>
<td>(0.62-1.13)</td>
<td>(0.76-1.72)</td>
<td>(0.86-1.94)</td>
<td>(0.73-1.66)</td>
<td></td>
</tr>
<tr>
<td>Always smoker</td>
<td>1.14</td>
<td>1.04</td>
<td>1.49</td>
<td>1.34</td>
<td>1.39</td>
</tr>
<tr>
<td>(0.98-1.32)</td>
<td>(0.85-1.26)</td>
<td>(1.15-1.94)</td>
<td>(1.01-1.77)</td>
<td>(1.07-1.81)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-5422.49</td>
<td>-3290.08</td>
<td>-1741.75*</td>
<td>-1570.40</td>
<td>-1753.59*</td>
</tr>
</tbody>
</table>

| **BMI** |     |     |    |        |     |
| Elevated/ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Always smoker| 1.14 | 1.04 | 1.49 | 1.34 | 1.39 |
| (0.98-1.32)  | (0.85-1.26) | (1.15-1.94) | (1.01-1.77) | (1.07-1.81) |
| **Log likelihood** | -5405.10*** | -3274.87** | -1734.46*** | -1569.86 | -1739.66*** |

| **SBP** |     |     |    |        |     |
| Elevated/ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Always smoker| 1.14 | 1.04 | 1.49 | 1.34 | 1.39 |
| (0.98-1.32)  | (0.85-1.26) | (1.15-1.94) | (1.01-1.77) | (1.07-1.81) |
| **Log likelihood** | -5374.76*** | -3261.48** | -1724.91*** | -1562.90** | -1730.60*** |

| **DBP** |     |     |    |        |     |
| Elevated/ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Always smoker| 1.14 | 1.04 | 1.49 | 1.34 | 1.39 |
| (0.98-1.32)  | (0.85-1.26) | (1.15-1.94) | (1.01-1.77) | (1.07-1.81) |
| **Log likelihood** | -5369.59*** | -3247.56** | -1733.11*** | -1555.34*** | -1735.79*** |

| **BP** |     |     |    |        |     |
| Elevated/ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Always smoker| 1.14 | 1.04 | 1.49 | 1.34 | 1.39 |
| (0.98-1.32)  | (0.85-1.26) | (1.15-1.94) | (1.01-1.77) | (1.07-1.81) |
| **Log likelihood** | -5373.37*** | -3261.41*** | -1728.73*** | -1563.08** | -1730.82*** |

| **SCL** |     |     |    |        |     |
| Elevated/ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Always smoker| 1.14 | 1.04 | 1.49 | 1.34 | 1.39 |
| (0.98-1.32)  | (0.85-1.26) | (1.15-1.94) | (1.01-1.77) | (1.07-1.81) |
| **Log likelihood** | -5412.77*** | -3275.07*** | -1738.73*** | -1569.92 | -1672.47* |

*p-value: *<0.05; **<0.01; ***<0.001
Multivariate analysis

We studied the relative risk of cardiovascular disease or death associated with a risk factor adjusted for the presence of other risk factors in the multivariate analysis. We assumed that risk factors act independently, which meant that the level of a risk factor had no influence on the levels of the other risk factors that were present. The absence of interaction between risk factors is an assumption that was made because of inadequate sample size. The number of observations was not sufficient to study interaction effects between risk factors. Table 8.3 shows the relative risk of CVD and its subtypes associated with each risk factor when the effect of the risk factor is controlled or adjusted for the level of other risk factors. Results are presented separately for males and females.

Consider males. Always smoking increased the risk of CVD by 34% when the level of other risk factors (RR is 1.34 in univariate analysis) was not controlled for. When we controlled for the level of other risk factors, it increased to 44%, which is an upward CIER of 29%. The upward CIER is 28% for CHF and 33% for stroke. The RR of CVD and its subtypes for obese males was lower when the levels of other risk factors were controlled for than when they were not. The univariate analysis yielded an RR of 1.81; the multivariate analysis produced an RR of 1.48, which is a 41% downward CIER, indicating that part of the effect of obesity could be attributed to other risk factors. Obese males were likely to have unfavorable levels of other risk factors. Otherwise than might be expected, being obese did not increase the RR of CVD, as long as the other risk factors were at their optimal level. The downward CIER was 36% for MI and 58% for CHF. Similarly, for hypertension the downward CIER was 16% for CHF and 31% for CHD. For high cholesterol level, the downward CIER was about 20% for CVD, CHD or MI.

Similarly, when we controlled for the level of other risk factors for female smokers, the upward CIER was 53% for stroke and 136% for CVD. The RR of CVD and its subtypes for obese females was lower when the levels of other risk factors were controlled for than when they were not, indicating that part of the effect of obesity could be attributed to other risk factors. The downward CIER ranged from 32% for MI to 53% for CVD. Similarly, the downward CIER for hypertension was about 2% for stroke and 31% for CHD. For high cholesterol level, the downward CIER was about 28% for MI and 37% for CVD.
Table 8.3 Relative risk of cardiovascular disease and its subtypes (including sudden death) by multiple risk factor (95% confidence intervals in parentheses) status during age 30-50

Male

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.09</td>
<td>0.89</td>
<td>0.83</td>
<td>2.05 ( +22%)</td>
<td>0.96</td>
</tr>
<tr>
<td>(0.85-1.40)</td>
<td></td>
<td>(0.67-1.19)</td>
<td>(0.58-1.20)</td>
<td>(1.21-3.47)</td>
<td>(0.60-1.55)</td>
</tr>
<tr>
<td>Always smoker</td>
<td>1.44 ( +29%)</td>
<td>1.21</td>
<td>1.24</td>
<td>2.06 ( +28%)</td>
<td>1.40 ( +33%)</td>
</tr>
<tr>
<td>(1.20-1.74)</td>
<td></td>
<td>(0.98-1.49)</td>
<td>(0.96-1.60)</td>
<td>(1.32-3.19)</td>
<td>(1.00-1.95)</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.15</td>
<td>1.13</td>
<td>0.98</td>
<td>1.52 ( -27%)</td>
<td>1.02</td>
</tr>
<tr>
<td>High normal</td>
<td>(0.99-1.35)</td>
<td>(0.94-1.36)</td>
<td>(0.77-1.21)</td>
<td>(1.07-2.15)</td>
<td>(0.76-1.38)</td>
</tr>
<tr>
<td>High</td>
<td>1.48 ( -41%)</td>
<td>1.49 ( -42%)</td>
<td>1.53 ( -36%)</td>
<td>1.82 ( -40%)</td>
<td>1.51 ( -58%)</td>
</tr>
<tr>
<td>(1.18-1.85)</td>
<td></td>
<td>(1.15-1.93)</td>
<td>(1.19-2.08)</td>
<td>(1.12-2.95)</td>
<td>(1.08-2.26)</td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.25 ( -26%)</td>
<td>1.31 ( -26%)</td>
<td>1.35 ( -17%)</td>
<td>1.34</td>
<td>1.42</td>
</tr>
<tr>
<td>High normal</td>
<td>(1.03-1.52)</td>
<td>(1.04-1.64)</td>
<td>(1.02-1.79)</td>
<td>(0.85-1.09)</td>
<td>(0.96-2.10)</td>
</tr>
<tr>
<td>High</td>
<td>1.89 ( -25%)</td>
<td>1.75 ( -31%)</td>
<td>1.69 ( -28%)</td>
<td>2.76 ( -19%)</td>
<td>2.86 ( -16%)</td>
</tr>
<tr>
<td>(1.54-2.33)</td>
<td></td>
<td>(1.37-2.24)</td>
<td>(1.25-2.30)</td>
<td>(1.74-4.36)</td>
<td>(1.90-4.30)</td>
</tr>
<tr>
<td>SCL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.48 ( -9%)</td>
<td>1.49 ( -13%)</td>
<td>1.36</td>
<td>1.31</td>
<td>0.77</td>
</tr>
<tr>
<td>high normal</td>
<td>(1.18-1.87)</td>
<td>(1.13-2.98)</td>
<td>(0.97-1.91)</td>
<td>(0.82-2.08)</td>
<td>(0.52-1.14)</td>
</tr>
<tr>
<td>High</td>
<td>1.97 ( -20%)</td>
<td>2.20 ( -19%)</td>
<td>1.85 ( -21%)</td>
<td>1.21</td>
<td>1.23</td>
</tr>
<tr>
<td>(1.56-2.49)</td>
<td></td>
<td>(1.67-2.92)</td>
<td>(1.32-2.60)</td>
<td>(0.76-1.94)</td>
<td>(0.83-1.81)</td>
</tr>
</tbody>
</table>

Log likelihood -5230.19*** -3944.18*** -2573.45*** -1188.39*** -1457.43***

*p-value: *<0.05; **<0.01; ***<0.001*
Continuation of Table 8.3…

<table>
<thead>
<tr>
<th>Smoking status</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never smoker</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>(0.94-1.49)</td>
<td></td>
<td>(0.74-1.37)</td>
<td>(0.92-2.12)</td>
<td>(0.99-2.25)</td>
<td>(0.93-2.13)</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.18</td>
<td>1.00</td>
<td>1.40</td>
<td>1.49</td>
<td>1.41</td>
</tr>
<tr>
<td>(1.14-1.56)</td>
<td></td>
<td>(0.99-1.48)</td>
<td>(1.38-2.39)</td>
<td>(1.14-2.02)</td>
<td>(1.36-2.33)</td>
</tr>
<tr>
<td>Always smoker</td>
<td>1.33(+136%)</td>
<td>1.21</td>
<td>1.82(+67%)</td>
<td>1.52(+53%)</td>
<td>1.78(+100%)</td>
</tr>
<tr>
<td>(1.14-1.56)</td>
<td></td>
<td>(0.99-1.48)</td>
<td>(1.38-2.39)</td>
<td>(1.14-2.02)</td>
<td>(1.36-2.33)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BMI</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.13</td>
<td>1.19</td>
<td>1.04</td>
<td>1.22</td>
<td>1.43(-30%)</td>
</tr>
<tr>
<td>high normal</td>
<td>(0.96-1.32)</td>
<td>(0.97-1.46)</td>
<td>(0.78-1.39)</td>
<td>(0.92-1.64)</td>
<td>(1.08-1.90)</td>
</tr>
<tr>
<td>High</td>
<td>1.43(-53%)</td>
<td>1.60(-46%)</td>
<td>1.93(-32%)</td>
<td>1.11</td>
<td>2.04(-40%)</td>
</tr>
<tr>
<td>(1.15-1.79)</td>
<td>(1.21-2.11)</td>
<td>(1.34-2.77)</td>
<td>(0.72-1.73)</td>
<td>(1.42-2.94)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BP</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.52(-13%)</td>
<td>1.47(-24%)</td>
<td>1.76(-11%)</td>
<td>1.29</td>
<td>1.45(-18%)</td>
</tr>
<tr>
<td>high normal</td>
<td>(1.27-1.82)</td>
<td>(1.15-1.86)</td>
<td>(1.26-2.47)</td>
<td>(0.92-1.79)</td>
<td>(1.03-2.04)</td>
</tr>
<tr>
<td>High</td>
<td>2.34(-16%)</td>
<td>2.08(-31%)</td>
<td>2.35(-13%)</td>
<td>2.08(-2%)</td>
<td>2.64(-23%)</td>
</tr>
<tr>
<td>(1.91-2.88)</td>
<td>(1.59-2.72)</td>
<td>(1.61-3.44)</td>
<td>(1.44-3.02)</td>
<td>(1.84-3.80)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SCL</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.09</td>
<td>1.44(-10%)</td>
<td>1.12</td>
<td>0.79</td>
<td>0.85</td>
</tr>
<tr>
<td>high normal</td>
<td>(0.89-1.35)</td>
<td>(1.07-1.94)</td>
<td>(0.76-1.66)</td>
<td>(0.54-1.13)</td>
<td>(0.60-1.22)</td>
</tr>
<tr>
<td>High</td>
<td>1.32(-37%)</td>
<td>1.83(-26%)</td>
<td>1.57(-28%)</td>
<td>1.01</td>
<td>1.06</td>
</tr>
<tr>
<td>(1.07-1.62)</td>
<td>(1.37-2.46)</td>
<td>(1.07-2.28)</td>
<td>(0.71-1.44)</td>
<td>(0.75-1.50)</td>
<td></td>
</tr>
</tbody>
</table>

Log likelihood: -5358.38*** -3245.01*** -1710.11*** -1556.23** -1715.32***
p-value: *<0.05; **<0.01; ***<0.001

8.3.2 Death

The risk of death after having experienced cardiovascular disease is heavily dependent on the severity of the disease. The transition to death is classified into three categories: alive to death (irrespective of disease status), free of disease to death and post-disease death. Here, alive to death transitions are considered, irrespective of any cardiovascular diseases. Using a two-state model (model 3.1(a)), the risk factor status at middle age and the burden on overall mortality is estimated in terms of the relative risk of dying. The transition from free of disease to death refers to the death that occurs without experiencing any cardiovascular disease. Post-disease death is the death that occurs after having experienced any cardiovascular disease. Both the univariate and multivariate analysis was performed for all three types of transitions. The effect of risk factors on three different types of transitions to death might not be same.

---

3 Cause of death is not necessarily cardiovascular disease


Univariate analysis
The relative risk of dying for each single risk factor category is presented in Table 8.4. For males, the rates of death (any type, except from MI or stroke) for smokers were considerably higher. The relative risk was 1.63 (CI: 1.39-1.99) for the transition from alive to death, 1.84 (CI: 1.34-2.52) for free of CVD to death and 1.24 (CI: 1.01-1.54) for post-disease death. The effect of obesity on overall mortality was significantly higher (RR=1.47, CI: 1.20-1.80) compared to normal BMI. The relative risk of death, whether from alive or from CVD, from CHD or from MI was significantly higher for the males with high SBP, DBP or BP at middle age. For males with high BP, this ranged from 1.33 (CI: 1.00-1.85) for MI to death to 1.85 for alive to death. Cholesterol levels at middle age and mortality at older ages were only significant for post-CHF death (RR=1.57, CI: 1.03-2.41).

For female always smokers, the risk of any type of death except from MI was significantly higher than for non-smokers. RR ranged from 1.34 (CI: 0.99-1.83) for CHF to 1.85 (CI: 1.45-2.34) for CHD. Like males, the overall mortality (irrespective of disease status) in the obese females was significantly higher (RR=1.54, CI: 1.29-1.85) than the normal BMI group. The effect of high SBP was significantly larger for the transition alive to death, CVD to death or CHD to death. The effect of elevated or high DBP and BP was significantly higher only for overall deaths (irrespective of disease status). For females, high cholesterol levels at middle age had significant impact on overall mortality and post-disease deaths. The RR ranged from 1.32 (CI: 1.10-1.58) for alive to death to 1.95 (CI: 1.28-2.96) for death from CHF. High normal cholesterol levels also had a significant impact on post-stroke deaths (RR=1.75, CI: 1.12-2.74) or CHF (RR=1.57, CI: 1.01-2.43).
A MULTIVARIATE ANALYSIS OF CARDIOVASCULAR RISK FACTOR 203

Table 8.4 Relative risk of deaths (total death and post-disease death) by single risk factor (95% confidence intervals in parentheses) status during age 30-50

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Alive</th>
<th>NO-CVD</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.12</td>
<td>1.34</td>
<td>1.00</td>
<td>0.92</td>
<td>0.74</td>
<td>1.20</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>(0.88-1.43)</td>
<td>(0.74-1.34)</td>
<td>(0.65-1.29)</td>
<td>(0.48-1.13)</td>
<td>(0.67-2.18)</td>
<td>(0.47-1.37)</td>
<td></td>
</tr>
<tr>
<td>Always smoker</td>
<td>1.63</td>
<td>1.84</td>
<td>1.24</td>
<td>1.20</td>
<td>0.91</td>
<td>1.03</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>(1.39-1.99)</td>
<td>(1.01-1.54)</td>
<td>(0.95-1.52)</td>
<td>(0.69-1.21)</td>
<td>(0.65-1.66)</td>
<td>(0.96-2.05)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td><strong>-6202.52</strong>*</td>
<td><strong>-2078.23</strong></td>
<td><strong>-3345.88</strong></td>
<td><strong>-2390.01</strong></td>
<td><strong>-1409.08</strong></td>
<td><strong>-540.22</strong></td>
<td><strong>-640.20</strong>*</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.08</td>
<td>0.84</td>
<td>1.00</td>
<td>1.02</td>
<td>0.97</td>
<td>0.81</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>(0.94-1.24)</td>
<td>(0.84-1.19)</td>
<td>(0.83-1.25)</td>
<td>(0.76-1.24)</td>
<td>(0.56-1.18)</td>
<td>(0.60-1.13)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1.47</td>
<td>1.07</td>
<td>1.03</td>
<td>1.11</td>
<td>1.05</td>
<td>0.74</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>(1.20-1.80)</td>
<td>(0.81-1.32)</td>
<td>(0.84-1.47)</td>
<td>(0.76-1.46)</td>
<td>(0.44-1.22)</td>
<td>(0.46-1.11)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td><strong>-6216.73</strong></td>
<td><strong>-2085.62</strong></td>
<td><strong>-3349.09</strong></td>
<td><strong>-2392.27</strong></td>
<td><strong>-1409.97</strong></td>
<td><strong>-539.60</strong></td>
<td><strong>-643.19</strong></td>
</tr>
<tr>
<td><strong>SBP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.17*</td>
<td>0.91</td>
<td>1.08</td>
<td>1.23</td>
<td>1.13</td>
<td>0.79</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>(0.94-1.28)</td>
<td>(0.88-1.13)</td>
<td>(0.82-1.31)</td>
<td>(0.84-1.18)</td>
<td>(0.51-1.22)</td>
<td>(0.39-0.86)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>2.05</td>
<td>1.12</td>
<td>1.59</td>
<td>1.87</td>
<td>1.69</td>
<td>1.13</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>(1.71-2.45)</td>
<td>(1.27-1.99)</td>
<td>(1.43-2.44)</td>
<td>(1.21-2.36)</td>
<td>(0.71-1.80)</td>
<td>(0.57-1.33)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td><strong>-6190.84</strong>*</td>
<td><strong>-2086.07</strong></td>
<td><strong>-3349.51</strong></td>
<td><strong>-2381.19</strong></td>
<td><strong>-1404.44</strong></td>
<td><strong>-538.46</strong></td>
<td><strong>-639.47</strong></td>
</tr>
<tr>
<td><strong>DBP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.11</td>
<td>0.94</td>
<td>1.08</td>
<td>1.16</td>
<td>1.28</td>
<td>0.92</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>(0.95-1.28)</td>
<td>(0.89-1.32)</td>
<td>(0.94-1.44)</td>
<td>(0.98-1.66)</td>
<td>(0.61-1.42)</td>
<td>(0.53-1.09)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1.80</td>
<td>1.08</td>
<td>1.40</td>
<td>1.42</td>
<td>1.43</td>
<td>1.31</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>(1.53-2.12)</td>
<td>(1.15-1.70)</td>
<td>(1.13-1.77)</td>
<td>(1.08-1.90)</td>
<td>(0.86-1.99)</td>
<td>(0.59-1.23)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td><strong>-6181.88</strong>*</td>
<td><strong>-2079.95</strong></td>
<td><strong>-3339.56</strong></td>
<td><strong>-2387.85</strong></td>
<td><strong>-1406.76</strong></td>
<td><strong>-538.71</strong></td>
<td><strong>-642.72</strong></td>
</tr>
<tr>
<td><strong>BP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.15</td>
<td>0.94</td>
<td>1.09</td>
<td>1.20</td>
<td>1.03</td>
<td>0.97</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>(0.97-1.37)</td>
<td>(0.87-1.36)</td>
<td>(0.93-1.56)</td>
<td>(0.75-1.42)</td>
<td>(0.59-1.60)</td>
<td>(0.42-0.97)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1.85</td>
<td>1.01</td>
<td>1.46</td>
<td>1.52</td>
<td>1.33</td>
<td>1.33</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>(1.55-2.21)</td>
<td>(1.16-1.83)</td>
<td>(1.16-1.98)</td>
<td>(1.00-1.85)</td>
<td>(0.81-2.17)</td>
<td>(0.47-1.10)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td><strong>-6194.29</strong>*</td>
<td><strong>-2086.89</strong></td>
<td><strong>-3341.45</strong></td>
<td><strong>-2387.15</strong></td>
<td><strong>-1407.65</strong></td>
<td><strong>-538.78</strong></td>
<td><strong>-642.42</strong></td>
</tr>
<tr>
<td><strong>SCL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.03</td>
<td>0.70</td>
<td>0.90</td>
<td>1.09</td>
<td>1.00</td>
<td>1.28</td>
<td>2.12</td>
</tr>
<tr>
<td></td>
<td>(0.85-1.24)</td>
<td>(0.70-1.17)</td>
<td>(0.79-1.50)</td>
<td>(0.68-1.48)</td>
<td>(0.76-2.01)</td>
<td>(1.36-3.30)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1.18</td>
<td>0.58</td>
<td>0.82</td>
<td>1.01</td>
<td>1.01</td>
<td>1.25</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>(0.97-1.42)</td>
<td>(0.64-1.07)</td>
<td>(0.73-1.39)</td>
<td>(0.69-1.50)</td>
<td>(0.74-2.12)</td>
<td>(1.03-2.41)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td><strong>-6220.93</strong></td>
<td><strong>-2081.00</strong></td>
<td><strong>-3347.86</strong></td>
<td><strong>2392.24</strong></td>
<td><strong>-1410.10</strong></td>
<td><strong>-540.01</strong></td>
<td><strong>-638.49</strong></td>
</tr>
</tbody>
</table>

*p-value: *<0.05; **<0.01; ***<0.001
Continuation of Table 8.4...

