INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Re-analyses of Framingham data

using time-dependent covariates

Guilin Li

Department of Epidemiology and Biostatistics
McGill University, Montreal

July, 1999

A thesis submitted to the Faculty
of Graduate studies and Research in partial fulfillment of the requirements for
the degree of Master of Science

Abstract

I propose a new approach, based on time-dependent covariates, to assess the impact of within-subject changes in predictors on subsequent mortality, and apply it to re-evaluate the impact of changes in serum cholesterol and smoking status on the coronary heart mortality in the Framingham Heart Study. Time-dependent covariates, representing updated risk factor value or its changes from either the baseline or the most recent measurement are included in two types of multivariable Cox regression analyses. The results reveal that in order to avoid confounding of the effects of changes in risk factor, the model should include a time-dependent variable identifying subjects who developed coronary disease during the follow-up. After adjusting for this variable, a within-subject decrease in cholesterol was associated with a significant reduction of coronary mortality, in contrast to the results of previous studies that did not prevent such confounding.

Résumé

Je propose une nouvelle approche, basée sur des variables dépendantes du temps, à modélisation de changements dans des facteurs de risque, et l'applique pour ré-évaluer les effets du cholestérol et du tabagisme sur la mortalité coronaire dans l'étude de Framingham. Les variables dépendantes du temps, représentant soit la valeur mise-à-jour, soit leur changement, en comparaison de le valeur initiale ou la plus récente, sont incluses dans les deux types de modèles multivariés de Cox. Les résultats démontrent que, afin d'éviter la confusion des effets des changements, le modèle doit inclure une variable dépendante du temps, qui identifie les sujets qui le modèlent ont développé la maladie coronaire pendant le suivi. Une fois cette variable prise en compte, il est démontré qu'une réduction du cholestérol est associée à une réduction significative de mortalité, ce qui contredit des études publiées dans lesquelles on n'a pas éliminé le risque de confusion des résultats.

Acknowledgements

I would like to take this opportunity to express my deepest gratitude to my supervisor—Dr. Michal Abrahamowicz, who gave me so much help during my thesis progress. His encouragement, his dedicated teaching, his patience, his support help me finish this thesis. Without his help, I can say there would not be this thesis. I would also like to give thanks to Dr. Louise Pilote for her valuable comments and for serving as my thesis committee member. Her interest, time and advice were very beneficial and greatly appreciated.

I wish to thank Ms. Gillian Bartlett, MSc. for her help in getting familiar with the Framingham data, for her patience with my questions. I want to thank Ms. Christine Beck for her help in proofreading my thesis.

I also wish to take this opportunity to thank my friend Mr. Jia Zheng, MSc. who helped me in some aspects of data management. Finally, I wish to thank my family whose support, encourage and consideration enabled me to progress through my graduate studies with confidence!

Table of Contents

Chapter 1

Introduction

In this chapter, I first review basic facts about risk factors for cardiovascular disease and, in particular, for coronary heart disease (CHD). I then comment on some methodological issues relevant for the assessment of risk factors' role as well as for the evaluation of preventive interventions, aimed at risk factors modification. However, statistical methodology directly relevant for analyses carried out in this thesis is reviewed in greater detail in chapter 2.

1.1 Overview of Cardiovascular Risk Factors

According to the third monitoring report of the World Health Organization, cardiovascular disease causes 12 million deaths in the world each year (World health statistical quarterly, 1993&1995). Cardiovascular diseases represent half of all deaths in several developed countries. They are one of the main causes of death in many developing countries, and the major cause of death in adults.

CHD seems to be the predominant type of cardiopathy encountered in many countries. Its direct and indirect costs to the society are enormous. For example, in the United States, these costs may exceed 100 billion US dollars per year (Fact Book, 1988). Based on data from the Atherosclerosis Risk in Communities (ARIC) study of the

National Heart, Lung, and Blood Institute (NHLBI), CHD caused 476,124 deaths in the United States in 1996, or 1 out of every 4.9 deaths. From 1986 to 1996, rates of death due to CHD declined by 27.0%, but the actual number of deaths declined only by 8.6% (American Heart Association statistical update). Accordingly, designing effective strategies for CHD prevention is among the highest priorities of public health authorities (Anderson et al, 1991).

The concept of risk factors constitutes a major advance in the development of strategies for CHD prevention (Scott et al, 1998). The Framingham Heart Study (Kannell et al, 1971) played a vital role in defining the contribution of risk factors to CHD occurrence in the general population of the United States. The major risk factors identified included cigarette smoking, hypertension, advancing age, diabetes mellitus, sex, family history of CHD, high serum cholesterol and various cholesterol fractions, such as low levels of high-density lipoprotein (HDL) cholesterol. Potential risk factors other than those listed as major risk factors which have been studied in the Framingham or other studies include obesity (Hubert et al, 1983), physical inactivity (Kannel et al, 1985), family history of premature CHD (Myers et al, 1990), hypertriglyceridemia (Austin, 1991), small low-density lipoprotein (LDL) particles (Lamarche et al, 1996), increased lipoprotein (a) (Lp[a]) (Dahlen, 1994), increased serum homocysteine and abnormalities in several coagulation factors (Kannel et al, 1987).

Identification of CHD risk factors is important not only for etiology, but also for prevention of CHD morbidity and mortality. The reason is that some of the major CHD

risk factors can be modified through deliberate interventions. In the next section, I present a brief overview of the common type of intervention aimed at CHD prevention.

1.2 Examples of Interventions Aimed at Risk Factor Modification

Some risk factors are not modifiable; that is, they cannot be changed or altered. Risk factors such as family history of heart disease and age are examples of these non-modifiable risk factors. By contrast, serum cholesterol, blood pressure, smoking and obesity presumably are modifiable risk factors. Several population-based interventions and clinical practice guidelines have been suggested to reduce the incidence of CHD by changing the levels of these modifiable risk factors (Rosenberg et al, 1985; Shea et al, 1985; 4S, 1994; Shepherd et al, 1995). The assessment of the effectiveness and cost-effectiveness of such interventions, in terms of changes in life expectancy or prolongation of CHD-free survival has gained considerable importance as the restricted health care budgets in many countries and has required optimal allocation of limited resources (Weinstein et al, 1985;Tsevat et al, 1991; Grover et al, 1994; Hamilton et al, 1995).