Female

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Alive</th>
<th>NO-CVD</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.09</td>
<td>1.24</td>
<td>0.98</td>
<td>1.13</td>
<td>1.03</td>
<td>0.69</td>
<td>1.60</td>
</tr>
<tr>
<td>(0.87-1.36)</td>
<td>(0.90-1.71)</td>
<td>(0.72-1.34)</td>
<td>(0.76-1.69)</td>
<td>(0.63-1.69)</td>
<td>(0.40-1.22)</td>
<td>(0.98-2.62)</td>
<td></td>
</tr>
<tr>
<td>Always smoker</td>
<td>1.73</td>
<td>1.93</td>
<td>1.68</td>
<td>1.85</td>
<td>1.25</td>
<td>1.64</td>
<td>1.34</td>
</tr>
<tr>
<td>(1.51-1.98)</td>
<td>(1.59-2.36)</td>
<td>(1.40-2.02)</td>
<td>(1.45-2.34)</td>
<td>(0.91-1.72)</td>
<td>(1.19-2.26)</td>
<td>(0.99-1.83)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-6643.01***</td>
<td>-3034.25***</td>
<td>-2824.33**</td>
<td>-1545.66**</td>
<td>-725.18</td>
<td>-632.80**</td>
<td>-691.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BMI</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.06</td>
<td>0.92</td>
<td>0.93</td>
<td>0.78</td>
<td>0.88</td>
<td>1.23</td>
<td>1.17</td>
</tr>
<tr>
<td>High normal</td>
<td>(0.93-1.22)</td>
<td>(0.75-1.13)</td>
<td>(0.77-1.12)</td>
<td>(0.61-1.00)</td>
<td>(0.63-1.23)</td>
<td>(0.89-1.71)</td>
<td>(0.85-1.61)</td>
</tr>
<tr>
<td>High</td>
<td>1.54</td>
<td>1.11</td>
<td>1.15</td>
<td>1.06</td>
<td>1.26</td>
<td>1.00</td>
<td>1.34</td>
</tr>
<tr>
<td>(1.29-1.85)</td>
<td>(0.83-1.49)</td>
<td>(0.91-1.45)</td>
<td>(0.79-1.43)</td>
<td>(0.87-1.82)</td>
<td>(0.63-1.59)</td>
<td>(0.92-1.96)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-6666.37***</td>
<td>-3054.86</td>
<td>-2840.06</td>
<td>-1555.54</td>
<td>-724.70</td>
<td>-639.37</td>
<td>-693.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SBP</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.27</td>
<td>1.04</td>
<td>0.97</td>
<td>0.94</td>
<td>1.86</td>
<td>1.11</td>
<td>1.07</td>
</tr>
<tr>
<td>High normal</td>
<td>(1.09-1.47)</td>
<td>(0.84-1.28)</td>
<td>(0.78-1.20)</td>
<td>(0.70-1.25)</td>
<td>(0.58-1.28)</td>
<td>(0.75-1.64)</td>
<td>(0.73-1.58)</td>
</tr>
<tr>
<td>High</td>
<td>1.90</td>
<td>1.11</td>
<td>1.15</td>
<td>1.06</td>
<td>1.26</td>
<td>1.00</td>
<td>1.34</td>
</tr>
<tr>
<td>(1.61-2.23)</td>
<td>(0.86-1.42)</td>
<td>(1.08-1.69)</td>
<td>(1.09-1.95)</td>
<td>(0.95-2.09)</td>
<td>(0.83-1.84)</td>
<td>(0.92-1.98)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-6646.84***</td>
<td>-3055.30</td>
<td>-2840.06</td>
<td>-1555.54</td>
<td>-724.70</td>
<td>-639.37</td>
<td>-692.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DBP</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.23</td>
<td>0.97</td>
<td>1.01</td>
<td>0.95</td>
<td>0.91</td>
<td>1.05</td>
<td>1.07</td>
</tr>
<tr>
<td>High normal</td>
<td>(1.07-1.42)</td>
<td>(0.80-1.23)</td>
<td>(0.83-1.23)</td>
<td>(0.73-1.24)</td>
<td>(0.64-1.28)</td>
<td>(0.74-1.50)</td>
<td>(0.76-1.51)</td>
</tr>
<tr>
<td>High</td>
<td>1.69</td>
<td>1.14</td>
<td>1.16</td>
<td>1.20</td>
<td>1.39</td>
<td>1.02</td>
<td>1.02</td>
</tr>
<tr>
<td>(1.44-1.99)</td>
<td>(0.89-1.47)</td>
<td>(0.94-1.43)</td>
<td>(0.91-1.57)</td>
<td>(0.98-1.98)</td>
<td>(0.68-1.48)</td>
<td>(0.72-1.43)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-6648.40***</td>
<td>-3055.35</td>
<td>-2840.06</td>
<td>-1555.57</td>
<td>723.32*</td>
<td>636.30</td>
<td>694.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BP</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.28</td>
<td>1.04</td>
<td>1.05</td>
<td>1.05</td>
<td>0.90</td>
<td>1.32</td>
<td>1.09</td>
</tr>
<tr>
<td>High normal</td>
<td>(1.10-1.50)</td>
<td>(0.84-1.29)</td>
<td>(0.84-1.32)</td>
<td>(0.77-1.41)</td>
<td>(0.60-1.35)</td>
<td>(0.87-1.98)</td>
<td>(0.73-1.63)</td>
</tr>
<tr>
<td>High</td>
<td>1.83</td>
<td>1.10</td>
<td>1.23</td>
<td>1.28</td>
<td>1.41</td>
<td>1.07</td>
<td>1.24</td>
</tr>
<tr>
<td>(1.55-2.15)</td>
<td>(0.86-1.41)</td>
<td>(0.98-1.55)</td>
<td>(0.94-1.72)</td>
<td>(0.94-2.10)</td>
<td>(0.71-1.62)</td>
<td>(0.84-1.84)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-6649.92***</td>
<td>-3055.27</td>
<td>-2839.34</td>
<td>-1556.50</td>
<td>-722.34*</td>
<td>-639.12</td>
<td>-693.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SCL</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/</td>
<td>1.07</td>
<td>0.98</td>
<td>1.25</td>
<td>1.16</td>
<td>1.41</td>
<td>1.75</td>
<td>1.57</td>
</tr>
<tr>
<td>High normal</td>
<td>(0.89-1.29)</td>
<td>(0.75-1.26)</td>
<td>(0.95-1.64)</td>
<td>(0.79-1.70)</td>
<td>(0.87-2.28)</td>
<td>(1.12-2.74)</td>
<td>(1.01-2.43)</td>
</tr>
<tr>
<td>High</td>
<td>1.32</td>
<td>0.98</td>
<td>1.43</td>
<td>1.47</td>
<td>1.33</td>
<td>1.62</td>
<td>1.95</td>
</tr>
<tr>
<td>(1.10-1.58)</td>
<td>(0.75-1.27)</td>
<td>(1.10-1.85)</td>
<td>(1.02-2.12)</td>
<td>(0.83-2.11)</td>
<td>(1.07-2.48)</td>
<td>(1.28-2.96)</td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>-6670.01**</td>
<td>-3055.57</td>
<td>-2837.36*</td>
<td>-1555.00</td>
<td>-725.17</td>
<td>-636.59*</td>
<td>-689.11**</td>
</tr>
</tbody>
</table>

*p-value: *<0.05; **<0.01; ***<0.001*
Multivariate analysis

The multivariate analysis of mortality describes the change in excess risks of death that are associated with a risk factor, after controlling for the levels of the other risk factors. Male always smokers had an excess risk of death (irrespective of disease status) of 63% without controlling for the levels of other risk factors and 72% if the levels of the other risk factors were controlled for. When other risk factors were controlled for, the CIER of death for always smoking was 14% upward. The upward CIER of death for always-smoking males with CVD was 25% after controlling for the levels of other risk factors (RR increased from 1.24 to 1.30). For obesity, the effect of controlling for the levels of risk factors was different. The RR of death, irrespective of whether the person had CVD or not, declined from 1.47 in the absence of any control to 1.27 if the levels of other risk factors were controlled for. The downward CIER of mortality was 43%.

Table 8.5 Relative risk of deaths (total death and post-disease death) by multiple risk factor (95% confidence intervals in parentheses) status during age 30-50

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Alive</th>
<th>NO-CVD</th>
<th>CVD</th>
<th>CHD</th>
<th>MI</th>
<th>Stroke</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>1.17</td>
<td>1.31</td>
<td>1.04</td>
<td>0.98</td>
<td>0.73</td>
<td>1.55</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>(0.92-1.49)</td>
<td>(0.86-1.99)</td>
<td>(0.77-1.40)</td>
<td>(0.70-1.38)</td>
<td>(0.47-1.13)</td>
<td>(0.83-2.90)</td>
<td>(0.48-1.44)</td>
</tr>
<tr>
<td>Always smoker</td>
<td>1.72(+14%)</td>
<td>1.86 (+2%)</td>
<td>1.30(+33%)</td>
<td>1.26(+30%)</td>
<td>0.94</td>
<td>1.12</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>(1.44-2.05)</td>
<td>(1.35-2.55)</td>
<td>(1.05-1.62)</td>
<td>(1.00-1.60)</td>
<td>(0.71-1.25)</td>
<td>(0.68-1.84)</td>
<td>(0.80-1.80)</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/high normal</td>
<td>1.03</td>
<td>0.89</td>
<td>0.94</td>
<td>0.92</td>
<td>0.89</td>
<td>0.71</td>
<td>0.86</td>
</tr>
<tr>
<td>High</td>
<td>1.27 (+42%)</td>
<td>1.18</td>
<td>0.91</td>
<td>0.92</td>
<td>0.85</td>
<td>0.57</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>(1.03-1.58)</td>
<td>(0.81-1.56)</td>
<td>(0.70-1.18)</td>
<td>(0.68-1.25)</td>
<td>(0.59-1.23)</td>
<td>(0.33-1.00)</td>
<td>(0.47-1.21)</td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/high normal</td>
<td>1.16</td>
<td>1.02</td>
<td>1.16</td>
<td>1.27</td>
<td>1.07</td>
<td>1.13</td>
<td>0.69</td>
</tr>
<tr>
<td>High</td>
<td>1.81 (+5%)</td>
<td>1.13</td>
<td>1.59(+28%)</td>
<td>1.64(+23%)</td>
<td>1.4(+30%)</td>
<td>1.57</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>(1.50-2.19)</td>
<td>(0.81-1.56)</td>
<td>(1.25-2.03)</td>
<td>(1.22-2.20)</td>
<td>(1.00-2.06)</td>
<td>(0.92-2.68)</td>
<td>(0.49-1.27)</td>
</tr>
<tr>
<td>SCL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Elevated/high normal</td>
<td>0.97</td>
<td>0.69 (-3%)</td>
<td>0.84</td>
<td>1.00</td>
<td>0.95</td>
<td>1.41</td>
<td>1.93(-17%)</td>
</tr>
<tr>
<td>High</td>
<td>1.04</td>
<td>0.56 (-3%)</td>
<td>0.77 (-6%)</td>
<td>0.93</td>
<td>0.97</td>
<td>1.29</td>
<td>1.55</td>
</tr>
<tr>
<td></td>
<td>(0.86-1.26)</td>
<td>(0.41-0.76)</td>
<td>(0.59-1.00)</td>
<td>(0.68-1.29)</td>
<td>(0.65-1.44)</td>
<td>(0.74-2.24)</td>
<td>(0.98-2.44)</td>
</tr>
</tbody>
</table>

Log likelihood = \(-6167.58^{***} \quad -2070.14^{***} \quad -3335.11^{**} \quad -2383.44^* \quad -1406.18 \quad -535.82 \quad -633.84^{**}\)

*p-value: \(<0.05; **<0.01; ***<0.001*
Similarly, female always smokers had an excess risk of death of 73% if the levels of other risk factors were not considered and 94% if the levels of the other risk factors were controlled for. The upward CIER of death for females who always smoke and have suffered a stroke was 27% after controlling for the levels of other risk factors. The downward CIER of death, irrespective of disease status for females who are obese, was 39% after controlling for the levels of other risk factors.

8.4 Discussion

In this chapter, we performed both univariate and multivariate analyses of the major cardiovascular risk factors in middle age and their effects on cardiovascular disease, its subtypes and mortality over a long follow-up. Both for men and women, the effect of being in a high risk category between the ages of 30 and 50 and the disease incidence in older ages of life was highly significant. It was also observed that the impact of moderately normal risk profiles at middle age also had a significant effect on the occurrence of CVD and its subtypes. This reinforces Grundy et al., (1999), who stated that while the "short-term risk may not be high in
adult population who have multiple risk factors of only moderate severity, long-term risk can be unacceptably high. The relative risk of experiencing cardiovascular disease and its subtypes was the highest for men or women who had high SBP or hypertension at middle age compared to being in high-risk of other risk factors. The relative risk of experiencing CVD or CHD or MI was also higher for males who had high cholesterol levels at middle life. For females, the second highest important risk factor to experience CVD or CHD was the DBP.

In this chapter, we investigated the long-term nature of relationships, not only of primary relationships as shown in univariate analyses, but also of confounding relationships as shown in the changes of relative risk after adjusting for the levels of other risk factors. Both for men and women, the adjusted relative risk and the risk of experiencing congestive heart failure were consistent with the previous study by Kenchaiah et al., (2002), who examined the relation of BMI to the risk of heart failure. In the multivariate model, the estimated risk of coronary heart disease by blood pressure and total cholesterol categories was consistent with the previous study (Wilson et al., 1998).

The association between the risk factors at middle age and post-disease death at older age is complex. However, in the multivariate model, the CIER of dying associated with smoking was upward; for obesity, downward. A male with high a cholesterol level was shown to have a significantly lower risk of post-disease death than a male with an optimal cholesterol level. By contrast, a female with high cholesterol had significantly higher post-disease (CHD, MI or CHF) mortality. While there have been reports that optimal cholesterol may be associated with increased mortality (Schatz et al., 2001), it is unclear as to what extent the results presented here support this or are merely due to chance.

The effect of a risk factor in a multiple risk factor context was controlled for the confounding effects of other risk factors. Therefore, the RR in multiple risk factor contexts could go up or down compared to the RR estimated in univariate analysis. For example, when we controlled for the level of other risk factors in male smokers the upward change in excess risk was 29% for CVD. The relative risk of CVD and its subtypes for obese males was lower when the levels of other risk factors were controlled for than when they were not. The direction in which the relative risk moves (e.g. upward for smoking and downward for obesity) is likely to be the result of the pattern of co-occurrence of risk factors in the population. In the FHS population, smokers were probably less exposed to other risk factors than non-smokers, which would explain why the RR goes up; obese people more than non-obese people, which is why the RR goes down. The substantive interpretation could be that smoking prevents obesity, and obesity leads to high serum cholesterol and high blood pressure.

We tested several interaction effects among the risk factors states. We found no risk factor interaction that significantly affected the occurrence of cardiovascular
disease and post-disease death rates. The main reason could be the small sample size. Although interactions between risk factors did exist, the sample size did not permit the study of the interaction effects. We analyzed the data separately for males and females since cardiovascular risk factors and gender differentials in all-cause and cardiovascular disease mortality are not same (Janghorbani et al., 1993; Castelli 1984; Thom et al., 1992; WHO MONICA Project, 1994).

The major strength of this chapter is that the Framingham Heart Study has accurate long follow-up of the cardiovascular disease incidence and mortality and the record of risk factor by biannual interview. Using this long-time follow-up information, it was possible to measure the impact of risk factor status in middle age and the disease incidence and mortality at older ages.

Upward or downward changes in excess risk of CVD and mortality affect the outcomes of life table analysis, which is the subject of chapter 9. For example, always smoking was shown to lead to a greater excess risk of CVD and mortality for males of normal weight, with an optimal blood pressure and an optimal cholesterol level than for average males. Therefore, a larger effect on life expectancy of smokers and non-smokers might be expected after adjusting for potential confounders than in the absence of adjustments. Similarly, the difference in expected lifetime between obese persons and persons with normal weight might be smaller when other risk factors are controlled for (multivariate analysis) than when they are not (univariate analysis).

The findings of this chapter highlight the need for early prevention of high levels of major cardiovascular risk factors. Appropriate prevention has the potential to bring about a significant reduction in long-term risk.

References


Technical Appendix 8.1

Risk factor characteristics, observed state occupancies, and number of possible transitions in FHS

<table>
<thead>
<tr>
<th>Male</th>
<th>Smoking status</th>
<th>BMI</th>
<th>SBP</th>
<th>DBP</th>
<th>BP</th>
<th>SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>E</td>
<td>A</td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td>Number</td>
<td>235</td>
<td>207</td>
<td>887</td>
<td>480</td>
<td>683</td>
<td>166</td>
</tr>
<tr>
<td>Mean</td>
<td>-</td>
<td>--</td>
<td>-</td>
<td>23</td>
<td>27</td>
<td>32</td>
</tr>
</tbody>
</table>

| Alive to death | 157 | 119 | 695 | 344 | 495 | 132| 229 | 490 | 252| 328 | 362 | 279| 193 | 440 | 338|

**CVD model**

| NoCVD to death | 47  | 42  | 235 | 135 | 151 | 38 | 98  | 165 | 61 | 126 | 131 | 66 | 84  | 158 | 82 |
| NoCVD to CVD   | 144 | 115 | 556 | 269 | 427 | 119| 180 | 423 | 212| 266 | 309 | 239| 149 | 377 | 289|
| CVD to death   | 110 | 77  | 460 | 209 | 344 | 94 | 131 | 325 | 191| 202 | 231 | 213| 109 | 282 | 256|

**CHD model**

| NoCHD to death | 63  | 64  | 353 | 191 | 235 | 54 | 137 | 226 | 117| 170 | 191 | 117| 116 | 214 | 150|
| NoCHD to CHD   | 115 | 80  | 408 | 196 | 313 | 94 | 126 | 331 | 146| 201 | 220 | 182| 104 | 288 | 211|
| CHD to death   | 94  | 55  | 342 | 153 | 260 | 78 | 92  | 264 | 135| 158 | 171 | 162| 77  | 226 | 188|

**MI model**

| NoMI to death | 92  | 87  | 473 | 236 | 338 | 78 | 169 | 317 | 166| 227 | 242 | 181| 141 | 292 | 219|
| NoMI to MI    | 77  | 48  | 267 | 136 | 190 | 66 | 82  | 216 | 94 | 130 | 152 | 110| 69  | 190 | 133|
| MI to death   | 65  | 32  | 222 | 108 | 157 | 54 | 60  | 173 | 86 | 101 | 120 | 98 | 52  | 148 | 119|

**Stroke model**

| NoCVM to death | 136 | 95  | 585 | 302 | 406 | 108| 201 | 418 | 197| 293 | 307 | 214| 172 | 380 | 264|
| NoCVM to CVM   | 24  | 34  | 126 | 48  | 106 | 30 | 35  | 89  | 60 | 43  | 73  | 68 | 27  | 77  | 80 |
| CVM to death   | 21  | 24  | 110 | 42  | 89  | 24 | 28  | 72  | 55 | 35  | 55  | 65 | 21  | 60  | 74 |

**CHF model**

| NoCHF to death | 117 | 96  | 551 | 275 | 390 | 99 | 194 | 389 | 181| 271 | 284 | 208| 163 | 354 | 247|
| NoCHF to CHF   | 46  | 28  | 161 | 76  | 118 | 41 | 40  | 119 | 76 | 66  | 89  | 79 | 35  | 100 | 100|
| CHF to death   | 40  | 23  | 144 | 69  | 105 | 33 | 35  | 101 | 71 | 57  | 78  | 71 | 30  | 86  | 91 | 28  | 76  | 103|

*N- never smoker; E- ever smoker; A- always smoker
++ I- optimal, II- high normal or elevated, III- high


A multivariate analysis of cardiovascular risk factor

<table>
<thead>
<tr>
<th>Female</th>
<th>Smoking status</th>
<th>BMI</th>
<th>SBP</th>
<th>DBP</th>
<th>BP</th>
<th>SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>E</td>
<td>A+</td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td>Number</td>
<td>829</td>
<td>220</td>
<td>667</td>
<td>973</td>
<td>543</td>
<td>200</td>
</tr>
<tr>
<td>Mean</td>
<td>-</td>
<td>--</td>
<td>-</td>
<td>22</td>
<td>27</td>
<td>34</td>
</tr>
</tbody>
</table>

| Alive to death | 457 | 96  | 446 | 518 | 330 | 151 |

<table>
<thead>
<tr>
<th>CVD model</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoCVD to death</td>
</tr>
<tr>
<td>NoCVD to CVD</td>
</tr>
<tr>
<td>CVD to death</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHD model</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoCHD to death</td>
</tr>
<tr>
<td>NoCHD to CHD</td>
</tr>
<tr>
<td>CHD to death</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MI model</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoMI to death</td>
</tr>
<tr>
<td>NoMI to MI</td>
</tr>
<tr>
<td>MI to death</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stroke model</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoCVM to death</td>
</tr>
<tr>
<td>NoCVM to CVM</td>
</tr>
<tr>
<td>CVM to death</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHF model</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoCHF to death</td>
</tr>
<tr>
<td>NoCHF to CHF</td>
</tr>
<tr>
<td>CHF to death</td>
</tr>
</tbody>
</table>

*N- never smoker; E- ever smoker; A- always smoker
++ I- optimal, II- high normal or elevated, III- high
Risk factors and compression of cardiovascular morbidity

Abstract

The objective of this chapter is to investigate whether being at optimal risk for the major cardiovascular risk factors at middle age (age 30-50) compresses cardiovascular morbidity. We used the first 48 years of follow-up of the Framingham Heart Study original cohort. We considered several risk factors separately (univariate analysis) and in combination (multivariate analysis). We distinguished two risk status levels: optimal (also referred to as favorable) and high-risk. In the univariate analysis, we constructed multistate life tables for optimal and high-risk categories of each single risk factor. In the case of the multivariate analysis, the optimal risk profile was defined as a never smoker, with, on average, an optimal blood pressure (SBP<120 and DBP<80), optimal cholesterol level (SCL<200) and optimal BMI (BMI<25) between age 30 and 50. The high-risk profile was defined as a smoker, with high blood pressure (SBP>140 or DBP>90), high serum cholesterol (SCL>240) and obesity (BMI>30). The effects of an optimal (high) risk factor profile combined the effects of optimal (high) single risk factors. These were not based on observations of individuals with such combinations, as they did not exist in the data set. Age-specific transition rates were estimated using a Gompertz regression model both for the univariate and multivariate analysis. The univariate analysis indicated that the difference in life expectancy free of CVD for males is the highest between optimal and high SBP (7.29 years). For females, this difference was the highest between optimal blood pressure and hypertension (7.84 years). While the differences in life expectancy with CVD between low cholesterol and high cholesterol levels are the highest for males (3.71 years more for high cholesterol), it is the highest between optimal BP and hypertension categories for females (3.00 years more for hypertension). Univariate analysis indicated that the major cardiovascular risk factors (except smoking) at middle age compressed cardiovascular morbidity. Multivariate analysis shows that a fifty year-old male with an optimal risk profile can expect to survive 6 additional
years compared to the total male population in the FHS. Similarly, fifty year-old females in the optimal risk group, can expect to survive 5 additional years compared to the total females in the FHS. At age 50, a male with an optimal risk profile can expect to survive 17 years more free of cardiovascular disease compared to a subject with a high-risk profile; a similarly aged female can expect to survive an extra 16 years compared to females with a high-risk profile. A high-risk profile at middle age shortens the duration of life, increases the lifetime probability of experiencing CVD and extends the period of life spent with cardiovascular disease. As the optimal risk profile increases the number of years lived free of cardiovascular disease and decreases the years lived with cardiovascular disease, we can conclude that an optimal risk profile for major cardiovascular disease risk factors in middle age compresses cardiovascular morbidity.

9.1 Introduction

The impact of risk factors on the cardiovascular disease incidence and mortality has been estimated in the past (Stamler et al., 1999; Anderson et al., 1991). However, what is not known is whether optimal risk status for the most common cardiovascular risk factors compresses cardiovascular morbidity. The primary objective of this chapter is to investigate whether being at optimal risk (both in single risk factor and combined risk factor status) for the major cardiovascular risk factors (both single risk factor status and multiple risk factor profiles) at middle age compresses cardiovascular morbidity. Henceforth, the term risk factor “profiles” will refer solely to a status defined by combination of risk factors. Here, we estimate the population burden of cardiovascular risk factors, expressed as life years lost to disease, life with disease, and life years lost through death i.e. life expectancy.

Experience has shown that a multifactorial approach, one that takes into consideration most of the risk factors, is probably the best strategy for the prevention of cardiovascular disease (Anderson et al., 1991; Lowe et al., 1998) and premature death. The multifactorial risk factor impact on the incidence of cardiovascular disease and mortality has been measured in earlier studies (Stamler et al., 1999; Anderson et al., 1991). Using the Framingham Heart Study, several prediction equations (both parametric and non-parametric) for CVD have been developed to date to predict CVD and its different manifestations. Anderson et al. (1991) used an accelerated failure time model for the prediction of CVD (including its different manifestation) within a short period of time. Wilson (1998) developed a simple coronary heart disease prediction algorithm using risk factors as categorical variables, which allows physicians to predict coronary heart disease (CHD) risk in patients without overt CHD. All these estimates refer to the risk of experiencing CVD within a short period of time (usually less than 12 years) under certain
specified risk factor profiles. They say nothing about the time spent in a specific disease state associated with those risk factors. For instance, life expectancy with cardiovascular disease and without cardiovascular disease by risk profile at middle age is unknown. That is, although the impact of cardiovascular disease risk factors on the risk of cardiovascular disease has been measured, the impact of a single risk factor and of multifactoral risk factor profiles on the life expectancy with and without cardiovascular disease has not been investigated.

A recent review study by Ben-Shlomo and Diana Kuh (2002) concluded that life course epidemiology has challenged the contentment with the adult lifestyle model of chronic disease risk. Risk factors in early life like tobacco use and obesity, for example, not only affect an individual’s own later health but also the health of the next generation (WHO, 2002). Recent studies have concluded that overweight or obesity in adulthood decreases life expectancy by a similar magnitude to smoking (Peeters et al., 2003; Fontaine et al., 2003). The impact of a single risk factor status and risk profiles at middle age (e.g. 30-50 years) on later ages of life is important and preventable compared to the single risk factor status and risk profiles at older ages. The relationship of the risk profiles at older ages with morbidity and mortality are also more complex (Kaplan et al., 1999). People with favorable risk profiles of CVD at middle age survive several years longer compared to those with unfavorable risk factor profiles (Lowe et al., 1998; Daviglus et al., 1998; Stamler et al., 1999). Low risk at middle age is also associated with lower average annual costs for medical care at older age (Daviglus et al., 1998). It is unknown whether such risk factors also decrease the lifetime probability of CVD and the duration of life with and without CVD. Therefore, in this chapter, we have focused on the important CVD risk factors during middle age and their impact on CVD and mortality at later ages.

Although there are many established risk factors for CVD, our focus will be on the following major CVD risk factors: smoking, systolic blood pressure (SBP), diastolic blood pressure (DBP), blood pressure (BP), serum cholesterol level (SCL) and body mass index (BMI). These are the standard cardiovascular risk factors (Pooling Project Research Group, 1978; Manson et al., 1990; Berenson et al., 1998; Stamler et al., 1998; Lowe et al., 1998). Improvements in these risk factors are likely to have a beneficial impact on all-cause mortality and incidences of cardiovascular disease (Norrish et al., 1995).

To measure the compression of cardiovascular morbidity for each risk factor, we constructed multistate life tables separately for each risk factor category (i.e. univariate analysis) by sex. To obtain precise estimates of the net or partial effect of a specific covariate on CVD, we needed to control for other covariates. Therefore, to estimate the combined effect of several risk factors, we constructed multistate life tables in multiple-covariates context (i.e. multivariate analysis) by bridging the multivariate regression model and the multistate life table. We estimated the age-
specific transition rates by means of both univariate and multivariate analysis using the Gompertz (1825; 1827) regression models described in Chapter 7 of this study. We used the estimated parameters to construct the life table. We took the opportunity afforded by the FHS data of using risk factor status between the ages of 30 and 50, and the first 48 years of follow-up of cardiovascular disease history and mortality in the Framingham Heart Study original cohort. Our analysis of the effects of risk factor profiles combines the effects of single risk factors. For instance, the analysis of the effects of an optimal risk factor profile combines the effects of optimal single risk factors. It is not based on observations of individuals with such combinations, as these did not exist in the data set.

We estimated the lifetime risk of CVD (any cardiovascular disease) and its subtypes (any coronary heart disease and any myocardial infarction) given the specific CVD risk factor status and risk profiles at middle age. These were thought to yield an indication of the potential to control the risk factors in adulthood and the consequences at later ages. We calculated the expected number of years lived free of CVD and with CVD given the risk factor status at middle age, in order to arrive at an indication of cardiovascular risk free life expectancy, i.e. such as health expectancy related to the cardiovascular disease risk factor at middle ages of life.

Section 9.2 describes the data and methods: the data source and risk factor exposures, state space and model selection procedure, the estimation of transaction rates using Gompertz model, and construction of life tables. Results are illustrated in Section 9.3. In subsection 9.3.1, we have described the results obtained from the univariate analysis. Results from the multistate life tables in multiple covariate contexts are described in subsection 9.3.2. We have concluded the chapter with a discussion in Section 9.4.

9.2 Data and methods

9.2.1 Data and risk factor exposure

We used the Framingham Heart Study, which began in 1948 and has followed 5209 participants (28 to 62 years of age at entry to the study) as part of a prospective epidemiologic study of cardiovascular disease. Enrollment criteria and study design have been previously published (Dawber et al., 1951). For the purpose of this chapter, we used the data regarding age at onset of cardiovascular disease or death over forty-eight years of follow-up (exam rounds 1 to 24) for the 3045 participants free of cardiovascular disease before age 50 and the risk factor status recorded at minimally two exams undergone by the subject between the ages of 30 and 50. Eligibility criteria for inclusion in our study are described in Chapter 8. Among many established cardiovascular risk factors, we considered four major risk factors: smoking status, systolic blood pressure (SBP), diastolic blood pressure (DBP), serum cholesterol
level (SCL) and body mass index (BMI). Methods for assessing risk factors have been published previously (Dawber et al., 1951; Cupples et al., 1988). All these risk factors cause premature death, increased cardiovascular disease, pulmonary disease, cancer and other diseases (Yusuf et al., 2001). The sample population and risk factor exposure used in this chapter are the same as the population and risk exposure defined in Chapter 8.

In the multivariate analysis, an optimal risk profile was defined as a non-smoker with, on average, optimal blood pressure (SBP<120 and DBP<80), optimal serum cholesterol level (SCL<200) and optimal BMI (BMI<25) throughout the period from age 30 to 50. If an individual was a smoker, had high blood pressure (SBP>140 and DBP>90), high serum cholesterol (SCL>240) and was obese (BMI>30), he or she was considered to have a high-risk profile.

9.2.2 State space and model

State space
The basic multistate life table has the state space \{NO-CVD, history of CVD, dead\} (model 3.1(b), Section 3.3.2, Chapter 3) where CVD is represented by one of the specific CVD states: all cardiovascular disease, all coronary heart disease and acute myocardial infarction. In the life table for all CVD, the possible transitions are, “NO-CVD” to “death”, “NO-CVD” to “history of CVD”, and “history of CVD” to “death”. In the life table for all CHD, the possible transitions are, “NO-CHD” to “death”, “NO-CHD” to “history of CHD”, and “history of CHD” to “death”. Similarly, in the life table for MI, the possible transitions are, “NO-MI” to “death”, “NO-MI” to “history of MI”, and “history of MI” to “death”.

Risk factor status throughout the age interval from 30 to 50 years, disease incidence and mortality throughout the 48 years of follow-up of the Framingham original cohort, from different states, were the same as in Table 8.2 of Chapter 8.

The model selection
We used the Gompertz regression model to specify the shape of the age dependence of the occurrences of an event. We described this model in Chapter 7 of this study. The model was based on risk factor levels and age until events or censored. Several models are distinguished for each transition. We considered the following three levels:

a. The null-model, which is the transition rate model without any covariate.

b. Model with covariates: this model is broken down into two types:
   - Covariates with the $b$ parameters: Fit the model incorporating covariates with the $b$ parameters and compare the change in likelihood ratio. If the change in likelihood ratio is significant, select the model.
• Covariates with \( b \) and \( c \) parameters: include covariates with \( b \) and \( c \) parameters and compare the changes in likelihood ratio with the model where covariates are included with \( b \) parameters. If the reduction is significant consider this third model, otherwise select the second model.

From the overall analysis, we found that incorporation of the time constant covariates with \( b \) parameters was significant compared to the null model. In most of the cases, we found that with the inclusion of the covariates with \( c \) parameters, reduction of the likelihood ratio was not significant. We therefore selected only the Gompertz model with covariates with the \( b \) parameters (Blossfeld and Rohwer, 2002). We assumed that the \( c \) parameter was not influenced by the covariates.

9.2.3 Estimation of transition rates

The basic parameter for constructing the multistate life table was the transition rate. In both univariate and multivariate analysis, the transition rates were estimated applying Gompertz regression models.

Univariate analysis
In the univariate analysis, we performed Gompertz regression analysis incorporating only one risk factor as an independent variable in the model for each transition defined in the previous section. We estimated the age-specific transition rate for each transition separately for males and females. Using these age-specific transition rates we constructed separate multistate life tables for the optimal and high-risk category of that risk factor, in order to establish whether each risk factor could independently compresses cardiovascular morbidity.

Multivariate analysis
Recent applications of the multistate life table technique with a few covariates have included applications in the area of work and retirement (Hayward and Grady, 1990), active life expectancy (Land et al., 1994) and life cycle model of labor force inequality (Hayward and Lichter, 1998). The authors used a piecewise-constant transition rates model for the panel data where transitions occurred at the middle of the interval or at a discrete time point, in the absence of exact timing of event or censoring. If the exact dates of transitions into and out of the states in continuous event histories are known, the estimation of hazard regression models at individual level with the computation multistate life tables would be directly applicable (Land et al., 1994). Here, we have described the multivariate rate models that we used to estimate the transition rates. We chose the Gompertz regression model, as it has fewer parameters (only two) and yields smoothed transition rates. Other features are described in Chapter 7. We assumed that any disease incidence or mortality rates
(following the basic model of cardiovascular disease, Figure 3.1(b), Chapter 3) follow Gompertz’s law.

We conducted multivariate analysis to construct multistate life tables in multiple covariate contexts. These we called multistate life tables in a multiple covariate contexts since each age-specific transition rate, which was the key component of the multistate life table, was adjusted with the described risk factors. The effect of the covariates on the transition rates can be given an interpretation similar to the conventional hazard rate regression models. For a better understanding, we have provided an example. Say we would like to estimate the age-specific transition rates for the transition- NO-CVD to CVD of male participants in FHS. We estimated the regression coefficients and standard errors, incorporating all mentioned covariates in the Gompertz regression models (Table 9.1). If we wanted to predict the transition rate-NO-CVD to CVD for the high-risk profile male at age 60, we could do so by using the estimated parameters in Table 9.1 as follows:

\[
\mu_{\text{nocvd,cvd}}(x, x+1) = \exp(-5.3958 + 0.3686 + 0.396 + 0.6421 + 0.6756) \exp(0.0504 \times (60 - 50)) \quad (9.1)
\]

\[
= 0.060
\]

We used a TDA (Transitional Data Analysis) program to estimate the age-specific transition rates for the transitions from NO-CVD to CVD or NO-CVD to death. Since TDA starts counting age or time from ‘0’, we needed to subtract 50 from 60 to estimate the transition rates at age 60. We used STATA-7 for the post-disease transitions. TDA cannot control the left-truncation, left entry or left censoring.