Prevention and treatment of CHD are listed as priorities in the Health People 2000 objectives. Education for prevention starts early in life. In particular, taking action to control blood pressure, reducing dietary fat intake, as well as reducing cigarette smoking and increasing moderate physical activity are each preventative actions against CHD (IPLAN Community Health Committee). Efforts to modify risk factors or prevent their development with the aim of delaying or preventing onset of new CHD in originally

asymptomatic individuals are classified as the primary prevention interventions (Scott, 1998). Various types of primary prevention interventions have been already used to reduce CHD risks. These interventions include:

Medication According to the National Cholesterol Education Program (NCEP) guidelines, cholesterol-lowering drugs can be considered for middle-aged men with LDL-cholesterol levels >190 mg/dL or >160 mg/dL in the presence of two or more CHD risk factors (NCEP reports, 1993, 1994). The recent West of Scotland Coronary Prevention Study (WOSCOPS) confirms that cholesterol-lowering drugs will safely and effectively reduce CHD rates in middle-aged men at high risk (Shepherd et al, 1995). Several primary and secondary prevention clinical trials also have shown that people using Clofibrate and Niacin can reduce CHD risks (CDPRG, 1975; Carlson et al, 1977; COPI, 1978).

Smoking Cessation Cigarette smoking acts synergistically with hypertension and hyperlipidemia to increase markedly the risk of CHD. It is estimated that 29% of all deaths from coronary heart disease are attributable to smoking (Cardiac Prevention and Rehabilitation Research Center, CPRRC). The Framingham data further reveal that smoking is a powerful risk factor for myocardial infarction. In fact, smoking is an even stronger risk factor for myocardial infarction than for angina pectoris (Hubert et al, 1982). The mechanism of this effect is that smoking accelerates coronary plaque development (Strong, 1976). Compared with nonsmokers, smokers have a 70% increased risk of fatal CHD and a two- to four-fold higher risk of nonfatal CHD and sudden death

(CPRRC). Regardless of age, quitting smoking will decrease one's chances of developing heart disease. Smoking cessation rapidly and markedly reduces risk for myocardial infarction, Rosenberg et al. (1985) found that exsmokers (those who had last smoked at least one year previously) eventually have an incidence of myocardial infarction similar to that for people who have never smoked. It is estimated that the increment in risk due to cigarette smoking can be erased in 2 to 3 years by smoking cessation (Freund et al, 1992; Kawachi et al, 1994). The Multiple Risk Factor Intervention Trial – a major study in the U.K. showed that men at high risk for myocardial infarction significantly reduced their chances of dying from CHD if they stopped smoking.

Dietary intervention Since 1957, the American Heart Association (AHA) proposed that modification of dietary fat intake would reduce the incidence of coronary heart disease. The AHA also has issued several policy statements, such as eating a nutritionally balanced diet consisting of a variety of foods, reducing consumption of fat, especially saturated fat and cholesterol, increasing consumption of complex carbohydrates and dietary fiber. Each of these recommendations is based on evidence that modification of specific risk factors associated with CHD will reduce this risk and improve the quality of life of those with the disease.

Others interventions include nondrug therapy weight reduction and increased physical activity.

To evaluate the public health benefits of preventive interventions, aimed at risk factor modification, it is not sufficient to identify a given variable as a risk factor. Indeed, one has to estimate the expected effect of, for example, lowering serum cholesterol level by 20 mg/dL on the risk of CHD mortality. Obviously, such estimates have to be based on epidemiologic data. In section 1.3, I briefly review the main types of epidemiologic studies used for this purpose.

1.3 Methodological Considerations: Between-subjects Designs as a Basis to Evaluate the Effect of Within-subject Changes

Usually, the effectiveness of interventions aiming at lowering risk factor values is assessed based on one of two types of epidemiological studies: **randomized clinical trials** and **large cohort studies**. Both types of studies require some inference about the hypothetical effect of within-individual longitudinal changes in risk factor values based on some type of between-subjects comparisons.

Randomized Clinical Trials

Randomized clinical trials allow the comparison of outcomes among control participants, who are not targeted by any specific intervention, and outcomes among the participants of the active intervention arm. These trials hold the implicit assumption that the difference in observed risks is due to the ability of the intervention to modify some risk factors. Randomization is expected to balance the distribution of relevant risk factors in the two groups of participants, therefore minimizing the risk of confounding.

The Scandinavian Simvastatin Survival Study (4S) is an example of a well designed and influential randomized clinical trial of secondary prevention of CHD mortality (4S group, 1994). Simvastatin is an inhibitor of hydroxy-methylglutaryl coenzyme A (HMG-CoA) reductase, which reduces LDL cholesterol to a greater extent than that achieved in previous diet and in other drugs intervention trials (Todd et al, 1990, Illingworth et al, 1994). The trial was conceived in April 1987, with the main objective of testing the hypothesis that lowering cholesterol with simvastatin would improve survival of patients with CHD.

In 4S, 4444 patients with angina pectoris or previous myocardial infarction and serum cholesterol 5.5-8.0 mmol/L (213 to 310 mg/dL) with triglycerides ≤2.5 mmol/L (220 mg/dL) on a lipid-lowering diet were randomized to double-blind treatment with 20 to 40 mg of simvastatin or placebo once per day. The groups were well matched at baseline. The investigators used the log-rank test and Cox model to find that a mean reduction of serum cholesterol of 1.8 mmol/L (25%) was achieved over the median follow-up period of 5.4 years. In addition, they found that the risk of coronary death was reduced by 37% over the whole study in simvastatin group. The authors concluded that long-term treatment with simvastatin is efficacious and improves survival in CHD patients.