In this way, we can estimate the transition rates for any other specific age. We assume that each transition rate depends only on age and risk factor status before age 50, providing a common prediction model across types of events.
Table 9.1 Regression coefficients and standard error of Gompertz regression models for the transition NO-CVD to CVD for males

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Coefficients</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>0.0907</td>
<td>0.1256</td>
</tr>
<tr>
<td>Always smoker</td>
<td>0.3686</td>
<td>0.0943</td>
</tr>
<tr>
<td>Blood pressure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Elevated/high normal</td>
<td>0.2203</td>
<td>0.0983</td>
</tr>
<tr>
<td>High</td>
<td>0.6421</td>
<td>0.1057</td>
</tr>
<tr>
<td>Body mass index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Elevated/high normal</td>
<td>0.1453</td>
<td>0.0796</td>
</tr>
<tr>
<td>High</td>
<td>0.3960</td>
<td>0.1155</td>
</tr>
<tr>
<td>Cholesterol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Elevated/high normal</td>
<td>0.3867</td>
<td>0.1179</td>
</tr>
<tr>
<td>High</td>
<td>0.6756</td>
<td>0.1184</td>
</tr>
<tr>
<td>Constant</td>
<td>(a)</td>
<td>-5.3958</td>
</tr>
<tr>
<td>Constant</td>
<td>(b)</td>
<td>0.0504</td>
</tr>
<tr>
<td>-Log likelihood</td>
<td></td>
<td>-3530.29</td>
</tr>
</tbody>
</table>

\(^*\) reference category

9.2.4 Life table construction

Separate multistate life tables were constructed for each of the cardiovascular disease types described above, for optimal and for high-risk groups. For the purpose of the univariate and multivariate analysis, the transition rates for each single year of age were calculated applying the Gompertz regression models described in the previous section. In the univariate analysis, age-specific transition rates were unadjusted and in the multivariate analysis, age-specific transition rates were the combined effects of all risk factors.

To construct the life table, the transition rates were converted to probabilities by assuming that within each single year age interval, the hazard remained constant, taking into account the competing risk. Once we derived the age-specific transition probability, construction of the life tables was straightforward, as described in Section 3.3.4 of Chapter 3. The life tables constructed for each single risk factor category were the same as the life table constructed for always smokers and never smokers in Chapter 6. The main difference is that in Chapter 6, the empirical occurrence-exposure rates were smoothed using the Gompertz regression, while in
this chapter, we use Gompertz regression as a transition rate model to the individual level data to predict the age-specific transition rates. In chapter 6, the relative risk of disease incidence and mortality by smoking status for each one-year age band was age dependent. The relative risks did not differ by age. In the multivariate analysis, the age-specific rates are estimated for the combined effects of all risk factors. The procedures to construct other life table statistics in both the univariate and multivariate analysis are the same.

All life tables were constructed from age 50 and closed at age 90 using the Massachusetts life expectancy at age 90 for 1989-91 (males 3.93 years, females 4.76 years and total population 4.55 years). We assumed that mortality rates beyond age 90 were similar irrespective of risk factor profiles at middle age.

We constructed a total of 8 life tables (2 for alive to death model; 2 for CVD model; 2 for CHD model and 2 for MI model) for one category of each risk factor, where each disease model was constructed separately for males and females. To compare high and optimal risk for each univariate case, we therefore constructed a total of 96 life tables (16 for smoking status; 16 for SBP; 16 for DBP; 16 for BP; 16 for BMI and 16 for cholesterol). For the multivariate case, we constructed a total of 16 life tables (8 for optimal risk and 8 for high-risk profiles).

9.3 Results

The results are presented into two parts. The first part describes the life table outcomes obtained from the univariate analysis. Here, we investigate whether optimal risk status for each single risk factor compresses cardiovascular morbidity, ignoring potential confounding. We mainly present the life expectancies free of cardiovascular disease and with cardiovascular disease, the lifetime probability of cardiovascular disease and the differences in the number of years spent with disease between optimal and high-risk exposure of each risk factor.

In the second part, we investigated whether optimal risk profiles compress cardiovascular morbidity. We constructed multivariate multistate life tables that translate age-specific transition rates for the combined risk factors into life table estimates: survival probabilities, life expectancies free of cardiovascular disease and with cardiovascular disease, lifetime probability of cardiovascular disease and the differences in number of years spent with disease between optimal and high-risk profiles. We then compared the results between the optimal and high-risk profiles.

9.3.1 Univariate analysis

Life expectancy
Total life expectancy, life expectancy free of CVD and with CVD at age 50 and 70 are presented by the optimal and high-risk category of each risk factor separately
for males and females (Table 9.2). A 50-year male smoker can expect to survive 19.86 additional years free of CVD and 6.54 years (25%) with CVD. At the same age, a non-smoking male can expect 4.16 more years of life free of CVD, as compared to a male smoker. However, life expectancy with CVD for a male never smoker is also longer (0.26 years) compared to a male smoker. Similarly, total life expectancy of a 50-year female never smoker is 34.50 years, of which 28.16 years free of CVD and 6.35 years (18%) with CVD. At the same age, a female never smoker can expect to survive 1.44 years more with CVD compared to a female smoker. For males, similar trends are exhibited for coronary heart disease and myocardial infarction. At age 70, the trends seen in the results are similar to those at age 50. For females, life expectancy with MI at age 50 and 70 is slightly higher among female smokers.

A 50-year old male with optimal SBP can expect to survive 23.87 years free of CVD and 5.95 years (21%) with CVD. A male of that age with high SBP can expect to survive 16.58 years without CVD and 6.60 years (28%) with CVD. A male at age 50 with optimal SBP can expect to survive 6.64 more years compared to a male of that age with high SBP. At the same age, a male with high SBP can expect to survive 0.65 years more with CVD. Life expectancy with CHD at age 50 or 70 is marginally higher for males with optimal SBP compared to males with higher SBP, and it remains nearly same with MI. A fifty-year-old female can expect to survive 7.75 years more free of CVD if she had optimal SBP instead of high SBP before age 50. A female with high SBP can expect to survive 2.71 years more with CVD compared to a female with optimal SBP at the same age. Overall, females with high SBP at mid-life can expect to survive longer with disease at age 50 or 70 compared to the females with optimal SBP: from –0.75 years with MI at age 70 to –2.71 years with CVD at age 50.

The life expectancy with CVD or CHD or MI is higher for males or females with high DBP compared to males or females with optimal DBP. For males, this ranges from –0.13 years with MI at age 50 to –0.59 years with CVD at age 50. The difference is higher for females in whom this varies from –0.50 with MI or CHD at age 70 to 1.95 years with CVD at age 50.

A male with optimal blood pressure at middle age can expect to survive 6.75 years more free of cardiovascular disease at age 50 compared to a hypertensive male of that age. The life expectancy with CVD of a hypertensive male at age 50 is 1.05 years higher compared to a male with optimal blood pressure of that age. Likewise, a female with optimal blood pressure at middle age can expect to survive 7.84 (more than one year higher compared to a male) years more free of cardiovascular disease at age 50 compared to a hypertensive female of that age. The life expectancy with CVD of a hypertensive male at age 50 is 3.00 years (nearly 2 years higher compared to a male) higher compared to a female with optimal blood pressure. Similar trends are observed in the differences between life expectancy free of
Table 9.2  Total life expectancy (LE) and residual life expectancy free from cardiovascular disease based on a population free of cardiovascular disease at age 50, by single risk factor status

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Age</th>
<th>Total</th>
<th>Free of CVD</th>
<th>With CVD</th>
<th>Free of CHD</th>
<th>With CHD</th>
<th>Free of MI</th>
<th>With MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never smoker</td>
<td>50</td>
<td>30.82</td>
<td>24.02</td>
<td>6.80</td>
<td>25.43</td>
<td>5.36</td>
<td>27.68</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>14.80</td>
<td>9.31</td>
<td>5.49</td>
<td>10.71</td>
<td>4.15</td>
<td>12.43</td>
<td>2.25</td>
</tr>
<tr>
<td>Always smoker</td>
<td>50</td>
<td>26.40</td>
<td>19.86</td>
<td>6.54</td>
<td>21.60</td>
<td>4.65</td>
<td>23.13</td>
<td>2.50</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>11.87</td>
<td>6.59</td>
<td>5.28</td>
<td>8.14</td>
<td>3.62</td>
<td>9.40</td>
<td>1.94</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>50</td>
<td>4.42</td>
<td>4.16</td>
<td>0.26</td>
<td>3.83</td>
<td>0.72</td>
<td>4.55</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.93</td>
<td>2.72</td>
<td>0.20</td>
<td>2.57</td>
<td>0.53</td>
<td>3.03</td>
<td>0.31</td>
</tr>
<tr>
<td>SBP&lt;120</td>
<td>50</td>
<td>29.82</td>
<td>23.87</td>
<td>5.95</td>
<td>25.29</td>
<td>4.43</td>
<td>26.93</td>
<td>2.62</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>14.07</td>
<td>9.15</td>
<td>4.92</td>
<td>10.40</td>
<td>3.58</td>
<td>11.76</td>
<td>2.18</td>
</tr>
<tr>
<td>SBP 140+</td>
<td>50</td>
<td>23.19</td>
<td>16.58</td>
<td>6.60</td>
<td>18.88</td>
<td>4.20</td>
<td>20.32</td>
<td>2.64</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>9.74</td>
<td>4.64</td>
<td>5.10</td>
<td>6.56</td>
<td>3.11</td>
<td>7.48</td>
<td>2.06</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>50</td>
<td>6.64</td>
<td>7.29</td>
<td>-0.65</td>
<td>6.41</td>
<td>0.23</td>
<td>6.62</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>4.33</td>
<td>4.51</td>
<td>-0.18</td>
<td>3.84</td>
<td>0.47</td>
<td>4.29</td>
<td>0.13</td>
</tr>
<tr>
<td>DBP&lt;70</td>
<td>50</td>
<td>29.11</td>
<td>22.65</td>
<td>6.46</td>
<td>24.05</td>
<td>4.75</td>
<td>25.65</td>
<td>2.97</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>13.63</td>
<td>8.29</td>
<td>5.34</td>
<td>9.59</td>
<td>3.83</td>
<td>10.85</td>
<td>2.50</td>
</tr>
<tr>
<td>DBP 90+</td>
<td>50</td>
<td>24.27</td>
<td>17.22</td>
<td>7.05</td>
<td>18.88</td>
<td>5.30</td>
<td>21.00</td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>10.46</td>
<td>4.95</td>
<td>5.51</td>
<td>6.40</td>
<td>4.06</td>
<td>7.81</td>
<td>2.50</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>50</td>
<td>4.84</td>
<td>5.43</td>
<td>-0.59</td>
<td>5.18</td>
<td>-0.55</td>
<td>4.65</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>3.17</td>
<td>3.34</td>
<td>-0.17</td>
<td>3.19</td>
<td>-0.23</td>
<td>3.03</td>
<td>0.00</td>
</tr>
<tr>
<td>BP&lt;120</td>
<td>50</td>
<td>30.07</td>
<td>24.16</td>
<td>5.91</td>
<td>25.66</td>
<td>4.27</td>
<td>27.68</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>14.26</td>
<td>9.34</td>
<td>4.91</td>
<td>10.67</td>
<td>3.47</td>
<td>12.43</td>
<td>2.25</td>
</tr>
<tr>
<td>BP 140+</td>
<td>50</td>
<td>24.37</td>
<td>17.41</td>
<td>6.96</td>
<td>19.23</td>
<td>4.97</td>
<td>21.12</td>
<td>2.99</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>10.52</td>
<td>5.05</td>
<td>5.47</td>
<td>6.59</td>
<td>3.84</td>
<td>7.89</td>
<td>2.41</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>50</td>
<td>5.70</td>
<td>6.75</td>
<td>-1.05</td>
<td>6.43</td>
<td>-0.69</td>
<td>6.56</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>3.74</td>
<td>4.30</td>
<td>-0.56</td>
<td>4.09</td>
<td>-0.37</td>
<td>4.54</td>
<td>-0.16</td>
</tr>
<tr>
<td>SCL &lt;200</td>
<td>50</td>
<td>28.09</td>
<td>23.70</td>
<td>4.39</td>
<td>25.02</td>
<td>3.10</td>
<td>25.96</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>13.00</td>
<td>9.40</td>
<td>3.61</td>
<td>10.45</td>
<td>2.52</td>
<td>11.32</td>
<td>1.70</td>
</tr>
<tr>
<td>SCL 240+</td>
<td>50</td>
<td>26.98</td>
<td>18.89</td>
<td>8.09</td>
<td>20.67</td>
<td>6.14</td>
<td>23.07</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>12.38</td>
<td>5.80</td>
<td>6.58</td>
<td>7.45</td>
<td>4.84</td>
<td>9.15</td>
<td>2.94</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>50</td>
<td>2.13</td>
<td>4.82</td>
<td>-3.71</td>
<td>4.35</td>
<td>-3.04</td>
<td>2.90</td>
<td>-1.56</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>1.25</td>
<td>3.59</td>
<td>-2.97</td>
<td>3.00</td>
<td>-2.33</td>
<td>2.17</td>
<td>-1.24</td>
</tr>
<tr>
<td>BMI&lt;25</td>
<td>50</td>
<td>28.46</td>
<td>22.50</td>
<td>5.95</td>
<td>24.00</td>
<td>4.30</td>
<td>25.47</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>13.25</td>
<td>8.37</td>
<td>4.87</td>
<td>9.74</td>
<td>3.41</td>
<td>10.87</td>
<td>2.27</td>
</tr>
<tr>
<td>BMI 30+</td>
<td>50</td>
<td>24.66</td>
<td>17.10</td>
<td>7.57</td>
<td>19.03</td>
<td>5.82</td>
<td>20.73</td>
<td>4.04</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>10.74</td>
<td>4.70</td>
<td>6.04</td>
<td>6.43</td>
<td>4.57</td>
<td>7.49</td>
<td>3.34</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>50</td>
<td>3.79</td>
<td>5.41</td>
<td>-1.61</td>
<td>4.97</td>
<td>-1.52</td>
<td>4.74</td>
<td>-1.25</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.51</td>
<td>3.67</td>
<td>-1.16</td>
<td>3.32</td>
<td>-1.16</td>
<td>3.37</td>
<td>-1.07</td>
</tr>
</tbody>
</table>
### Table 9.2...

#### Female

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Age</th>
<th>Total</th>
<th>Free of CVD</th>
<th>With CVD</th>
<th>Free of CHD</th>
<th>With CHD</th>
<th>Free of MI</th>
<th>With MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never smoker</td>
<td>50</td>
<td>34.50</td>
<td>28.16</td>
<td>6.35</td>
<td>30.17</td>
<td>4.24</td>
<td>32.88</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>17.15</td>
<td>11.79</td>
<td>5.36</td>
<td>13.64</td>
<td>3.50</td>
<td>15.84</td>
<td>1.15</td>
</tr>
<tr>
<td>Always smoker</td>
<td>50</td>
<td>30.07</td>
<td>25.16</td>
<td>4.91</td>
<td>26.96</td>
<td>2.96</td>
<td>28.45</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>14.00</td>
<td>9.94</td>
<td>4.06</td>
<td>11.55</td>
<td>2.35</td>
<td>12.60</td>
<td>1.30</td>
</tr>
<tr>
<td>Difference</td>
<td>50</td>
<td>4.43</td>
<td>3.00</td>
<td>1.44</td>
<td>3.20</td>
<td>1.27</td>
<td>4.43</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>3.15</td>
<td>1.86</td>
<td>1.30</td>
<td>2.09</td>
<td>1.15</td>
<td>3.24</td>
<td>-0.14</td>
</tr>
<tr>
<td>SBP&lt;120</td>
<td>50</td>
<td>34.59</td>
<td>30.17</td>
<td>4.42</td>
<td>31.98</td>
<td>2.62</td>
<td>33.66</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>17.35</td>
<td>13.58</td>
<td>3.78</td>
<td>15.23</td>
<td>2.16</td>
<td>16.63</td>
<td>0.76</td>
</tr>
<tr>
<td>SBP 140+</td>
<td>50</td>
<td>29.55</td>
<td>22.42</td>
<td>7.13</td>
<td>24.83</td>
<td>4.55</td>
<td>27.48</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>13.66</td>
<td>7.75</td>
<td>5.91</td>
<td>9.97</td>
<td>3.68</td>
<td>11.98</td>
<td>1.51</td>
</tr>
<tr>
<td>Difference</td>
<td>50</td>
<td>5.04</td>
<td>7.75</td>
<td>-2.71</td>
<td>7.15</td>
<td>-1.91</td>
<td>6.18</td>
<td>-0.88</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>3.70</td>
<td>5.82</td>
<td>-2.13</td>
<td>5.26</td>
<td>-1.52</td>
<td>4.65</td>
<td>-0.75</td>
</tr>
<tr>
<td>DBP&lt;70</td>
<td>50</td>
<td>33.87</td>
<td>28.87</td>
<td>5.01</td>
<td>32.52</td>
<td>1.13</td>
<td>32.52</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>16.82</td>
<td>12.56</td>
<td>4.26</td>
<td>15.76</td>
<td>1.00</td>
<td>15.76</td>
<td>1.00</td>
</tr>
<tr>
<td>DBP 90+</td>
<td>50</td>
<td>29.54</td>
<td>22.58</td>
<td>6.95</td>
<td>27.86</td>
<td>1.73</td>
<td>27.86</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>13.65</td>
<td>7.85</td>
<td>5.81</td>
<td>12.27</td>
<td>1.49</td>
<td>12.27</td>
<td>1.49</td>
</tr>
<tr>
<td>Difference</td>
<td>50</td>
<td>4.34</td>
<td>6.28</td>
<td>-1.95</td>
<td>4.67</td>
<td>-0.60</td>
<td>4.67</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>3.17</td>
<td>4.72</td>
<td>-1.55</td>
<td>3.49</td>
<td>-0.50</td>
<td>3.49</td>
<td>-0.50</td>
</tr>
<tr>
<td>BP&lt;120</td>
<td>50</td>
<td>34.72</td>
<td>30.37</td>
<td>4.35</td>
<td>32.11</td>
<td>2.60</td>
<td>33.75</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>17.46</td>
<td>13.74</td>
<td>3.73</td>
<td>15.33</td>
<td>2.16</td>
<td>16.71</td>
<td>0.78</td>
</tr>
<tr>
<td>BP 140+</td>
<td>50</td>
<td>29.88</td>
<td>22.53</td>
<td>7.35</td>
<td>24.99</td>
<td>4.77</td>
<td>27.99</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>13.93</td>
<td>7.79</td>
<td>6.15</td>
<td>10.02</td>
<td>3.95</td>
<td>12.39</td>
<td>1.43</td>
</tr>
<tr>
<td>Difference</td>
<td>50</td>
<td>4.85</td>
<td>7.84</td>
<td>-3.00</td>
<td>7.12</td>
<td>-2.17</td>
<td>5.76</td>
<td>-0.76</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>3.53</td>
<td>5.95</td>
<td>-2.42</td>
<td>5.30</td>
<td>-1.79</td>
<td>4.32</td>
<td>-0.65</td>
</tr>
<tr>
<td>SCL &lt;200</td>
<td>50</td>
<td>34.06</td>
<td>28.39</td>
<td>5.67</td>
<td>31.13</td>
<td>2.74</td>
<td>32.46</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>17.06</td>
<td>12.13</td>
<td>4.93</td>
<td>14.57</td>
<td>2.34</td>
<td>15.67</td>
<td>1.13</td>
</tr>
<tr>
<td>SCL 240+</td>
<td>50</td>
<td>31.63</td>
<td>25.38</td>
<td>6.25</td>
<td>27.33</td>
<td>4.22</td>
<td>29.73</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>15.17</td>
<td>9.95</td>
<td>5.23</td>
<td>11.77</td>
<td>3.43</td>
<td>13.60</td>
<td>1.49</td>
</tr>
<tr>
<td>Difference</td>
<td>50</td>
<td>2.25</td>
<td>3.02</td>
<td>-0.58</td>
<td>3.80</td>
<td>-1.47</td>
<td>2.73</td>
<td>-0.44</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>1.65</td>
<td>2.18</td>
<td>-0.30</td>
<td>2.80</td>
<td>-1.09</td>
<td>2.08</td>
<td>-0.36</td>
</tr>
<tr>
<td>BMI&lt;25</td>
<td>50</td>
<td>33.29</td>
<td>28.18</td>
<td>5.11</td>
<td>30.25</td>
<td>2.94</td>
<td>31.93</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>16.39</td>
<td>12.07</td>
<td>4.33</td>
<td>13.95</td>
<td>2.40</td>
<td>15.29</td>
<td>1.05</td>
</tr>
<tr>
<td>BMI 30+</td>
<td>50</td>
<td>29.52</td>
<td>22.61</td>
<td>6.92</td>
<td>24.66</td>
<td>4.83</td>
<td>27.39</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>13.66</td>
<td>7.93</td>
<td>5.73</td>
<td>9.79</td>
<td>3.93</td>
<td>11.81</td>
<td>1.77</td>
</tr>
<tr>
<td>Difference</td>
<td>50</td>
<td>3.77</td>
<td>5.58</td>
<td>-1.81</td>
<td>5.59</td>
<td>-1.89</td>
<td>4.54</td>
<td>-0.85</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.74</td>
<td>4.14</td>
<td>-1.40</td>
<td>4.16</td>
<td>-1.53</td>
<td>3.47</td>
<td>-0.72</td>
</tr>
</tbody>
</table>

At age 50, the total life expectancy differences between optimal BMI and obese males or females are nearly 4 years. Using a similar data set, Peeters et al. (2003) found a life expectancy difference between normal and obese males or females of about 6 years at age 40. Peeters et al., (2003) stratified the population by smoking status, used the first 40 years of follow-up and considered only the baseline BMI.
status. In our analysis, we excluded CVD and death before age 50. All these assumptions may have contributed to our arriving at the different estimate. At age 50, males can expect 5.41 years more free of CVD; this is 5.58 years for females. Life expectancy with CVD is higher (1.61 years for males and 1.81 years for females) in obese males and females compared to males and females with normal BMI at middle age. Similar patterns were found for CHD or MI at age 50 or 70.