One recent publication by Pederson et al. (1998) also confirmed same conclusion of the long-term treatment with simvastatin. In a study of lipoprotein changes and reduction in the incidence of major CHD events in the 4S trial, these authors used the Cox

proportional hazards model to assess the relationship between lipid values at baseline, and lipid values after 1 year (percent change from baseline at year 1) and major coronary events (MCEs). The results from this study showed that for each additional percentage point reduction in total cholesterol, the MCE risk was reduced by 1.9% (95% CI, 1.0% to 2.4%; p=0.00005). Changes in LDL and HDL cholesterol both contributed to the reduction in risk. The authors suggested that the beneficial effect of simvastatin in individual patients in the 4S trial were determined mainly by the magnitude of the change in LDL cholesterol. Their conclusions are consistent with recommendations in current guidelines, which emphasize the aggressive reduction of LDL in CHD patients.

The West of Scotland Coronary Prevention Study (WOSCOPS) was a randomized placebo controlled clinical trial for the investigation of primary prevention of CHD (1992). It was designed to test the hypothesis that a nightly dose of 40 mg of pravastatin would reduce the combined incidence of coronary morbidity and mortality in 45-64 year old men, who had elevated plasma cholesterol levels. Over an average period of 5 years of follow-up study, the authors found that pravastatin lowered plasma cholesterol levels and low-density lipoprotein cholesterol levels, whereas there was no change with placebo. As a result, a 22% reduction in the risk of death from any cause in the pravastatin group (95% CI, 0 to 40%; p=0.051) was observed. The authors concluded that treatment with pravastatin significantly reduced the incidence of myocardial infarction and death from cardiovascular causes (Shepherd et al, 1995).

Other randomized clinical trials included the World Health Organizations clofibrate trial (1978, 1980), the Lipid Research Clinical Coronary Primary Prevention Trial (LRC-CPPT) (1979, 1984), the Helsinki Heart Study (HHS) (Mänttäri, et al, 1987, Fric et al, 1987).

The results from each of these studies were consistent with epidemiological data (Stamler et al, 1986, Chen et al, 1991, Burchfiel et al, 1995), other intervention studies (LRCP, 1984, Manninen et al, 1988) and meta-analyses (Law et al, 1994, Gordon et al, 1995, Gould et al, 1998), in that greater reduction in serum cholesterol, and especially in LDL cholesterol, reduced the risk of CHD.

Large Cohort Studies

The cohort study is another epidemiological study design, used to estimate the effects of CHD risk factors. The Framingham Heart Study (Margolis et al, 1974), the Honolulu Heart Program (Trombold et al, 1966), and the Lipid Research Clinics – Follow-up study (Morris et al, 1994) are examples of influential CHD cohort studies. The results of such studies are analyzed by comparing outcomes of individuals with different baseline (initial) risk factor vectors, with the expectation that these between-subject comparisons can be generalized to predict the effects of longitudinal within-subject changes.

The Framingham Heart Study is one of the most important epidemiological studies in medical history. In this study, the investigators have collected data prospectively every two years since 1948 on a cohort of 5209 subjects, aged 30 to 62 years at the beginning of the study, to examine the relationship of potential risk factors to the development of cardiovascular disease (Dawber, 1980; D'Agostino et al, 1989). In this study, repeated measures of risk factors have been employed to evaluate their relationship to disease development. The outcomes examined in this study included CHD death, cancer death, non-fatal CHD events and strokes. Through this long-term cohort study, a number of potential cardiovascular risk factors have been identified, such as high cholesterol, high blood pressure and obesity.

Another ongoing, prospective, population-based cohort study of CHD and stroke is the Honolulu Heart Program. This study began in 1965 among men of Japanese ancestry born between the years 1900 and 1919 and living on the island of Oahu, Hawaii (Yano et al, 1984). A total of 8,006 men participated in the initial examination carried out from 1965 to 1968. A second examination was carried out 2 years after the first. A third comprehensive examination was carried out 6 years after the first, during the years 1971 through 1974, and 6,860 of the original cohort of 8,006 men participated. The fourth comprehensive examination of elderly survivors of the original cohort began in 1991 and continued through 1993. Several CHD risk factors were studied during this period. For example, Kagan et al. examined the relationship between baseline serum cholesterol levels and subsequent 9-year mortality. Those investigators found that the baseline serum cholesterol level was positively related to CHD mortality. The relation of baseline serum cholesterol level to total mortality in this cohort was quadratic; that is, there was an

excess of CHD death associated with serum cholesterol level at the high end of the distribution (mainly due to CHD) and at the low end (mainly due to cancer) (Kagan et al, 1981).

Reed et al also studied the relationship between lipids and lipoproteins and CHD, stroke and cancer in the Honolulu Heart Program. In their study, a group of 2,122 healthy men in the Honolulu Heart Program, who participated in the Cooperative Lipoprotein Phenotyping Study, from 1970 to 1972, were followed for 10 years. For this study, repeated examinations and surveillance of hospital discharge and mortality records were made in order to diagnose new cases of coronary heart disease, stroke, cancer, and other deaths. Through multivariate analyses, the authors found that total cholesterol and low-density lipoprotein cholesterol levels were significantly associated with all clinical types of CHD. High-density lipoprotein cholesterol levels were inversely associated with nonfatal myocardial infarction and total CHD, but not with fatal CHD or angina. In univariate analyses, triglyceride and very-low-density lipoprotein cholesterol levels were associated with total CHD, but this relationship was not observed in multivariate analyses. Thus, for total disease (CHD, stroke, cancer, and other deaths), the optimal range for lowest total disease incidence was about 200 to 220 mg/dl for total cholesterol, and 120 to 140 mg/dl for low-density lipoprotein cholesterol. A strong inverse relationship between total disease incidence and high-density lipoprotein cholesterol levels indicated that the highest cholesterol levels were the optimal levels (Reed et al, 1986).

The Honolulu Heart Program and most other CHD studies focused on the correlation of baseline risk factors and CHD. However, the values of some risk factors, such as cholesterol and diastolic or systolic blood pressure, can change with the time. The information about these variables at baseline may be less relevant to assess their average risk level over the whole follow-up time, and so a model using only baseline measurements may overestimate the risk for individuals with high initial risk levels (Cupples, 1988). In addition, it is important to ensure that the design of such a study permits the separation of effects of the baseline risk factor values from the effects of their changes during the follow-up (Abrahamowicz et al, 1997).

To assess the impact of updated, rather than initial value of serum cholesterol on CHD risk, it is necessary to measure this variable repeatedly, and employed adequate statistical methods enabling us to account for such repeated measurements. In the next section, I discuss the role of repeated measurements of serum cholesterol and summarize some published methods to handle such measurements.

1.4 The Role of Repeated Measures of Serum Cholesterol in the Epidemiology of Coronary Heart Disease

To effectively control blood cholesterol levels, the most important question is "do we already know enough about this risk factor? ". Obviously, we do not. Even if we could identify all of the CHD risk factors, we would need precise and accurate data on their levels at all times in the life spans of all individuals (Nieto, 1999). While continuous monitoring of risk factor values is impractical, some prospective CHD studies included

repeated measurements of major risk factors. For example, the Framingham Heart Study has been collected data prospectively at two years intervals for over 35 years since 1948 to examine the relationship of potential risk factors to the development of cardiovascular disease (Dawber et al, 1980). One challenge in medical and statistical research is how to best use the information contained in repeated measures for evaluating the role of a given risk factor in disease development (Cupples et al, 1988).

For the Framingham heart study, two different approaches were proposed to examine the relationship between potential risk factors and the development of CHD. One method was to measure the risk factors at a single moment in time and to follow individuals in time to observe the incidence of CHD or mortality (Cupples et al, 1988). For example, we could utilize the information on risk factors that was measured at the first exam and the development of disease over the following 30 years. That is, we could ignore the repeated measures and use only the baseline measures to evaluate the association. Since risk factor values in individual subjects change over time, potentially important information is lost if we take this approach (Cupples et al, 1988). As we know, the greatest overall change occurred in the men who exhibited the highest levels for CHD risk factors at the baseline examination. These men, who were at high risk of morbidity because of elevated risk factor values earlier in life, gradually modified those values through medical intervention or life-style changes into a greater alignment with values that associated with lower risk (Benfante et al, 1994). Indeed, it is the expectation of a changing risk profile that motivates researchers to record potential risk variables repeatedly during the follow-up period, particularly in long-term cohort projects such as

the Framingham study. A common hypothesis is that one's current risk profile may be more predictive of outcome than one's baseline measurements (Cupples et al, 1988).

The second approach utilizes the information obtained through repeated measures. One method to accomplish this is known as the pooling of repeated observations (PRO) method. The PRO method was originally proposed by Wu and Ware (1979) and further discussed by Cupples (1988) and D'Agostino (1990). For this approach, each 2-year examination interval is treated as an independent follow-up study. Observations over all intervals are pooled into a single sample, and logistic regression analysis is used to examine the association between the risk factors and development of disease during all follow-up periods (Hu et al, 1999). D'Agostino et al. showed the asymptotic equivalence of this approach to the Cox regression model with time-dependent covariates. A general pooled logistic model can be written as:

$$\text{Logit } p_i\,(X\,(t_{i-1})) = \alpha_i + \beta_1 x_1\,(t_{i-1}) + \ldots + \beta_p x_p\,(t_{i-1}) \qquad \textbf{[1.1]}$$

Where $p_i\,(X\,(t_{i-1}))$ is the conditional probability of observing an event by time *ti*, given that the individual is event-free at time t_{i-1}, and $X\,(t_{i-1})$ is the vector of p independent variables measured at time t_{i-1}.

This method is a generalized person-years approach and has been a conventional method of analysis employed in the Framingham Study cohort (Kahn et al, 1966; Shurtleff et al, 1974; Schatzkin et al, 1984). Here all time-dependent measurements recorded at repeated intervals are considered in the evaluation of the relationship between

risk factors and outcome. It treats each observation interval (of equal length) as a mini-follow-up study, in which the current risk factor measurements are employed to predict an event of interest in the interval among persons free of the event at the beginning of the interval. Observations over multiple intervals are pooled into a single sample to predict the short-term risk of an event (Cupples et al, 1988).

The two approaches (one time versus repeated measures) provide different insights into the etiology of disease and should not be compared in a statistical manner. Some risk factors may carry long-term significance and be less important in the short term. For example, in the examination of the risk of MI, age at the first exam in 1950 appeared to be a more important predictor for CHD than current age (Cupples et al, 1988). On the other hand, some risk factors may only be important in the short term, as they may indicate the beginning of the development of disease, and they may have little prognostic value for longer periods.

The above discussion points out to the importance of taking into account the timing of both risk factor measurements and of events of interest in the epidemiologic studies of CHD. This is important for both the continuous risk factors discussed in this section, such as total serum cholesterol, and binary risk factors, such as smoking. The role of smoking is briefly described in "Smoking Cessation", the sub-section of section 1.2. In chapter 2, I describe, in some detail, statistical methodology that incorporates information on the timing of events of interest and allows the model to appropriately changes over time in the values and/or in the effects of the risk factor.

Chapter 2

Proportional Hazards Model and its Applications in Coronary Heart Disease Studies

This chapter continues a review of the conventional Cox Proportional Hazard model and its various generalizations and thus, provides the methodological background for statistical modeling carried out in this thesis. When reviewing the relevant statistical methods, I have also discussed some of its application in epidemiology, with particular emphasis on the studies of CHD risk factors.

Current perception of the effects of coronary heart disease (CHD) risk factors is largely based on the results of statistical analyses from large prospective CHD studies, such as the LRC program (Jacobs et al, 1990), the Framingham heart study (Dawber, 1980), MRFIT (1979) and the Honolulu Heart Program (Yano et al, 1984).