The total life expectancy difference between a fifty year-old male with an optimal or with a high cholesterol level is 2.13 years. However, at the same age, a male with low cholesterol can expect to survive 4.82 years more free of CVD and 3.71 years less with CVD. A female of the same age and category can expect to survive 3.02 years more without CVD and 0.58 year less with CVD. Males with high cholesterol levels consistently survive longer with CVD or CHD or MI at age 50 or 70. However, females at that age survive longer with CHD.

**Lifetime risk**

The lifetime probability (before death) and the probability of developing CVD before the age of 70 of the male and female cohorts free of CVD at age 50 are presented in Table 9.3 by risk factor status for the age interval extending from 30 to 50 years of age. Both men and women smokers have a higher chance of developing CVD before age 70, as compared to non-smokers (men 42% vs. 35%; women 25% vs. 23%); non-smokers have a higher or similar chance (1% for men and 4% for women) of developing CVD during their lifetime.

The lifetime probability of a male with optimal SBP developing CVD is 58% and 78% for males with high BP; for CHD, these percentages are 41% vs. 53% for CHD and for MI 27% vs. 34% for MI. Before age 70, the difference in the probability of developing CVD between the high and optimal SBP categories is – 24% percent (31% vs. 55%). This difference is also huge for CHD (-15%) and MI (-9%). Similarly, females with high SBP have a 67% chance of developing CVD before death and a 35% chance of developing this before age 70. For developing CVD, the differences between optimal and high SBP are –19% (16% vs. 35%) before age 70 and –24% (42% vs. 67%) before death. Like SBP, DBP also has a huge impact on the chance of developing CVD or CHD or MI before age 70 or during lifetime. For males, the impact is higher before age 70. For females, the impact is higher in lifetime.

The probability of developing CVD, CHD or MI before age 70 or before death is, as expected, significantly higher for hypertensive males as compared to the males who had optimal blood pressure at middle age. Before age 70, the probability is –22% (30% vs. 52%) for CVD, -16% (22% vs. 38%) for CHD, and -9% (13% vs. 22%) for MI. Over lifetime, the difference is slightly lower than before age 70. Like males, hypertensive females have a significantly higher chance of developing CVD or CHD or MI before age 70 or in lifetime as compared to the females who have
optimal blood pressure at middle age. Before death, this is –24% (41% vs. 64%) for CVD, -18% (23% vs. 41%) for CHD, and -11% (12% vs. 23%) for MI. The difference is significantly higher in lifetime compared to before age 70.

Men and women with normal cholesterol levels before age 50 have a significantly lower risk of developing CVD or CHD or MI before age 70 or before death compared to individuals with high cholesterol levels at that age. For men, this ranges from –11% (12% vs. 22%) for MI before age 70 to –29% (32% vs. 60%) for CHD before death; For females, from –3% (4% vs. 7%) for MI before age 70 to –16% (21% vs. 37%) for CHD before death.

The probability of developing CVD or one of its subtypes either before age 70 or before death is significantly higher among obese males or females compared to males or females who have optimal levels of BMI between the ages of 30 and 50. For males, this ranges from –10% for MI before age 70 to –17% for CVD before age 70. For females, this probability runs from -6% for MI before age 70 to –16% for CHD before death. For females the difference is higher before death, while for males the difference is only higher for CVD before age 70, otherwise it is the same.

Table 9.3 Lifetime probability (%) of subjects without cardiovascular disease developing cardiovascular disease at age 50, by single risk factor

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Before age 70</th>
<th>Before death</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CVD</td>
<td>CHD</td>
<td>MI</td>
<td>CVD</td>
<td>CHD</td>
</tr>
<tr>
<td>Never smoker</td>
<td>35</td>
<td>29</td>
<td>17</td>
<td>67</td>
<td>54</td>
</tr>
<tr>
<td>Always smoker</td>
<td>42</td>
<td>31</td>
<td>19</td>
<td>67</td>
<td>49</td>
</tr>
<tr>
<td>Difference</td>
<td>-7</td>
<td>-2</td>
<td>-1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>SBP&lt;120</td>
<td>31</td>
<td>22</td>
<td>13</td>
<td>58</td>
<td>41</td>
</tr>
<tr>
<td>SBP 140+</td>
<td>55</td>
<td>37</td>
<td>22</td>
<td>76</td>
<td>53</td>
</tr>
<tr>
<td>DBP&lt;70</td>
<td>34</td>
<td>26</td>
<td>15</td>
<td>63</td>
<td>45</td>
</tr>
<tr>
<td>DBP 90+</td>
<td>54</td>
<td>41</td>
<td>23</td>
<td>77</td>
<td>59</td>
</tr>
<tr>
<td>Difference</td>
<td>-19</td>
<td>-15</td>
<td>-8</td>
<td>-14</td>
<td>-14</td>
</tr>
<tr>
<td>BP&lt;120</td>
<td>30</td>
<td>22</td>
<td>13</td>
<td>57</td>
<td>40</td>
</tr>
<tr>
<td>BP 140+</td>
<td>52</td>
<td>38</td>
<td>22</td>
<td>76</td>
<td>55</td>
</tr>
<tr>
<td>Difference</td>
<td>-22</td>
<td>-16</td>
<td>-9</td>
<td>-19</td>
<td>-15</td>
</tr>
<tr>
<td>SCL &lt;200</td>
<td>27</td>
<td>18</td>
<td>12</td>
<td>51</td>
<td>32</td>
</tr>
<tr>
<td>SCL 240+</td>
<td>49</td>
<td>39</td>
<td>22</td>
<td>75</td>
<td>60</td>
</tr>
<tr>
<td>Difference</td>
<td>-21</td>
<td>-21</td>
<td>-11</td>
<td>-25</td>
<td>-29</td>
</tr>
<tr>
<td>BMI&lt;25</td>
<td>35</td>
<td>26</td>
<td>16</td>
<td>62</td>
<td>45</td>
</tr>
<tr>
<td>BMI 30+</td>
<td>52</td>
<td>41</td>
<td>27</td>
<td>74</td>
<td>59</td>
</tr>
<tr>
<td>Difference</td>
<td>-17</td>
<td>-15</td>
<td>-10</td>
<td>-12</td>
<td>-15</td>
</tr>
</tbody>
</table>
Continuation of Table 9.3...

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Before age 70</th>
<th>Before death</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CVD</td>
<td>CHD</td>
</tr>
<tr>
<td>Never smoker</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>Always smoker</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>SBP&lt;120</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>SBP 140+</td>
<td>35</td>
<td>23</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>-19</td>
<td>-13</td>
</tr>
<tr>
<td>DBP&lt;70</td>
<td>19</td>
<td>12</td>
</tr>
<tr>
<td>DBP 90+</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>-14</td>
<td>-10</td>
</tr>
<tr>
<td>BP&lt;120</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>BP 140+</td>
<td>35</td>
<td>22</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>-19</td>
<td>-13</td>
</tr>
<tr>
<td>SCL &lt;200</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>SCL 240+</td>
<td>27</td>
<td>19</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>-8</td>
<td>-9</td>
</tr>
<tr>
<td>BMI&lt;25</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>BMI 30+</td>
<td>34</td>
<td>23</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>-14</td>
<td>-11</td>
</tr>
</tbody>
</table>

**Differences in years spent with disease**

Here, the burden of CVD, CHD or MI is measured in terms of the differences in the life-table person years lived in the age interval \([x, x+1]\) with disease between the optimal group and the high-risk group. We used the simple equation described in Chapter 2 (subsection 2.4.7) to estimate the life years lost to disease or number of years lived with disease of the life table population between optimal and high-risk groups. The number of age-specific person years lived with disease for the high-risk group is subtracted from the number of age-specific person years lived with disease for the optimal risk group (Figure 9.1). This approach enables us to establish the length of time spent with cardiovascular morbidity at the relevant age by risk factor status. For instance, in Figure 9.1 (male: never smoker—always smoker), a male who smoked at middle age, spent 0.1991 years with cardiovascular disease in the age interval 60 to 61. In the same age interval, a male who was non-smoker at middle age spent 0.1078 years with cardiovascular disease. The difference is –0.0913 years, i.e. a smoker spends 0.0913 years more with CVD in the age interval 60 to 61. A difference resulting in negative values indicates that a high-risk individual spends a
longer time with disease, i.e. elimination of that risk from the population at middle age will reduce the life years lost to disease at that age.

Male smokers spend more years with CVD before age 75 as compared to never smokers (Figure 9.1). Male never smokers spend more years with CHD or MI from age 65 onwards. Similarly, women smokers spend more time with CVD before age 65. Female never smokers spend more time with CHD. However, they spend less time with MI until age 83. While males with high SBP spend more time with CVD before age 74, females spend more time with CVD throughout life (at least until age 88). Similar trends are seen in males and females for DBP. Hypertensive males spend more years with CVD or CHD until age 78 (MI until 76). Hypertensive females spend more years with CVD and its subtypes over lifetime. Males with obesity at middle life spend more time with CVD before age 83. They spend more years with CHD or MI until age 75. Obese women spend a longer time with CHD or MI until the endpoint of life. While males with high cholesterol levels spend more time with CVD, CHD or MI throughout life, females spend slightly more time with CVD until age 82, compared to optimal cholesterol levels. However, females spend more years with CHD or MI throughout life.

Figure 9.1 The differences in the number of years lived with cardiovascular disease between optimal risk and high-risk category (difference: optimal–high)

a. Never smoker- always smoker
RISK FACTORS AND COMPRESSION OF CARDIOVASCULAR MORBIDITY

Female

\[ \text{Difference (in years)} \]

\[ \begin{array}{llllllll}
\text{CVD} & -0.12 & -0.09 & -0.06 & -0.03 & 0.00 & 0.03 & 0.06 & 0.09 & 0.12 \\
\text{CHD} & -0.09 & -0.06 & -0.03 & 0.00 & 0.03 & 0.06 & 0.09 & 0.12 & -0.09 \\
\text{MI} & -0.06 & -0.03 & 0.00 & 0.03 & 0.06 & 0.09 & 0.12 & -0.06 & -0.03 \\
\end{array} \]

\[ \begin{array}{llllllll}
\text{Age} & 50 & 55 & 60 & 65 & 70 & 75 & 80 & 85 \\
\end{array} \]

b. Optimal SBP-high SBP

Male

\[ \text{Difference (in years)} \]

\[ \begin{array}{llllllll}
\text{CVD} & -0.12 & -0.09 & -0.06 & -0.03 & 0.00 & 0.03 & 0.06 & 0.09 & 0.12 \\
\text{CHD} & -0.09 & -0.06 & -0.03 & 0.00 & 0.03 & 0.06 & 0.09 & 0.12 & -0.09 \\
\text{MI} & -0.06 & -0.03 & 0.00 & 0.03 & 0.06 & 0.09 & 0.12 & -0.06 & -0.03 \\
\end{array} \]

\[ \begin{array}{llllllll}
\text{Age} & 50 & 55 & 60 & 65 & 70 & 75 & 80 & 85 \\
\end{array} \]
Female

\[ \text{Age} \]

\[ \text{Difference (in years)} \]

\begin{array}{|c|c|c|c|}
\hline
& \text{CVD} & \text{CHD} & \text{MI} \\
\hline
50 & -0.14 & -0.07 & 0.00 \\
55 & -0.14 & -0.07 & 0.00 \\
60 & -0.14 & -0.07 & 0.00 \\
65 & -0.14 & -0.07 & 0.00 \\
70 & -0.14 & -0.07 & 0.00 \\
75 & -0.14 & -0.07 & 0.00 \\
80 & -0.14 & -0.07 & 0.00 \\
85 & -0.14 & -0.07 & 0.00 \\
\hline
\end{array}

c. Optimal DBP-high DBP

Male

\[ \text{Age} \]

\[ \text{Difference (in years)} \]

\begin{array}{|c|c|c|c|}
\hline
& \text{CVD} & \text{CHD} & \text{MI} \\
\hline
50 & -0.12 & -0.08 & -0.04 \\
55 & -0.12 & -0.08 & -0.04 \\
60 & -0.12 & -0.08 & -0.04 \\
65 & -0.12 & -0.08 & -0.04 \\
70 & -0.12 & -0.08 & -0.04 \\
75 & -0.12 & -0.08 & -0.04 \\
80 & -0.12 & -0.08 & -0.04 \\
85 & -0.12 & -0.08 & -0.04 \\
\hline
\end{array}
Female

d. Optimal BP-high BP

Male
e. Optimal cholesterol-high cholesterol

---

Female

---

Male

---
Female

Difference (in years)

CVD
CHD
MI

Age

50 55 60 65 70 75 80 85

f. Optimal BMI-obesity

Male

Difference (in years)

CVD
CHD
MI

Age

50 55 60 65 70 75 80 85
9.3.2 Multivariate analysis

In the univariate analysis, we measured whether each cardiovascular disease risk factor compressed cardiovascular morbidity. In the multivariate analysis, we aim to investigate whether optimal risk profiles compress cardiovascular morbidity. We constructed multistate life tables in a multiple covariate context that translated age-specific transition rates for the combined effects of several risk factors into life table estimates: survival probabilities, life expectancies, lifetime risk and the differences in number of years spent with disease between optimal and high-risk profiles.

Survival curves
The age profiles of survival from death and survival free of cardiovascular disease for the lowest and highest risk groups in our life table population are shown in Figure 9.2. This figure shows the survival free of cardiovascular disease, coronary heart disease, acute myocardial infarction, or death of fifty-year-old men and women with optimal and higher risk profiles. As expected, men and women with an optimal risk profile lead a longer life and survive longer free of any cardiovascular disease.
Both mortality and cardiovascular disease incidence led to large differences throughout life between optimal and high-risk profiles. Among fifty year-old males, 47% of those in the high-risk and 14% of those in the low risk groups die before the age of 70. In females, these figures are 36% for the high-risk, but only 7% for optimal risk profiles. While an average of 4 in 5 males with an optimal risk profile who are free of cardiovascular disease at age 50 will be alive and free of cardiovascular disease twenty years later, only 1 in 5 males with a high-risk profile will remain in this state (Figure 1). Among the females free of cardiovascular disease at age 50, 1 in 5 with an optimal risk profile compared to 2 in 5 in the high-risk group will be alive and free of cardiovascular disease at age 70.

Figure 9.2 Survival curves illustrating the probability of surviving and surviving free of cardiovascular disease (CVD), coronary heart disease (CHD), and myocardial infarction (MI), by sex and risk profile
Life expectancy
While survival probabilities are important indicators of the potential burden of a disease by risk factor status, they give no indication of the impact of that disease on life expectancy. To synthesize the consequences of age-specific disease incidence and mortality, the multivariate multistate life table estimates the life expectancies in specific disease states by risk profile (Table 9.4).

At age 50, the life expectancy of a male is 27.25 years, for a female, 32.41 years, in the selected sample population from the Framingham heart study. At age 50, a man with an optimal risk profile can expect to live another 33.04 years, compared to 20.74 years for a man who is in the high-risk group. A female of that age, with an optimal risk profile, can expect to live 37.57 additional years, while a female in the high-risk group can look forward to 23.30 years. The difference in number of additional survival years between an optimal risk and a high-risk profile is for males 12.29 years, and for females, 14.27 years. Even at age 70, this difference is 8.29 years for males and 10.20 years for females. The males in the optimal risk profiles can expect to survive 6 additional years compared to the total male population in the FHS. Similarly, females in the low risk group can expect to survive 5 additional years compared to the total female respondents of the FHS. This huge difference in life expectancy between optimal and high-risk profiles offers an the indication that risk factor status at an earlier stage of life has a significant impact at older ages.

The most impressive is the number of years lived free of disease for the optimal risk profiles. A 50-year-old man with an optimal risk profile can expect to live 17.19 years more free of cardiovascular disease than a member of the high-risk profiles.
A woman of that age with optimal risk can expect to survive 16.22 additional years free of cardiovascular disease. Importantly, at age 70, this difference is 11.37 years for males and 11.36 years for females. Both for males and females, similar patterns of life expectancy free of CHD or MI are also observed at age 50 and 70 (Table 9.4).

Concordant with the often-higher incidence of cardiovascular disease, the duration of disease is longer among members of the high-risk group than among those in the optimal risk group (Table 9.4). At age 50, males in the high-risk group can expect to live 38 percent (7.90 years) of remaining life in the CVD state; for the low risk group, this is 9 percent (3.0 years). Similarly, fifty year-old females with an optimal risk profile spend 11 percent (4.12 years) of the rest of their life with CVD; for the high-risk group, this is 26 percent (6.07 years). Seventy year-old males and females with optimal risk profiles can expect to spend nearly 83 percent and 81 percent of their remaining life free of cardiovascular disease. By contrast, only 27 percent of males and 48 percent of females in the high-risk group can at that same age expect to remain free of CVD. Similar patterns are exhibited for coronary heart disease and myocardial infarctions.