The early statistical analyses of these data relied on univariate methods, and their results may have been affected by confounding. More recent analyses typically use multiple regression methods. These multivariable analyses rely almost exclusively on two parametric models: multiple logistic regression (Cupples et al, 1987 and Efron, 1988) and Cox Proportional Hazards (Cox, 1972). Of the two models, the Cox Proportional Hazards

is used most often. It is designed specifically to analyze censored survival data by taking into account the differences in time-to-event and by accounting for losses to follow-up.

One important difference between logistic regression and the Cox model is that the former considers only who had and outcome of interest but ignores when the outcomes of individual subjects occur during the follow-up. By contrast, in the estimation of the Cox model, the ranking of times at which individuals has outcome plays an important role (Kalbfleisch and Prentice, 1980). By taking timing of events explicitly into account, the Cox model allows the analyst to discriminate between two types of covariates. "Fixed" covariates are those variables that remain constant during the entire follow-up period, while "time-dependent" covariates change their values over time. In sections 2.1 and 2.4, I discuss fixed and time-dependent covariates in more detail.

2.1 Cox Proportional Hazard Model with Fixed Covariates

In 1972, Cox introduced a statistical methodology based on longitudinal follow-up data that popularized the use of regression analysis methods in modeling the relationship between predictive covariates and the hazard function for the occurrence of an event. One of the most commonly quoted features of the hazard regression model proposed by Cox, which subsequently will be referred to as the Cox model, is the assumption that the hazard rates remain proportional over time. For this reason, the Cox model is often referred to as the Proportional Hazards model (Kalbfleisch and Prentice, 1980).

The Cox model is widely used in the analysis of survival data to estimate the effects of explanatory variables on survival times and to test statistical significance of

these effects. The survival time for each member of a population is assumed to follow its own hazard function $\lambda(t)$ (Kalbfleisch and Prentice, 1980). The hazard function specifies the instantaneous rate of failure at $T = t$, conditional upon survival to time t and is defined as:

$$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{P(t \le T < t+\Delta t)}{\Delta t} \qquad [2.1]$$

$$= f(t) / S(t)$$

In (2.1), f (t) is a probability density function. S(t) is the survival function, which indicates the probability that T is at least as great as a value t (here let T be a non-negative random variable representing the failure time of an individual from a homogeneous population). Then we have:

$$S(t) = P(T \ge t) = 1 - \int_0^t f(u)\, du \qquad [2.2]$$

Let $\lambda(t; z)$ represent the hazard function at time t for an individual with the covariate vector $\underline{z}$. The proportional hazards model (Cox, 1972) specifies that

$$\lambda(t; z) = \lambda_0(t) \exp(\underline{z}\,\beta) \qquad [2.3]$$

where $\lambda_0(t)$ represents the arbitrary, unspecified baseline hazard function corresponding to an individual with all covariate values equal to zero (z=0). $Z = (z_1, z_2, \ldots, z_k)'$ is the

vector of a measured explanatory variable, and $\beta = (\beta_1, \beta_2, \ldots \beta_k)'$ is the vector of logarithms of the hazard ratios, and is associated with a unit increase in subsequent explanatory variables. Thus, in the proportional hazards model, the effects of covariate Zj on survival is expressed by the hazard ratio $exp\ (\beta_j)$, j =1,2...k.

The Proportional Hazards model relies on the essential assumption that the ratio of the hazard functions for two individuals with different sets of covariate values does not depend upon time. That is, the hazard ratio is constant over time (Kalbfleisch and Prentice, 1980). This assumption is also referred to as the Proportional Hazards (PH) assumption. One advantage of the Cox implementation of the PH model is that whereas the ratio of hazard function is assumed to be constant, there are no restrictive assumptions on the shape of the actual hazard functions (Cox, 1972). In other words, whereas relative risks are restricted to be constant, the absolute risk of outcome may vary over time according to an arbitrary function. This feature of the Cox model makes it much easier to fit the actual survival data in various epidemiologic studies as it avoids additional restrictive assumptions required in classic, fully parametric, regression models for survival data such as Weibull or lognormal models (Kalbfleisch and Prentice, 1980). As a result, the Cox model has quickly become an increasing popular method for analyzing survival data in various field of medical research, including the epidemiology of CHD.

2.2Applications of Conventional Cox Model in Epidemiological Studies of Coronary Heart Disease Risk Factors

In the 1980's, the Cox Proportional Hazard model, which was cited in more than 500 clinical papers per year (Hanley, 1989), became the most popular method for studying the effects of risk factors on survival time. A review of survival analyses for a cancer journal by Altman et al (1995) indicated that only 1 among 46 papers using multivariate survival analyses did not use the Cox model, and relied on logistic regression. Accordingly, the Cox model is probably the most frequently applied method in multivariable survival analyses of CHD risks. The general tendency to restrict Cox model analyses to fixed-in-time covariates is also evident in the studies of CHD risk factors.

For example, Posner et al. (1991) used the Cox model to assess the relationship between dietary variables and the development of CHD in the Framingham population. The bivariate and multivariable regression models were estimated to determine if diet has independent associations with CHD mortality after adjusting for conventional cardiovascular risk factors, including serum total cholesterol level, systolic blood pressure, glucose intolerance and cigarette smoking. However, using the maximum partial likelihood estimates, there were no observed significant relationships between the dietary variables and CHD incidence among the older Framingham study male participants. The authors commented that one explanation might be related to the changes that may have occurred in dietary intake during the 16 years of follow-up. Whereas information on dietary intake was limited to initial measurement, it is possible that undetected changes in

the individuals ' nutrient intake may have affected the outcomes. This possibility is a rather general concern, for it is likely that risk factor values will change during follow-up in long-term epidemiological prospective studies of CHD.

Benfante et al (1994) reported convincing evidence of such changes occurring among the male participants of the Honolulu Heart Program during 25 years of follow-up time. In this study, risk factors, that included systolic blood pressure (SBP), diastolic blood pressure (DBP), serum cholesterol, cigarette smoking, body mass index and alcohol intake, were observed 4 times during 25 years. Comparing the distributions at the initial and the fourth examination, it was found that 65% of the men had moved into a different quartile, with 25% changing by more than one quartile for systolic blood pressure (SBP). For diastolic blood pressure (DBP), 68% had moved into another quartile, with 28% moving more than one quartile. For body mass index, 53% had moved into another quartile, with 14% moving more than a quartile. Accordingly, the correlation between initial risk factor values and their values taken 25 years later for the different variables was rather weak from the Spearman rank correlation coefficients: 0.280 for diastolic blood pressure, 0.356 for systolic blood pressure, 0.426 for cholesterol, 0.355 for smoking, and 0.642 for body mass index.

These results showed that during the 25 years of follow-up, there was a substantial redistribution of cardiovascular disease risk factor values, indicating that the baseline value became less and less representative of the current risk factor levels (Benfante et al, 1994). On the other hand, it may be expected that the relative risks at a given time during

the follow-up are mostly determined by the *current* risk factor values (Cupples et al, 1988). If so, then the ability of a baseline risk factor value to predict up-dated relative risks will deteriorate with an increasing follow-up time. By contrast, a different mechanism was postulated by Benfante et al (1994), who argued that among older men the earlier values of risk factors might be more predictive of long-term cardiovascular risk than current values. The reason for such a hypothesis would be that high values of risk factors in mid-life would likely induce atherosclerotic change, the impact of which would be unlikely to be reversed by later changes in risk factors. However, the mechanism postulated by Benfante et al (1994) has not been yet verified by adequate analyses of epidemiological data on changes in CHD risk factors.

Indeed, relatively few authors have attempted to address the issue of the impact of changes in CHD risk factors over time. In the next section, I present a brief review of such studies.

2.3 Review of Published Analyses Related to Temporal Changes over time in Coronary Heart Disease Risk Factors

Recently, a number of authors have addressed the issue of changes in CHD risk factors over time, and of the impact these changes may have on the risks of CHD mortality and/or morbidity. For example, Benfante et al (1994), Abbott et al, (1997) and Pekkanen et al, (1994) investigated the association between the direction of changes in total serum cholesterol over time and the CHD risk among elderly participants of the

Honolulu Heart Program and the Finnish cohorts of the Seven Countries study, respectively.

In the Honolulu Heart Program study, Benfante et al. (1994) used three approaches to assess the extent of change and redistribution of risk factor levels between middle and late life. Their results showed that the mean values for cardiovascular disease risk factor levels at both the initial and final examinations for the entire cohort had changed remarkably. The differences in mean values between the initial values and the values measured 25 years later were all highly significant (p-value < 0.001) (Benfante, 1994).

Another study of the Honolulu Heart Program by Abbott et al. compared the updated levels of total cholesterol and high-density-lipoprotein cholesterol (HDL-C) in a group of elderly men to those that were observed 20 years earlier. This study presented data for 971 men who participated in a separate fasting study of lipids and lipoproteins from 1970-1972, and those who received repeat examinations 10 and 20 years later. The men were aged 71-93 years at the time of the last examination. Over the 20-year period, total cholesterol declined by 1.6-1.8 mg/dL per year ($P < 0.001$), from average baseline values of 219-222 mg/dL. The mean reductions in total cholesterol in the second 10 years of follow-up (24 mg/dl) were more than double the reductions observed in the first 10 years (9 mg/dl). Levels of HDL-C rose 0.2-0.3 mg/dL per year ($P < 0.001$), from average baseline values of 44-46 mg/dL. Levels of total cholesterol declined and levels of HDL cholesterol increased regardless of initial levels of systolic blood pressure (SBP), body mass index, physical activity, cigarette smoking status, or the use of treatment for

hypertension or elevated total cholesterol. After adjustment for baseline cholesterol levels, men with CHD at the end of the 20-year follow-up (32 mg/dl) experienced significantly greater reductions in total cholesterol levels than men without this disease (22 mg/dl) ($P < 0.001$). Men who developed CHD within the first 10 years of follow-up had the greatest yearly decline in total cholesterol (1.9 mg/dL), as compared to men who developed heart disease later (1.8 mg/dL), and men who remained disease free (1.5 mg/dL). Differences between men with recent and earlier disease were not statistically significant, although men without coronary disease experienced a significantly smaller decrease in total cholesterol than either of these groups ($P < 0.05$). The authors concluded that changes in total cholesterol and HDL-C levels with advancing age might be part of a natural aging process. Some changes, however, such as large reductions in total cholesterol, may signal occult disease or declines in overall health.

These findings may be partly explained also by the selection mechanism through which healthier individuals have higher probability of survival since improvements in lipid and lipoprotein levels that are beneficial at younger ages were common in this long-lived cohort of men.

The cohort of the Finish Seven-Country study has been followed for up to 30-years. The association of past changes in serum cholesterol level with cause-specific mortality between 1974 and 1989 was examined in a cohort of 784 men aged 55-74 years who were free of symptomatic coronary heart disease in 1974. Changes in serum cholesterol level were computed based on measurements made in 1959, 1964, 1969, and

1974. Of the 405 deaths, 202 were due to cardiovascular diseases and 107 were due to cancer. Men who experienced declines in their serum cholesterol levels had higher serum cholesterol levels at the beginning of the follow-up in 1959/1964, but they had lower levels in 1969/1974. This result is to be expected, because, due to the mathematical properties of change, change is negatively associated (r=-0.707) with the baseline and follow-up measurement even if the baseline and follow-up measurement were completely uncorrelated (Oldham, 1962). Men with a decline in serum cholesterol level between 1959 and 1974 also experienced greater than average declines in body mass index and were more likely to be current smokers in 1974. The authors found that among 339 men aged 65-74 years in 1974, men in the lowest tertile of serum cholesterol change, i.e., with greatest declines, had increased cardiovascular (hazard ratio of 1.58; 95% confidence interval 1.00-2.50) and all-cause (hazard ratio of 1.46; 95% confidence interval 1.06-2.02) mortality compared with men in the middle tertile of change (Pekkanen et al, 1994).