Table 9.4 Total life expectancy and life expectancy free from diseases based on a population free of cardiovascular disease at age 50 by risk profiles at middle age

<table>
<thead>
<tr>
<th>Male</th>
<th>Age 50</th>
<th>Age 60</th>
<th>Age 70</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total LE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>33.03</td>
<td>24.21</td>
<td>16.29</td>
</tr>
<tr>
<td>High risk</td>
<td>20.74</td>
<td>13.56</td>
<td>8.00</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>12.29</td>
<td>10.64</td>
<td>8.29</td>
</tr>
<tr>
<td><strong>Without CVD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>30.03</td>
<td>21.06</td>
<td>13.56</td>
</tr>
<tr>
<td>High risk</td>
<td>12.84</td>
<td>5.77</td>
<td>2.19</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>17.19</td>
<td>15.29</td>
<td>11.37</td>
</tr>
<tr>
<td><strong>Without CHD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>30.71</td>
<td>21.77</td>
<td>14.19</td>
</tr>
<tr>
<td>High risk</td>
<td>14.69</td>
<td>7.59</td>
<td>3.61</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>16.02</td>
<td>14.18</td>
<td>10.57</td>
</tr>
<tr>
<td><strong>Without MI</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>31.96</td>
<td>23.05</td>
<td>15.35</td>
</tr>
<tr>
<td>High risk</td>
<td>16.64</td>
<td>9.37</td>
<td>4.67</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>15.32</td>
<td>13.68</td>
<td>10.69</td>
</tr>
</tbody>
</table>
**Table 9.4**

Continuation of Table 9.4...

<table>
<thead>
<tr>
<th></th>
<th>Total LE</th>
<th>Without CVD</th>
<th>Without CHD</th>
<th>Without MI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age 50</td>
<td>Age 60</td>
<td>Age 70</td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>37.57</td>
<td>28.23</td>
<td>19.50</td>
<td></td>
</tr>
<tr>
<td>High risk</td>
<td>23.30</td>
<td>15.53</td>
<td>9.30</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>14.27</td>
<td>12.70</td>
<td>10.20</td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>33.45</td>
<td>24.14</td>
<td>15.84</td>
<td></td>
</tr>
<tr>
<td>High risk</td>
<td>17.23</td>
<td>9.41</td>
<td>4.48</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>16.22</td>
<td>14.73</td>
<td>11.36</td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>35.59</td>
<td>26.26</td>
<td>17.77</td>
<td></td>
</tr>
<tr>
<td>High risk</td>
<td>19.04</td>
<td>11.27</td>
<td>6.06</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>16.56</td>
<td>14.99</td>
<td>11.71</td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>36.89</td>
<td>27.56</td>
<td>18.92</td>
<td></td>
</tr>
<tr>
<td>High risk</td>
<td>20.86</td>
<td>12.93</td>
<td>6.99</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>16.03</td>
<td>14.64</td>
<td>11.93</td>
<td></td>
</tr>
</tbody>
</table>

**Lifetime probability of disease**

The lifetime probability of a fifty year-old without cardiovascular disease, developing cardiovascular disease is presented by risk factor status in Table 9.5. This probability is a life table probability, as it applies to a synthetic cohort of people aged 50 and free of CVD at that age.

The lifetime probability of developing cardiovascular disease, including sudden cardiovascular death is 82 percent for males with high-risk profile vs. 38 percent for males with optimal risk profiles. Females with optimal risk profiles have a lifetime chance of experiencing CVD of 35 percent; for high-risk women this is 69 percent. It is estimated that within 10 years (i.e. from age 50 to 60) 37 percent of the high-risk males and 24 percent of females will experience CVD. Within 20 years, 67 percent of the males and 45 percent of the females with high-risk profiles will experience CVD. Fifty year-old males and females with optimal risk profiles have a chance of respectively 14 and 11 percent of experiencing CVD before age 70. For males and females, similar patterns are also observed in relation to coronary heart disease and myocardial infarction.
Table 9.5  Lifetime probability (%) of developing cardiovascular disease, for the cardiovascular disease free person of age 50 by optimal and high-risk profiles

<table>
<thead>
<tr>
<th></th>
<th>Before age 60</th>
<th>Before age 70</th>
<th>Before death</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Optimal risk</td>
<td>High risk</td>
<td>Optimal risk</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVD</td>
<td>6</td>
<td>37</td>
<td>14</td>
</tr>
<tr>
<td>CHD</td>
<td>5</td>
<td>34</td>
<td>11</td>
</tr>
<tr>
<td>MI</td>
<td>3</td>
<td>17</td>
<td>8</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVD</td>
<td>5</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>CHD</td>
<td>2</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>MI</td>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

**Differences in years spent with disease**

The difference between optimal and high-risk profiles (optimal-high-risk profiles) in the number of years spent with disease is presented in Figure 9.3. The enormous impact of cardiovascular disease on the human life course is translated into life years lost to disease. Importantly, the life years lost by those in the higher risk profiles to cardiovascular disease fall predominantly before age 80. The optimal risk group survives longer with cardiovascular disease after age 80 (Figure 9.3). Only at older ages (after age 80) do both males and females in the optimal risk group spend more years with cardiovascular disease. After age 80, males and females with an optimal risk profile spend more years with CVD than the high-risk group. Only the subjects in the low risk group survive longer, which indicates that, as people age, the tendency to experience CVD increases. That is, aging itself causes cardiovascular disease.
Figure 9.3 Differences of the number of years lived with diseases: optimal risk profiles—high-risk profiles

Male

Female
9.4 Discussion

The aim of this chapter was to investigate whether optimal levels of cardiovascular risk factors (single and multiple risk factors) at middle age compressed cardiovascular morbidity. To measure this, we constructed multistate life tables of cardiovascular disease history by risk factor status for the 30-50 age interval. In the following sub-sections, the major findings are discussed by comparing our methods and findings with existing studies, in addition to examining the strengths and limitation of our proposed method. In the last section, we briefly discuss the public health implications of this study.

Principal findings

Across all single risk factor categories, males who did not smoke and females who had optimal blood pressure at middle age were shown to have the highest life expectancy at age 50, namely 30.82 years for males and 34.72 years for females. The difference in life expectancy free of CVD for males was highest between optimal SBP and high SBP (7.29 years). For females, this difference was highest between optimal blood pressure and hypertension (7.84 years). While the differences in life expectancy with CVD were highest for males with low and males with high cholesterol levels (3.71 years more for high cholesterol), for females this was highest between the optimal BP and hypertension categories (3.00 years more for hypertension). The probability of developing CVD, CHD, or MI before age 70 was higher for the higher risk categories than for the optimal risk groups. The lifetime probability of a male smoker developing cardiovascular disease was equal to a male non-smoker. The lifetime probability of developing CVD or CHD was higher for female non-smokers, although the reverse holds for MI. These results are consistent with Chapter 6. For males and females, the major cardiovascular risk factors at middle age compress cardiovascular morbidity.

We demonstrated the possibility of using multivariate Gompertz regression models for long time continuous event histories to estimate the state transition rates, after which these estimated transition rates could be used to construct a multistate life table in multiple covariate contexts. Using this multistate life table, we found that, on average, 4 in 5 males with a low risk profile who were free of cardiovascular disease at age 50 would still be alive and free of cardiovascular disease twenty years later. By contrast, only 1 in 5 of high-risk profile males would still be in that state at age 70. Of the females free of cardiovascular disease at age 50, 1 in 5 in the optimal risk profile, as compared to 2 in 5 of low risk profile would be alive and remain free of cardiovascular disease.

A female of age 50 with an optimal risk profile could expect to survive another 37.57 years, as compared to another 23.30 years for a female in the high-risk group. These findings are consistent with Peeters et al., (2003) who estimated these figures
for non-smoking, non-obese females (13 years at age 40) using the Framingham Heart Study original cohort. The differences in total life expectancies at age 50 for males and females with optimal risk and high-risk profiles were 12.29 and 14.27 years. The males in the low risk groups could expect to survive 6.0 additional years compared to the total male respondents in the FHS. Similarly, females with optimal risk profiles could expect to survive 5 additional years compared to the total female respondents in the FHS. Overall, females with high-risk profiles were more vulnerable compared to high-risk profile males, which is consistent with a previous study (Jousilahti et al., 1999).

The most impressive findings concerned the number of years lived free of cardiovascular disease for subjects with optimal risk profiles. A man who had optimal risk profile at middle age, and was free of disease before age 50, could expect to survive 17.19 more years free of cardiovascular disease than a member of the high-risk group. A woman of that age with an optimal risk profile could expect to survive 16.22 years more free of cardiovascular disease. A similar pattern in life expectancy free of CHD or MI for males and females was also observed at age 50 and 70.

The duration of disease was generally longer among members of the high-risk profile than among those in the optimal risk profile. At age 50, a male in the high-risk profile could expect to live 7.90 years of remaining life with CVD; a male with an optimal risk profile, only 3.0 years. Similarly, a female of that age with a high-risk profile could expect to survive 6.07 years with CVD, a female with an optimal risk profile, 4.12 years. This indicates that high risk at adulthood not only shortens the duration of life but also extends life with cardiovascular disease. The optimal risk profile in middle age compresses the cardiovascular morbidity.

For individuals with favorable levels of cholesterol, blood pressure, BMI and for those who do not smoke, the long term incidence of cardiovascular disease and mortality is much lower and longevity is much greater, which is consistent with the previous study by Lowe et al (1998) and Stamler et al (1999). Stamler et al.(1999) estimated greater life expectancy (6 years) of low risk sub-cohorts vs. others based on multiple risk factor intervention trial for a cohort aged 35-39 years or 40-49 years from the Chicago Heart Association detection project population. This was underestimated, because of the greater risk of misclassifying individuals by using a single measurement of the major risk factors. These findings also directly confirm earlier statistical estimates of the benefits of low-risk status (Daviglus, et al., 1998).

Another important finding of this study was the lifetime probability of CVD-free fifty year-olds developing cardiovascular disease. The lifetime probability of developing cardiovascular disease, including sudden cardiovascular death, was, for the synthetic CVD-free cohort at age 50, 82 percent for males, and 69 percent for females in the high-risk profile. For the optimal risk profile, this was 38 percent for males and 35 percent for females. Remarkably, we also found that within 20 years,
67 percent of the males and 45 percent of the females in the high-risk group would experience CVD. It further became evident that the number of life years lost to cardiovascular disease for the higher risk profile fall dramatically before age 80, with the optimal risk group living more years with cardiovascular disease after age 80. After age 80, males and females with an optimal risk profile spent more years with CVD than did those with a high-risk profile. The low risk group survived longer, indicating that as the people age, the tendency to experience CVD also increases. That is, aging itself causes cardiovascular disease.

**Strengths and limitations**

The main strength of this chapter is that it is based on the prospective surveillance of a community-based cohort over a period of 48 years. Over this period, risk factors were measured and accurate data on the incidence and mortality of CVD and its sub-types were gathered consistently, routinely and accurately. Without such an extended follow-up period, it is not possible to empirically analyze the burden of cardiovascular disease risk factors throughout life.

To estimate the transition rates, we chose the Gompertz model since it could capture the biological effect of morbidity and mortality well and give us smoothed transition rates. However, the Gompertz law is not applicable to the oldest-old (usually age 85+) mortality (Olshansky and Carnes, 2001; Yue, 2002). Olshansky and Carnes (2001) suggested that this might be the result of population heterogeneity at advanced ages. We did not extend our life table that was based on the Gompertz model beyond age 90, since Gompertz may overestimate the rates after that age. The life tables were closed at age 90 using the Massachusetts life expectancy for 1989-91 population, assuming that life expectancy remained same after age 90 irrespective of disease incidence and risk factor status. This assumption would not influence our major findings. Life expectancy of the optimal risk group would to some extent be under estimated and overestimated for the higher risk group, but other measures would remain the same, e.g. life time probability, number of years lived with disease and so on. However, by closing the life table at age 80, we dealt with this assumption of the Gompertz model. The observed life expectancy of Massachusetts males, females and total population at age 80 was 7.17, 9.25 and 8.56 years, respectively (Centers for Disease Control, 1989-91). Closing the life table at age 80 based on this observed life expectancy of males, we found that the total life expectancy for the low risk profiles decreased about one-and-a-half years and increased about one-and-a-half years for the high-risk profiles. This slight difference was largely due to our assumption that both the optimal and the high-risk group had the same observed life expectancy at age 80. Similar patterns were also exhibited for females.
We derived total life expectancy and life expectancy with cardiovascular disease by risk factor profiles based on the multivariate regression models. To validate the results, an empirical investigation was made, in which the optimal and high-risk profiles were defined slightly differently from the present approach. We selected the sample population from the same population analyzed in this chapter. We assumed, in those who had never smoked or who smoked less that 20 percent of the recorded time, a mean blood pressure of less than 140 (mm Hg), mean cholesterol level of less than 240 (mg per deciliter) and mean BMI \((w/h^2)\) under 30 before age 50. We found 471 (100 males and 371 females) participants who complied with these conditions. A basic multistate life table (Figure 3.1(b), Chapter 3) was constructed using this sample population. It was found that total life expectancy (male-female combined) at age 50 would be is 35.3 years, while life expectancy with cardiovascular disease would be 4 years. These estimates are slightly lower for total life expectancy (1.5 years) but higher (around 0.5 years) for life expectancy with cardiovascular disease compared to the estimated values obtained in this chapter. This difference is mainly due to the different definition of optimal risk profiles and population selection. The sample population (15% of the 3045 sample population) consisted of a mixture of optimal and high normal risk profiles, indicating that the total life expectancy of this selected sample would be less, and longer with cardiovascular disease as compared to the sample population of 3045.

Another major strength of this study is the multistate life tables in multiple covariate contexts. The transition rate transfers into dwelling times give more transparent information about the consequences of risk factors. This study adds the consequences of risk factor status for cardiovascular disease in terms of incidence and years free of CVD and years with CVD throughout the life course.

Instead of only the baseline risk factor (Stamler et al., 1999; Lowe et al., 1998) status, we also considered at least two risk factor measurements, recorded between the ages of 30 and 50. We defined the risk factor statuses before age 50 and measured their impact on later ages of life. We were able to do so only because of the long period of follow-up of the FHS. Usually, single measures have a large variance, and relating a single (as opposed to multiple) measure of a risk factor to an outcome leads to substantial underestimation of the strength of association (Yusuf et al., 2001). Since the effect of several risk factors for vascular disease may take several years to fully manifest (Yusuf et al., 2001), cohort studies of relatively short duration may therefore only identify a part of this effect, while the extended period covered by the FHS might uncover a larger effect.

The relationships of major cardiovascular disease risk factors (e.g. smoking, blood pressure, cholesterol level) have been assumed to be “… strong, continuous, graded, consistent, independent, predictive, and etiologically significant for those with and without coronary heart disease” (Stamler et al., 1999). We defined the risk factors as categorical variables, since our intention was to construct multistate life
tables by risk factor profiles. We constructed a life table for a specific group (single or combined). Wilson et al., (1998) used the Framingham Heart Study to predict coronary heart disease using risk factors as categorical variables.

We addressed only the major cardiovascular risk factors between age 30 and 50 and not the risks of other diseases of the elderly. Therefore, we can only make conclusions regarding cardiovascular disease, not disability. The optimal risk profile in the MSLT in multiple covariate contexts is the (unobserved) combination of optimal levels of four risk factors, ignoring possible interactions between them. However, we found no evidence in Chapter 8 of non-independence.

The most important outstanding question is to what extent the observed association between the major risk factors and the indicators derived using multistate life tables are causal and applicable to today’s populations. The risks associated with risk factor status that were defined 50 years ago may not be same as those defined today because of differences in the risk factor status in population and treatment. However, the hazard ratios estimated in chapter 8 are consistent with those derived from the more recent studies (Kenchaiah et al., 2002; Wilson et al., 1998).

There have been great improvements in mortality over the past 50 years. However, the Framingham Heart Study cohort is relatively healthy, and its life expectancy is similar to that of the 1990 Massachusetts population (Peeters et al., 2002; Centers for Disease Control and Prevention, 1989-91). Although the results presented here will not represent the absolute figures of today’s populations, they are a robust estimation of the relative magnitude of the life table outcomes (e.g. number of years lived with CVD) by risk profiles.

The major limitations of this chapter are that we could not identify what proportion of the increased life expectancy with cardiovascular diseases or decreased life expectancy free of cardiovascular disease has directly resulted from a high risk at middle age, and would therefore be preventable through high-risk prevention. In the multivariate analysis, we did account for major risk factors of cardiovascular disease.

**Public health implications**

For individuals with favorable levels of blood pressure, total cholesterol levels, BMI and those who do not smoke, the long-term incidence of cardiovascular disease and mortality is much lower and longevity is much greater. On average, 4 in 5 males or females with favorable levels of cholesterol, blood pressure, BMI who do not smoke throughout the age interval of 30-50, and who remain free of cardiovascular disease until age 50 would still be alive and free of cardiovascular disease twenty years later. The same holds for a mere 2 or 3 in 10 members of the high-risk group. This reemphasizes the fact that the presence of major risk factors such as high BP, a high SCL and BMI poses a major public health burden.
These remarkable differences in life expectancy with and without cardiovascular disease between low and high-risk profiles indicate that risk factor status at middle age of life has a significant impact at later ages. Low risk in adulthood compresses cardiovascular morbidity and high risk expands cardiovascular morbidity. The findings reinforce the guidelines of American Public Health Association (Pearson et al., 2002) and the recent life course epidemiology approach (WHO, 2002). We have provided some empirical evidence on the impact of major risk factors in middle age on the cardiovascular life course. It is widely accepted that these risk factors are preventable through lifestyle modification (Pearson et al., 2002, WHO, 2002). Therefore, the high-risk group of adults needs to be motivated to change their lifestyle, keeping in mind that “prevention is better than cure”, in order to stay healthy, save lives and increase healthy life.

References


10

Discussion and conclusion

10.1 Introduction

The focus of this book was to explore the utility of the multistate life table technique in public health research and to investigate the life history of cardiovascular disease, its risk factors and the compression of morbidity. A substantial portion of this book has been devoted to delineating of the methodological issues in public health research. The multistate life table method was investigated and consistently applied to the lengthy period of follow-up of the original ‘Framingham Heart Study’ cohort. The general approach was to couple demographic methods with an epidemiological framework to develop new public health and life course indicators, which would be transparent and policy relevant. To some extent, this approach also shed light on the empirical investigation of the life course epidemiology of cardiovascular disease. Life course analysis is a new tool with a huge potential for public health epidemiology. This study pioneered the use of life course analysis on the Framingham Heart Study. This study exemplified that risk is an inevitable part of the human life course (Ben-Shlomo and Kuh, 2002).

The cardiovascular risk factors at middle age and their long-run impact on the disease history and post-disease mortality were investigated explicitly. The burden of cardiovascular disease was analyzed by risk factor status. We tried to identify which risk factor compressed cardiovascular disease and at which age compression occurred. We also estimated the multifactorial effects in the cardiovascular life history. The major strength of this study lies in its focus on a well- documented epidemiologically defined community-based population and 48 years of prospective, consistent follow-up of the same cohort in the Framingham Heart Study, and the combination of well-developed techniques from the discipline of epidemiology and demography.

This concluding chapter summarizes the major contributions of this research and draws attention to the future potentials of the cardiovascular disease history. An elaborate discussion of each chapter study can be found in the discussion sections of the separate chapters. In Section 10.2, we have summarized the major
findings of this study. The important contributions of this research, both methodological and substantive, are illustrated in Section 10.3. The implications of this study for public health research are briefly described in Section 10.4. Some of the future prospects of this study are indicated in Section 10.5. Finally, we bring our study to a close look with a concluding remark in Section 10.6.

10.2 Summary of findings

We applied multistate life table technique consistently throughout this study. The major strength of the multistate life table is its ability to capture the implications of age-related declines and improvements in health (i.e. inflow and outflow). This approach provides a more accurate assessment of the expected life cycle of disease history in the population compared to the prevalence-based measure (Sullivan method). We found the multistate life table technique to be an elegant method to address the debate of compression or expansion of morbidity. We illustrated the multistate life table technique with its generalization and statistical precision in a simplified way. The methodological features of the MSLT that might serve as an added contribution to the field of public health research was pointed out in Chapter 2 of this book.

From a methodological point of view, the analyses performed in Chapter 3 were of interest because:

- the chapter provided a detailed description of how the exact occurrence-exposure rates could be derived from the micro level information from longitudinal follow-up data;
- they transferred the transition rates into a multistate life table to analyze the life history of cardiovascular disease;
- the results exhibited the utility of transforming epidemiological data into time-based health policy measures; and
- it analyzed the possible model structure or state space for cardiovascular disease processes.

By transferring age-specific occurrence-exposure rates into a life table population, it was found that at age 40, FHS participants could expect to survive an average of 38.5 years, of which 84 percent free of cardiovascular disease. From the age of 40 on, nearly two-thirds of the men (67 percent) and more than half (55 percent) of the women would go on to develop cardiovascular disease in their lifetime. A disparity in the life trajectory of cardiovascular disease between males and females was observed. It was found that males spend more years of life with cardiovascular disease and its subtypes at middle age, but that the burden of cardiovascular disease at older ages (80+) was higher for females. We showed that the greater longevity of women was the primary cause of both their greater lifetime
DISCUSSION AND CONCLUSION

probability of stroke and the greater number of years of life lost to an equivalent
disease as compared to men. While epidemiological data enables prediction of the
number of coronary heart disease events and deaths, the multistate life table
technique enables estimation of the overall potential burden of specific diseases in
terms of years of life lost to and lived with disease. In chronic disease modeling,
this approach and the use of a life course indicator are new, and we found these to
provide very transparent and robust outcomes that are furthermore relevant to
public health policy.

We proposed a potential method to impute missing values of risk factors in a
repeated measurement study (Chapter 4). The idea behind the proposed
methodology is relevant to any study design subject to incomplete responses and
having at least three repeated measurements. Using the proposed methods, we can
easily reconstruct the risk career and measure the impact of this risk career on
disease incidence and mortality. We applied this technique to impute missing values
for smoking status in the FHS and to reconstruct the smoking career of the FHS
population (Chapter 5). The analysis was intended to illustrate the usefulness of
multistate life tables for describing how smoking evolves over the life course. Our
smoking status definition included the updated information on smoking status
throughout life rather than the status at a specific point in time, for instance,
incidence of smoking. We described the smoking experience of an American cohort
from the 1950s and translated the smoking career of this cohort into life years spent
as non-smokers and life years spent smoking, within a synthetic cohort. At age 10,
the total life expectancy of the FHS cohort was 66 years for males and 72 years for
females. Males smoked for considerably more years than females. Males spent two-
thirds of their expected lifetime (68%) smoking. For females, this was less than one
third (28%). The difference declined at higher ages when males with a history of
smoking had either quit or died. The smoking careers of males started 2 years
earlier than that of females. Males started, on average, at age 15 (median) and
females at age 17. Fifty-eight percent of the American males who started smoking
manages to quit at some point. For females, the proportion of quitters was
considerably lower, namely 37%. Compared to men, women smokers were more
persistent. An individual who quit smoking was likely to do so between the ages of
50 and 70. The probability of a relapse was 26% for males and females. This
knowledge of the historical smoking and quitting patterns is important for current
studies interested in the cumulative smoking histories of current populations. Our
approach demonstrated one of the ways the changes in risk factors throughout life
could be explicitly taken into account. This method could also be used for other
risk factors to indicate current risk factors in life courses.

When investigating cardiovascular disease in the life course of smokers and
non-smokers (Chapter 6), we found that the risk of developing cardiovascular
disease before age 70 was higher among smokers. We classified never smokers as
those with all available smoking records coded as a “non-smoker” and always smokers as those with all available smoking records coded as a “smoker”. Associated with the longer life expectancy of male non-smokers were higher lifetime risks of coronary heart disease, myocardial infarction, stroke and congestive heart failure. Female non-smokers had higher lifetime risks of coronary heart disease and congestive heart failure. Non-smokers spent more years with cardiovascular disease as well as free of cardiovascular disease over the life course because they live longer. Not smoking will not compress cardiovascular morbidity, but it will postpone it to older ages. Smoking, by shortening life, decreases the years lived with cardiovascular disease throughout the life course. Paradoxically, in a non-smoking society, more people will live with cardiovascular disease but this will be concentrated at the end of life.

Two well-known rate models, the Cox and the Gompertz model, were compared with the empirical estimates of age-specific transition rates (Chapter 7). In the null model, the Cox and the observed rates were similar. When covariates were added to the Cox model, the estimated age-specific rate did not overlap with the observed rates. Instead, it fit less within the bounds because of its proportionality assumption. The Gompertz estimate behaved the same in the null model and in the model with covariates. Gompertz has fewer parameters, gives smooth transition rates and reduces the variability. On the multistate life tables of cardiovascular disease, with and without the presence of covariates or risk factors, the Gompertz model could be a better option compared to the observed rates and the Cox model.

We investigated the association of the risk factor status between the ages of 30 and 50, with the incidence of cardiovascular disease and its subtypes, and post-disease mortality over a long follow-up (Chapter 8). We considered the following major cardiovascular disease risk factors: smoking, body mass index, serum cholesterol level and blood pressure. Smoking status for each participant was allocated based on the current smoking status recorded at each available exam from age 30 to 50. We classified never smokers as those with all available smoking records coded as a “non-smoker” and always smokers as those with all available smoking records coded as a “smoker”. Ever smokers were the rest of the participants, characterized by a mixture of smoking and non-smoking throughout the period from entry into the survey to age 50. Blood pressure (systolic or diastolic) at each exam was defined based on the mean value of recorded blood pressure from different examiners. We took an average of the recorded mean blood pressure between age 30 and 50 years. Likewise, the average levels of serum cholesterol and body mass index (BMI) recorded between age 30 and 50 were taken as predictors. Univariate analysis indicated that each risk factor had a substantial effect on the risk of CVD and mortality. For instance, obese males (BMI ≥ 30) had an 81% excess risk of CVD and a 47% excess mortality compared to males of normal weight. We investigated the long-term nature of relationships, not only of
primary relationships as shown in univariate analyses, but also of confounding relationships as shown in the changes of relative risk after adjusting for the levels of other risk factors. The effect of smoking is aggravated by the presence of other risk factors. For example, when we controlled for the level of other risk factors in male smokers, the upward change in excess risk was 29% for CVD. The presence of other risk factors reduced the effect of obesity, hypertension, or high cholesterol level on CVD (in the absence of any interaction effect). The direction in which the relative risk changes (e.g. upward for smoking and downward for obesity) is likely to be the result of the pattern of co-occurrence of risk factors in the population.

We examined whether optimal risk factor status throughout the age interval 30-50 compresses cardiovascular morbidity. Since one risk factor could compress cardiovascular morbidity and others might not, we performed both a univariate and multivariate analysis to examine the independent and partial impact on cardiovascular disease life history in terms of life table estimates (Chapter 9). In the case of the univariate analysis, across all risk factor categories, it was found that a male who did not smoke and a female who had optimal blood pressure at middle age had the highest life expectancies at age 50: 30.82 years for males and 34.72 years for females. For both males and females, the absence of major cardiovascular risk factors at middle age compressed cardiovascular morbidity. Non-smoking at middle age does not necessarily compress cardiovascular morbidity. Apparently, the balance between effects on incidence, case fatality and recovery of cardiovascular disease differs between smoking and the other risk factors.

We demonstrated the possibility of using multivariate Gompertz regression models for the long time continuous event histories to estimate the state transition rates and then to use the estimated transition rates in constructing a multistate life table in multiple covariate contexts. We compared optimal and high-risk profiles. In the multivariate case, an optimal risk profile was defined as an individual who was a never smoker, on average had optimal blood pressure (SBP<120 and DBP<80), optimal cholesterol level (SCL<200) and optimal BMI (BMI<25) between age 30 and 50. If an individual was a smoker, had high blood pressure (BP>140 or DBP>90), high cholesterol (SCL>240) and was obese (BMI>30), he or she was considered to have a high-risk profile. Multivariate analysis showed that an average of four in five males or females belonging to the optimal risk group, who were free of cardiovascular disease at age 50, would still be alive and free of cardiovascular disease twenty years later; only 1 or 2 in 5 of those considered to be at high risk, would remain in this state. The differences in additional number of survival years between males and females by optimal and high-risk profiles were found to be 12 and 14 years. A fifty-year-old male with an optimal risk profile could expect to survive 17 more years free of cardiovascular disease and a female of that age with same risk profile could expect to survive 16 more years, compared to individuals with high-risk profiles. For the synthetic cohort that was free of CVD at age 50, the
The lifetime probability of developing cardiovascular disease was 82 percent for males (female 72 percent) in the high-risk group and 38 percent for males (35 percent females) in the low risk group.

The significant differences between optimal and high-risk profiles in life expectancy with and without cardiovascular disease indicates that risk factor status at middle age has a significant impact on later ages of life. High-risk profiles in adulthood shorten the duration of life, increase the lifetime probability of experiencing CVD, and extend the length of time spent living with cardiovascular disease. An optimal risk profile significantly increases the number of years lived without cardiovascular disease and decreases the years lived with disease, i.e. optimal risk profiles in middle age are associated with the compression of cardiovascular morbidity.

The findings provide empirical evidence of the importance of the prevention of major cardiovascular risk factors at adulthood and reconfirm the guidelines of the American Public Health Association (Pearson et al., 2002) and the recent life course epidemiology approach (WHO, 2002). In public health research, major priority should be given to targeting the high-risk group of middle-aged adults and motivating them to change their lifestyle.

10.3 Major contributions of this study

The present study has attempted to contribute to the methodology of public health research and to derive life course indicators for the cardiovascular disease history, focusing on the compression of morbidity hypothesis. We initiated a new public health approach to the compression of morbidity debate, coupling demographic techniques to the framework of epidemiology. Relating risk factor status with the disease incidence in the debate on the compression of morbidity is relatively new. Researchers hitherto mostly addressed this debate by linking either disability (Nusselder, 1998; Rogers et al., 1989; Crimmins et al., 1994) or degenerative disease (Barendregt and Bonneux, 1998) to the older population and their mortality. We focused on the adult risk factor and the compression of cardiovascular morbidity, and were first with the change of risk factor status at adult age and the compression of cardiovascular morbidity. Some of the methodological and substantive contributions that are made in this study are described below.
Methodological
The methodological contributions of this research can be summarized in at least six points:

1. We made an explicit distinction between prevalence and incidence based measures of population health status, and shed light on the utility of the multistate life table to address the compression or expansion of morbidity. We generalized the multistate life table technique and consistently applied it to estimate the compression of cardiovascular morbidity by risk factor status. The statistical competence of the multistate life table outcomes was tested using a non-parametric bootstrapping technique. We consistently analyzed the 48 years of follow-up data from the Framingham Heart Study. To our knowledge, no one has hitherto analyzed the follow-up data over such a lengthy period in a multistate life table framework.

2. A methodology to measure the combined effect of several risk factors in the life history of cardiovascular disease was proposed, which is the added contribution in both the life table and cardiovascular risk factor analysis. Using the methodology developed by Anderson et al., (1991) and Wilson et al., (1998) the probability of developing cardiovascular or coronary heart disease in the shorter term can be measured, using our approach estimates can be made of the lifetime probability of developing cardiovascular disease, the number of years lived with and without disease.

3. We proposed potential techniques to impute missing values for both the categorical and continuous risk factors in the repeated measurement study.

4. We assessed the well-known and widely used transition rate models, Cox and Gompertz, against the empirical occurrence-exposure rates and life table outcomes. On the multistate life table of cardiovascular disease, with and without the presence of covariates or risk factors, the Gompertz model could be a better option than the observed rates and the Cox model.

5. One way in which the risk career can be analyzed was demonstrated using the very common risk factor of smoking. Our novel approach demonstrated the importance of explicitly taking into account the changes in risk factors throughout life. Other risk factors can be analyzed in the same fashion, enabling researchers to gain more insight into the impact of other risk factors throughout life.

6. We used age as a time variable, both in the multistate life table and when estimating the relative risks. Disease incidence, risk factor exposure and their relationship, depend strongly on age. To estimate the relative risk of disease or mortality, most researchers use age at baseline, an independent variable that they call age adjusted relative risk. In survival analysis, most of the researchers use follow-up time as the time variable, instead of using age at transition.
Substantive
The substantive contributions made in this study are summarized in the following points:

1. Using a multistate life table, we derived new indicators for the cardiovascular disease life history. At age 40, an FHS participant could expect to live an average of 38.5 years, of which 84 percent free of cardiovascular disease. At age 50, a male could expect to spend 25 percent of his remaining life expectancy with the disease; females, 18 percent. Although males survive longer with CVD during middle age, the burden of CVD at later ages is higher for females.

2. Although the lifetime risk of developing coronary heart disease was previously estimated by Lloyd-Jones and Levy (1999), in the present study we have estimated the lifetime risk of developing cardiovascular disease and its sub-types: coronary heart disease, acute myocardial infarction, angina pectoris, congestive heart failure and stroke. On the basis of the data obtained from the FHS, the lifetime probability for a 40-year-old male and female American without CVD of developing CVD, CHD, MI, CHF and stroke was calculated to be 67% vs. 55%, 50% vs. 33%, 33% vs. 17%, 18% vs. 16% and 20% vs. 18%.

3. By analyzing the smoking career, we derived different episodes of smoking, patterns of quitting, of restarting, of number of years lived as smokers and non-smokers and the lifetime probability of quitting smoking which are the new indicators in public health research. We re-created the smoking careers of American males and females followed from the 1950s. Over the life course, females quit smoking 21 percent less compared to males. The restarting rate in males was twice that of females. Over the lifetime, more than 50 percent of the American smokers did not quit smoking; individuals who did quit were likely to do so between the ages of 50 and 70. We found that a knowledge of the historical smoking and quitting patterns could be important for current studies interested in the cumulative smoking histories of current populations.

4. The risk of developing any cardiovascular disease before age 70 is higher among smokers. Associated with their longer life expectancy, male non-smokers have higher lifetime risks of coronary heart disease, myocardial infarction, stroke and congestive heart failure, while female non-smokers have higher lifetime risks of coronary heart disease and congestive heart failure. Non-smokers spend more years with cardiovascular disease over the life course. Non-smokers also live more years free of cardiovascular disease. Not smoking will not eliminate cardiovascular disease, but it will postpone it to older ages. Smoking, by shortening life, decreases the years lived with cardiovascular disease throughout the life course. Paradoxically, in a non-smoking society, more people will live with cardiovascular disease but this will be concentrated at the end of life.

5. We investigated the association of the risk factor status between ages 30 and 50, with the incidence of cardiovascular disease and its subtypes, and post-disease
mortality over a long-follow-up. The major risk factors at middle age are strong
predictors of incidence of cardiovascular disease and mortality at older ages. The
effect of smoking is upward when the confounding effect of other risk factors is
present. The effect of obesity, hypertension or high cholesterol is downward
when the confounding effect of other risk factors is removed. In the FHS
population, smokers are probably less exposed to other risk factors than non-
smokers, which is why the relative risk goes up; obese people more than non-
obese people, which is why the relative risk goes down. Smoking prevents
obesity, and obesity leads to high serum cholesterol and high blood pressure.

6. Contributions from the multistate life table analyzed by univariate risk factor
status at middle age can be summarized as follows:

- Tobacco use at middle age decreased life expectancy at age 50, by 4.4 years
  for males or females. Male or female non-smokers survive 4.2 or 3.0 years
  longer without CVD at age 50, and 0.3 or 1.4 years more with CVD. Male
  or female smokers spend more years with CVD before age 75 or 65. Non-
  smoking at middle age does not necessarily compress cardiovascular
  morbidity.

- High blood pressure at middle age decreased life expectancy at age 50, by
  5.7 years for males and 4.9 years for females. Males or females with optimal
  BP survive 6.8 or 7.8 years longer without CVD at age 50 and 1.1 or 3.0
  years less with CVD. Males with high blood pressure spend more years with
  CVD before age 78. Females with high blood pressure spend more years
  with CVD throughout life. Optimal BP at middle age compresses
  cardiovascular morbidity.

- High cholesterol at middle age decreased life expectancy at age 50, by
  2.2 years for males or females. Males or females with optimal SCL survive 4.82
  or 3.0 years longer without CVD at age 50 and 3.7 or 0.6 years less with
  CVD. Males with high serum cholesterol levels spend more years with
  CVD throughout life and females with high serum cholesterol spend more
  years with CVD before age 82. Optimal SCL at middle age compresses
  cardiovascular morbidity.

- Obesity at middle age decreased life expectancy at age 50, by 4 years for
  males or females. Males or females with normal weight survive 5.4 or 5.6
  years longer free of CVD at age 50 and 1.6 or 1.8 years less with CVD.
  Both obese males and females spend more years with CVD before age 82.
  Normal weight at middle age compresses cardiovascular morbidity.
7. Contributions made by the multivariate risk factor analysis are:

- Fifty-year-old males with an optimal risk profile at middle age can expect to survive 6 additional years compared to the total male respondents in the FHS. This indicates that American male life expectancy could be extended by 6 years (i.e. male life expectancy would be 83 years) if males were all to belong to the optimal risk group between the ages of 30 and 50. Similarly, fifty-year-old females with an optimal risk profile at middle age, can expect to survive 5 additional years compared to the total female respondents in the FHS. This indicates that American females can expect to survive 5 years more (i.e. female life expectancy would become be 87 years) if they were all to belong to the optimal risk group between the ages of 30 and 50.

- A 50-year-old man with an optimal risk profile can expect to live 17 years longer without cardiovascular disease than a member of the high-risk group. A woman of that age with an optimal risk profile can expect to survive 16 additional years free of cardiovascular disease.

- Individuals who are free of CVD at age 50 have a lifetime probability of developing cardiovascular disease of 82 percent if they are a high-risk male, and 69 percent if they are female and belong to the high-risk group, and 38 or 35 percent if they belong to the optimal risk group.

- High-risk profiles at adulthood shorten the duration of life, increase the lifetime probability of experiencing CVD, but extend the period of life spent with cardiovascular disease. An optimal risk profile importantly increases the number of years lived free of cardiovascular disease i.e. an optimal risk profile compresses the cardiovascular morbidity. Males with high-risk profiles spend more years with CVD before age 83 and females with high-risk profiles spend more years with CVD before age 81.

Limitations

This research had some limitations as well. The limitations are listed below:

1. To construct an MSLT for the cardiovascular disease or chronic disease, we need micro-level longitudinal information. The Framingham Heart Study is unique in its length of follow-up. However, the problem with the FHS is precisely its length; a very long follow-up study means that the first stages of data are very old, leading to questions about the validity for current populations. The innovative ways to combine other data sources will be a new area of research if policymakers wish to make decisions based on something other than cardiovascular disease.