In accordance with these findings, a decrease in the serum cholesterol level was strongly associated with increased cardiovascular mortality in the Framingham Study (Anderson et al, 1987). In their study, falling cholesterol levels and their association with subsequent survival was studied. For each individual with five or more cholesterol measurements at the first eight visits, a least-squares line was fit to estimate the change in serum cholesterol values per year during this period. Analyses were stratified by five-year age groups and performed separately for men and women. Two variables were created from the slope measurement to test if positive or negative changes in cholesterol values (vs no change) were associated with subsequent elevated mortality. The first

variable was defined as the value of the individual's slope if positive, and 0 otherwise. The other was defined as the absolute value of the slope if negative and 0 otherwise.

The results showed that the falling cholesterol levels – the negative slope variable – were associated with elevated overall mortality and CVD mortality in both men and women. The analysis combined men and women free of CVD and cancer, and was stratified by sex and age groups. Based on the model with negative slope only, it was estimated that for each 1 mg/dL per year drop in serum cholesterol values over the 14-year period of cholesterol measurement, there is an 11% increase in both the overall death rate ($p<0.01$) and the CVD death rate ($p<0.01$) during the following 18 years. That is, a person whose cholesterol levels dropped 14 mg/dL during the initial 14 years would be expected to have an 11% higher death rate during the subsequent 18 years than a person whose cholesterol levels remained constant or rose during the same period.

When interpreting these findings, the authors suggested two possible explanations (Anderson et al, 1987). One explanation is that a spontaneously occurring decrease in serum cholesterol would by itself cause increased coronary disease mortality among the elderly. This explanation might imply that cholesterol levels are falling mostly due to diseases predisposing to death. The second explanation is that decrease in serum cholesterol level and the associated high disease risk may be both caused by one or several other factors. Since the changes in serum cholesterol in the Framingham study are probably mostly spontaneous and less a result of intentional changes in diet, other living habits, or drug treatment, the authors suggest that both the decline in serum cholesterol

level and the associated high mortality may be caused by a third factor, such as increased prevalence of chronic diseases or other changes associated with aging. This factor would help to explain why several studies have not found an association between serum cholesterol and coronary risk among the elderly (Anderson et al, 1987).

However, it should be noted that Anderson et al. (1987) did not associate updated cholesterol level and its recent changes with simultaneous occurrence of CHD death. In their study, change in cholesterol was measured in the initial 14 years while outcomes were ascertained in the following 18 years. Accordingly, there is a possibility that the estimated associations are partly confounded by more recent changes in cholesterol, occurring during the 18-year follow-up period. To access this possibility and to estimate the effects of updated changes in cholesterol during follow-up, one can not, however, limit the analysis to the conventional fixed-in-time covariates. Indeed, the Cox model (1975) allows incorporating information on changes in risk factors during follow-up by including time-dependent covariates. Therefore, the next section contains a review of methodological issues pertinent to the use of time-dependent covariates.

2.4 Time-dependent Covariates in the Cox Model

To facilitate estimation of the proportional hazards (PH) model, Cox introduced the partial maximum likelihood approach (Cox, 1975), which eliminates the need to specify the baseline hazard function $\lambda_0(t)$ and accounts for censoring of survival times. Partial likelihood approach also allows for incorporating information on the changes in

the risk factor values during the follow-up time, which are represented by time-dependent variables (Cox, 1972).

A time-dependent variable is one whose value for any given individual can change over time. Accordingly, the value of a risk factor becomes a function of time rather than a single constant (Kalbfleisch et al, 1979). Let $Z_i(t)$ denote the covariate vector at time t for the ith individual under study. It is also convenient to introduce $Z_i(t)$ to denote the covariate path up to time t, $\{Z_i(u); 0<u<t\}$, and Z_i to denote the whole covariate process to the end of the study. The data for the ith individual are $(t_i, \delta_i, Z_i(t_i))$, $i=1, \ldots, n$. Time-dependent covariates fall into two broad categories: external covariates and internal covariates.

An *external* covariate is one that is not directly involved with the failure mechanism (Kalbfleisch et al, 1979). This includes covariates that may change over time but in such a way that these changes are either determined a priori or not related to any changes in variables measured directly on individuals under study. There are three types of external covariates, one type is the fixed covariate, whose value is measured in advance and fixed for the duration of study. An example of this type of covariate is a dose of radiation to which a given individual was exposed before the start of follow-up. The second type of external covariate corresponds to the situation when the total path of changes in covariate: Z_i, although not constant, is determined in advance for each individual in the study. Age, for example, could be determined for an individual in a trial of long duration. The third type of external covariate is the ancillary covariate. An

ancillary covariate is the output of a stochastic process, which is external to the individual under study. It has the property that the marginal probability distribution of Zi, i=1,..., n does not involve the parameters of the failure time model. An example of this type of covariate is one that measures daily airborne pollution as a predictor for the frequency of asthma attacks. Ancillary covariates play the role of ancillary statistics for the failure time model, and the conditioning principal would suggest conditioning on their whole-observed path Z. Ancillary covariates are characterized by the condition

$$P[Z(t + dt) \mid H(t)] = P[Z(t + dt) \mid Z(t)] \qquad \textbf{[2.4]}$$

This condition is a formalization of the idea that the path of the covariate process may influence, but is not influenced by, the failure experience of the trial.