2. We used longitudinal follow-up data on a historical cohort. The risk factors status and disease incidence changes over the time. However, we took this into account in different ways: smoking status throughout life (Chapter 6) and risk factor status between the ages 30-50 (Chapters 8 and 9). A key area of future
research will be to define risk factor status prior to outcome, taking into account changes in risk factor status over time. Defining risk factor status is almost impossible if there is more than one outcome, such as CVD and death, where smoking status is only defined at middle age and may represent status prior to CVD, but not death.

3. In modeling cardiovascular disease, we did not consider back transitions. Return transitions were possible in the model structures illustrated in Chapter 3, except for models 3(a) and 3(b). If such data is available, however, the method we have described could be used for any situation (as demonstrated in Chapter 5).

4. Disease severity, co-morbidity and the quality of life after having cardiovascular disease was not considered. However, we considered the fatal cases. Our conclusions regarding the compression of morbidity are restricted to cardiovascular morbidity and do not tell us whether these risk factors would also lead to a general compression of morbidity.

5. To measure the impact of risk factors, we mainly considered the major adult risk factors. Many of the established risk factors are acquired in childhood; exposures in foetal life, infancy, childhood, and adolescence, can strongly affect risk of chronic diseases later in life, such as cardiovascular diseases (WHO, 2002; Ben-Shlomo and Kuh, 2002). This was not considered in this study. However, we believe that our described novel approach can be applied for any population in the same fashion.

10.4 Implications for public health

In this research, we have initiated a new public health approach to the compression of morbidity debate, in which demographic techniques are coupled to the framework of epidemiology. We recommend using such methods to assess the health impact of risk factors and interventions. The life table approach can integrate the effect of risk factors (at a specific point or at different stages of life) on disease occurrences, age at developing disease, years lived with disease, years free of disease, lifetime risks of developing disease and mortality. This approach can provide a synthetic life course measure of population health status. Using short-term or long-term follow-up data this method can be used to reconstruct an individual biography (micro perspectives) or the age spectrum (population health perspectives). This method also provides the range of outcome measures necessary to better understand the individual and population health consequences of potential prevention or intervention of risk exposures. The life table indicators, such as those presented in this study, can help inform choices between strategies for improving the health of aging populations. These indicators are also important from the perspective of an individual making choices about a change in lifestyle.
The findings presented in this thesis indicate that people with a high risk profile at middle age spend much time with CVD and survive fewer years than those in the optimal risk group. Risk factors in middle age strongly affect lifetime CVD, which is why we suggest that more emphasis should be given to prevention prior to middle age. However, the effect of eradicating a risk factor cannot be simplistically thought to lead to less chronic disease in society or throughout life. The effect obviously differs for different risk factors. For instance, this study shows that the impact of smoking on cardiovascular morbidity is different from other major risk factors, such as serum cholesterol, blood pressure and body mass index. Therefore, in addition to the balance between increased risks of chronic disease, disability and mortality, a greater understanding of the role of risk factors throughout life is needed to decrease the burden of chronic disease, disability and mortality.

We recommend that the new indicators such as life expectancy with disease, life expectancy free of disease or lifetime probability of disease given the risk factor status at a specific age or at different stages of life, be taken into account when drafting public health campaigns. These types of indicators form the health benefits in the context of the life course and are more easily understood than traditional risk measures such as odds ratios or relative risks. Our approach offers a new tool with great potential for public health epidemiology. We need similar approaches to be applied to changes in risk factor status, to predict the most beneficial public health strategies.

10.5 Some future prospects

This study demonstrated that the multistate life table could be used as a tool to analyze the life history of cardiovascular disease and its risk factors. Using the same method, it would be possible to derive life course indicators in any other life course research. Adding health interventions, health care cost, health care quality, disease severity and co-morbidity to the illustrated multistate life tables, public health researchers, clinical decision makers and health policymakers can easily transfer the epidemiological or disease information into time–based health policy measures. Different risk factors have different treatments and health benefits. The number of health benefits from a certain treatment that could possibly be achieved could be an extension of our present study.

This study shows that the presence of any major (single or multiple) risk factor at middle age has enormous impact on the life history of cardiovascular morbidity in terms of lifetime probability of disease, number of years spent free of disease and number of years spent with disease. We observed that more than one third (38% for males and 35% for females) of the people who had optimal risk profiles as adults developed cardiovascular disease at older ages. This development could be
caused by other adult risk factors (e.g. diabetes), genetic factor, environmental factors, and aging itself. However, this study has demonstrated the methods and shown how to estimate the impact of risk factor status in disease incidence and mortality in terms of the life table outcomes. Therefore, when the information on other risk factors becomes available, it can be investigated in the same way. For instance, by combining the three Framingham generations (Framingham original cohort, Offspring cohort and third generation) and using our methodology, more insight and mechanisms could be gained into the life course of cardiovascular disease.

We did not investigate the mechanisms of why smoking was different from the other risk factors. This is something to be investigated in the future, using our methodology in order to identify characteristics of risk factors which predict whether they will be associated with compression or expansion of morbidity. We investigated only cardiovascular disease, but the risk factors of cardiovascular disease are also accountable for many other chronic diseases and disability (Norrish et al., 1995). However, using the same approach and with the appropriate data, we can analyze the life history of other chronic diseases and address the debate of compression of morbidity hypothesis.

The life course effects of smoking, obesity, hypertension and cholesterol have been evaluated in a stationary cohort life table; further research exploring the same methodology will be able to extend life course methodology to assess changes in the life course: quitting smoking, losing weight, changing diet, treating hypertension.

A recent study conducted in the United States found that, although serum cholesterol levels declined significantly during the period 1970s and 1980s across the population, no further decline in serum cholesterol has occurred over the past decade (Arnett et al., 2002). There have been upsetting trends in the health status of teenagers (Surgeon General, 2001), among whom there are troubling increases in the prevalence of cigarette smoking (Murray and Lopez, 1996), obesity (Mokdad et al., 2001; Fontaine et al., 2003; WHO, 1998) and decreases in participation in physical activity programs (Surgeon General, 1996). Recent investigation shows that southern Asia (representing one quarter of the world population) is likely to be more strongly affected by the increase in cardiovascular disease in the future (Nishtar, 2002). Therefore, globally, cardiovascular disease will play an ever-growing role as a major cause of morbidity and mortality (WHO, 2002; Hennekens, 1998; Murray and Lopez, 1996). Thus the alarming trends and backslide in the health status of growing teenagers, the risk factor profiles of young adults and the emerging pandemic of cardiovascular disease all underscore the critical need to intensify and to give high priority to more efficient prevention and treatment in public health.
10.6 Concluding remarks

Except for smoking, other major cardiovascular risk factors demonstrated that cardiovascular disease could be postponed by adopting a healthy lifestyle. This was the second proposition of Fries’ (1980) compression of morbidity hypothesis. Not smoking will not compress cardiovascular morbidity, but it will postpone it to older ages. Belonging to the optimal categories of blood pressure, optimal categories of serum cholesterol levels or normal weight at adult life compresses cardiovascular morbidity. Similarly, belonging to the combined optimal risk (optimal blood pressure, optimal cholesterol, normal weight and never smokers) profiles at middle age also compresses cardiovascular morbidity to a larger extent compared to the single risk factor.

In chronic disease modeling, our approach and the use of a life course indicator is new, and these provide very transparent outcomes that are public health policy relevant. This study also shows the health benefits of belonging to optimal risk profiles as an adult and the relationship of risk factor status to the compression of morbidity hypothesis. Since life course research is a new focus in scientific community, our described methodology can be used as a tool to measure the life course indicators, not only in public health but also in any other discipline.

With consistent application of the multistate life tables in public health research, the population health status, health care costs, quality of life and other dimensions of health could be measured. We need appropriate crucial prevention or intervention on major cardiovascular risk factors at middle age. This prevention or intervention will have the potential to bring about a significant reduction in long-term risk and to increase the number of years spent without cardiovascular disease in population health.

References


DISCUSSION AND CONCLUSION


Samenvatting

Inleiding
Deze studie wil bijdragen aan de ontwikkeling van een instrumentarium voor volksgezondheidsonderzoek in het algemeen en voor onderzoek naar hart- en vaatziekten in het bijzonder. Uitgangspunt is dat een integratie van methoden en technieken uit de demografie en de epidemiologie tot inzichten kan leiden die daarvoor niet of moeizaam konden worden verkregen. In deze studie worden ziekten en gezondheid benaderd vanuit een levensloopperspectief. De keuze voor een levensstijl kan de kans op chronische ziekten op latere leeftijd beïnvloeden. De levensloopbenadering biedt belangrijke perspectieven voor epidemiologisch onderzoek en volksgezondheid. De data voor het onderzoek komen van de Framingham Heart Study (FHS), een longitudinaal epidemiologisch onderzoek over een periode van 48 jaar van 5.209 inwoners van Framingham, Massachusetts, VSA. Het doel van dat onderzoek was het inzicht te vergroten in het onstaan en beloop van hart- en vaatziekten. De FHS heeft geresulteerd in meer dan 1.000 artikelen in wetenschappelijke tijdschriften, maar de levensloopbenadering waarbij demografische en epidemiologische technieken worden gecombineerd, is nieuw. In deze studie wordt de invloed van vier risicofactoren onderzocht: rookgedrag, bloeddruk (BP), serumcholesterolgehalte (SCL) en zwaarlijvigheid uitgedrukt in body mass index (BMI). Vier typen hart- en vaatziekten worden onderscheiden: myocardinfarct (MI), coronaire hartziekten (coronary heart disease CHD), chronische reumatische hartaandoeningen (congestive heart failure CHF) en beroerte (stroke). Nieuw in dit onderzoek is de situering van risicofactoren in de levensloop. Bijvoorbeeld, zwaarlijvigheid op jonge leeftijd heeft andere effecten op de incidentie van hart- en vaatziekten dan zwaarlijvigheid op hogere leeftijd. Het rookgedrag wordt bepaald door de leeftijd waarop iemand begint met roken, het aantal jaren dat wordt gerookt, en het aantal sigaretten per dag. Op basis van die longitudinale data en aan de hand van aangepaste modellen voor longitudinaal onderzoek wordt de invloed van risicofactoren op de incidentie van de verschillende typen hart- en vaatziekten en van hart- en vaatziekten in het algemeen. De individuele consequenties van veranderingen in incidentie worden geraamd aan de hand van de meerdimensionele overlevingstafel (multistate life table) uit de demografie.

Methode
Drie typen modellen voor de raming van effecten van risicofactoren op incidentie worden onderscheiden: de niet-parametrische, de semi-parametrische en de parametrische methoden. Tot de eerste categorie behoren de Kaplan-Meier schatter van de overlevingsfunctie en de Cox regressiemodel dat wordt gebruikt om relatieve risico’s ter schatten
of te voorspellen zonder daarbij rekening te houden met de leeftijdsafhankelijkheid van het relatieve risico. Tot de derde categorie behoort het Gompertz model dat het leeftijdsafhankelijke relatieve risico voorspelt. Het model gaat uit van een exponentiële verandering van incidentie met leeftijd. Het Gompertz model levert een goede beschrijving van het incidentiecijfer naar leeftijd en daardoor van relatieve risico’s in de verschillende fasen van de levensloop. De studie van de leeftijdsafhankelijkheid van relatieve risico’s voegt een nieuwe dimensie toe aan epidemiologisch onderzoek. Relatieve risico’s veranderen omdat zowel risicofactoren veranderen (bijvoorbeeld stoppen met roken, lijnen, gebruik bloeddrukverlagers) als de effecten (bijvoorbeeld stoppen met roken resulteert in een snelle daling van het relatieve risico op hart- en vaatziekten). Na 10 jaar bestaat geen verhoogd risico meer.

De individuele consequenties van veranderingen in incidentie worden geraamd aan de hand van de meerdimensionele overlevingstafel. In vergelijking tot gangbare methoden biedt de meerdimensionele overlevingstafel een aantal voordelen. In de eerste plaats wordt goed rekening gehouden met concurrerende gebeurtenissen die de gevolgen van risicofactoren op hart- en vaatziekten mede beïnvloeden. Concurrerende gebeurtenissen zijn vooral van invloed op de kans op hart- en vaatziekten. De kans op een aandoening kan aanzienlijk worden overschat wanneer geen rekening wordt gehouden met concurrerende gebeurtenissen. In de tweede plaats geeft de overlevingstafel een raming van de leeftijd bij hart- en vaatziekte en van het verwacht aantal jaren met de aandoening. Dit laatste is van belang niet alleen voor de patient maar ook voor de uitgaven voor gezondheidszorg.

Resultaten
Het is bekend dat roken een grote invloed heeft op de kans op hart- en vaatziekten. Roken vermindert ook de levensverwachting. Een persoon van leeftijd 30-50 die rookt leeft gemiddeld 4,4 jaren korter dan niet-rokers. Dat geldt voor zowel mannen als vrouwen. Aangezien rokers minder lang leven, hebben zij minder jaren te lijden aan hart- en vaatziekten niettegenstaande de kans op de aandoening groter is dan bij niet-rokers. Een gevolg is dat niet-rokers meer jaren lijden aan de aandoening dan rokers. *Niet roken leidt dus niet tot een compressie van de morbiditeit.*

Hoge bloeddruk is een andere risicofactor. Hoge bloeddruk (systolische bloeddruk > 140 mmHg en/of een diastolische bloeddruk > 90 mmHg) op middelbare leeftijd (30–50 jaar) vermindert de levensverwachting op leeftijd 50 met 5,7 jaren voor mannen en 4,9 jaren voor vrouwen. Mannen en vrouwen met een normale bloeddruk (systolische bloeddruk minder dan 120 mmHg en diastolische bloeddruk minder dan 80 mmHg) hebben aanzienlijk minder last van hart- en vaatziekten. Een 50-jarige man met normale boeddruk heeft 6,8 jaren minder last van CVD dan een man met een hoge bloeddruk. Voor vrouwen bedraagt het verschil 7,8 jaren. *Een normale bloeddruk op middelbare leeftijd leidt tot een compressie van cardiovasculaire morbiditeit.*

Hoog cholesterolgehalte (serumcholesterol van 240 mg of meer per deciliter) op middelbare leeftijd vermindert de levensverwachting met meer dan 2 jaren en vergroot de kans op hart- en vaatziekten. Mannen van middelbare leeftijd met een norma cholesterolgehalte (serumcholesterol van 160 – 199,9 mg per deciliter) blijven na hun vijftigste 4,8 jaren langer gezond (vrij van CVD) en wanneer zij de aandoening krijgen gebeurt dat op een hogere leeftijd minder (3,7 jaren later). Voor vrouwen bedragen de cijfers 3,0 en 0,6 jaren. *Normaal cholesterolgehalte op middelbare leeftijd leidt tot een compressie van cardiovasculaire morbiditeit.*

Overgewicht leidt tot een kortere levensduur en meer hart- en vaatziekten. De levensverwachting vermindert met 4 jaren voor zowel mannen als vrouwen. Overgewicht wordt gedefinieerd als een body mass index (BMI) van 30 kg/m² of hoger. Mannen en vrouwen met een normaal gewicht blijven na hun vijftigste langer gezond (vrij van CVD): 5,4 jaren voor mannen en 5,6 jaren voor vrouwen. Krijgen ze toch CVD, dan gebeurt dat op latere leeftijd. *Een gevolg is dat een normaal gewicht leidt tot compressie van cardiovasculaire morbiditeit.*

De vaststelling dat gezonde mensen langer leven is niet nieuw. Nieuw is echter de quantificering op basis van een longitudinaal onderzoek van bijna 50 jaren en de berekening van de bijdrage van de verschillende risicofactoren. Wat mag een gezonde vijftigjarige verwachten? Wat betekent een gezonde levensstijl en het gevrijwaard blijven van bijzondere risico’s zoals hypertensie? Voor mannen van 50 betekent dat 6 jaar langer leven dan de gemiddelde man van die leeftijd (d.w.z. gemiddelde van alle 50-jarigen). De FHS geeft aan dat mannen met een gezonde levensstijl en vrij van risico’s als hypertensie een levensverwachting hebben van 83 jaar, aanzienlijk meer dan die van mannen met een minder gezonde levensstijl en/of

De gepresenteerde indicatoren, zoals de kans op hart- en vaatziekten in aanwezigheid van andere aandoeningen en doodsoorzaken, en de levensverwachting vrij van een bepaalde aandoening, zijn belangrijke indicatoren van volksgezondheid. De relatie die in dit onderzoek wordt vastgesteld tussen het niveau van deze indicatoren en de aan- of afwezigheid van risicofactoren in verschillende fasen van het leven biedt nuttige informatie voor een volksgezondheidsbeleid. De effecten van een verandering in rookgedrag en levensstijl, en de effecten van controle van cholesterolgehalte en bloeddruk, kunnen aan de hand van het hier gepresenteerde instrumentarium worden geraamd. Dit onderzoek toont dat een gezonde levensstijl niet automatisch leidt tot een lagere kans op chronische ziekte, maar wel tot een uitstel van de ziekte naar een hogere leeftijd. Een gezonde levensstijl betekent een langer leven en een langer leven in goede gezondheid. Het onderzoek biedt ook goede aanknopingspunten voor verdere analyse waaronder kostenramingen. Aangezien voor iedere leeftijd de ziekteëntoestand wordt beschreven en de rol van risicofactoren in kaart wordt gebracht, en aangezien ziektekosten sterk samenhangen met ziekteëntoestand én leeftijd, kan van het hier gepresenteerde instrumentarium (regressiemodel en meerdimensionele overlevingstafel) een betere kostenraming verwacht worden.
The book series Population Studies aims at disseminating results of research on population trends, in the broadest sense.

Series editorial board: Melinda Mills, Anton Oskamp & Harrie van Vianen.
In memory of Anton Kuijsten.

Manuscripts can be submitted to Rozenberg Publishers, Rozengracht 176A, 1016 NK Amsterdam, The Netherlands. E-mail: info@rozenbergps.com

The following books have been published in the Population Studies series:


This book aims at contributing to the study of chronic diseases and the debate on compression versus expansion of morbidity by applying a new methodology that integrates demography and epidemiology and by deriving indicators of cardiovascular disease history. The long-term impact of four risk factors on cardiovascular disease is investigated using the 48 years follow-up of the original Framingham Heart Study cohort. Four disease types are distinguished: coronary heart disease (CHD), myocardial infarction (MI), congestive heart failure (CHF) and stroke. The risk factors that are studied are smoking, blood pressure, body mass index, and serum cholesterol. These risk factors at middle age are strong predictors of the incidence of cardiovascular disease and mortality at older age. The study shows that a healthy life style postpones cardiovascular disease and may reduce the lifetime risk of the disease.

The human life course offers an exciting new paradigm for multidisciplinary scientific research. The study of individual disease histories and risk factors during different stages of life leads to new insights in the impact of lifestyle on health at older age. The adoption of the new paradigm necessitates longitudinal data and appropriate data analytic techniques. This book demonstrates that a combination of the multistate life table and multivariate techniques of event history modeling provides an effective way to describe, explain and predict disease histories.

Abdullah Al Mamun received MSc in Statistics, University of Dhaka, and MSc in Population Studies, University of Groningen. He worked as a senior research officer (1995-1999) at International Centre for Diarrhoeal Disease Research of Bangladesh (ICDDR,B). On the basis of this book he holds PhD in Demography from the University of Groningen, The Netherlands. Recently he was appointed as a Research Fellow at the School of Population Health, University of Queensland Medical School, Australia.