For external covariates, the hazard function is defined as

$$\lambda(t ; Z, \theta)dt = P\{T \in [t, t+ dt) \mid Z, \theta, T \geq t\} \qquad \textbf{[2.5]}$$

An *internal* covariate is represented by repeated measurements taken on the individual over a period of time. Thus, in contrast to external covariates, an internal time-dependent covariate represents changes over time in some characteristics of individual study subjects. Moreover, the patterns of changes in internal covariates are different for different subjects, i.e. cannot be determined a priori. For example, age is an external time-dependent covariate because the pattern of its changes is the same for all subjects. By contrast, total serum cholesterol will change in a different way for different individuals. Therefore, repeated measures of cholesterol will be represented by an internal time-

dependent covariate. The internal covariate has the following property that it requires the survival of the individual until time t for its value Z(t) to be defined, and thus, it carries information on the failure time. A simple example of this covariate could arise in an immunotherapy trial for cancer studies. In such a trial, it may be of interest to examine the effect of immunotherapy on the failure rate given a current measure of immune status such as white blood count. In this case, the covariate Z (t) may be taken to specify white blood count at time t. Models could also be constructed to evaluate treatment effects while adjusting for the current white blood count, or to allow treatment effects to depend on the current white blood count.

For internal covariates, the hazard function is defined by

$$\lambda(t\,;\,Z(t),\,\theta)dt = P\{T\in[t,\,t+dt)\,|\,Z(t),\,\theta,\,T\geq t\} \qquad [2.6]$$

We can see that in (2.6) we used Z(t) for internal covariates, while we used Z in (2.5) for external covariates. An internal covariate is the output of a stochastic process that is generated by the individual under study, and so it is only observed for as long as the individual survives and is uncensored. In consequence, its observed value carries information about the survival time of the corresponding individual. Z determines the survival information for the corresponding individual in the internal covariate, and not their whole-observed path Z in the external covariate. For example, a patient typically moves from one state to another over time and the hazard at time t depends markedly on Z(t).

We consider a vector $z(t) = (y(t),x(t))$ of time-dependent covariates where $y(t)$ is the vector of all fixed, defined or ancillary covariates and $x(t)$ is the vector of all internal covariates. Let $Y= \{y(u), u>0\}$ denote the complete covariate function over the whole study period for the external covariates and $X(t) = \{x(u), 0<u<t\}$ denote the covariate process up to time t for the internal covariates. Let $Z(t) = (Y, X(t))$. In the special and unusual case where no internal covariates are present, $z(t) = y(t)$ and $Z(t)$ specify the full covariate path. The hazard function is defined by

$$\lambda(t ; Z(t))\, dt = P\{T \in [t, t + dt) | Z(t), T \geq t\} \qquad [2.7]$$

As a special case of interest, we suppose

$$\lambda(t ; Z(t)) = \lambda_0(t) \exp(z(t)\beta) \qquad [2.8]$$

in which the hazard at t depends only on the current value, z (t).

To better illustrate the additional insights obtained from time-dependent covariates, in the next section I briefly describe how they were employed in selected epidemiologic studies.

2.5 Selected Applications of Time-dependent Covariates in Epidemiological Studies

Time-dependent variables have many useful applications in survival analysis. For example, in a study of the impact of medication, we can use a time-dependent variable to

model the effect of subjects changing their dose of medication during the follow-up period (Tamblyn & Abrahamowicz, 1998). Or we can include time-dependent variables, such as current blood pressure or blood chemistry measures, which vary with time during the course of study.

One of best known examples of the use of such strategies in survival analysis is the Stanford heart transplant program (Crowley and Hu, 1977). The objective of this study was to assess the effects of various explanatory variables on the survival of patients. Patients were accepted if physicians judged them suitable for a heart transplant. Then, when a donor became available, physicians chose transplant recipients according to various medical criteria. Thus, a patient's transplant status was changed from waiting for a transplant to a transplant recipient during the study. Accordingly, transplant status for these patients should be defined by a time-dependent covariate function $Z(t)$, with:

$$Z(t) = \begin{cases} 0 & \text{if the patient has not received the transplant by time } t \\ 1 & \text{if the patient has received the transplant at or before time } t \end{cases}$$

In the Stanford heart transplant study, the authors had presented a number of analyses to assess the effects of various explanatory variables on the survival of patients, including transplant status Z (t) and the age at acceptance, which were fixed-in-time covariates. Using Wald chi-square statistic test, with an α=0.05 significance level, time-dependent transplant variable Z(t) appeared to be associated with a slight decrease in risk, and the risk increased significantly with increasing age at acceptance.

The time-dependent analytic technique was also employed in various studies to identify prognostic factors for established disease or risk factors for the development of disease. For example, Culp et al. (1996) represented repeated measures of laboratory values as time-dependent covariates to model mortality risk in hemodialysis patients. Lewis et al. (1997) represented updated measurement of stenosis by time-dependent covariate to verify whether current stenosis is a more powerful predictor of asymptomatic carotid disease than the initial stenosis. Grohn et al. (1997) employed a time-dependent covariate, which represented whether or not the patient was currently sick to determine the general effect of disease on culling. Tsoukas et al. (1998) treated splenectomy as a time-dependent covariate, in order to account for the variation in its timing in their study that assessed the effects of splenectomy on survival and time to development of acquired immunodeficiency syndrome among HIV-positive patients. Merkel et al. (1998) used updated values of the Child-Pugh score and the aminopyrine breath test as time-dependent covariates in the study of current prognosis of cirrhosis by the Cox model. In each of these studies, the value of a risk factor became a function of time rather than a single constant.

Despite their use in the examples presented above, the use of time-dependent covariates in epidemiologic studies has been rather infrequent. For example, Altman et al (1995) carried out a systematic review of survival analyses in the research papers published in five major clinical oncology journals between October and December 1991. Of the 43 papers that he found which used a multivariable Cox model, only 1 employed time-dependent covariates. However, application of time-dependent covariates could

explain several of the unexpected findings in this study, such as the violation of the PH assumption which only takes into account baseline risk factor values (represented by fixed-in-time covariates) (Abrahamowicz et al, 1996). Moreover, the failure to account for subsequent changes in risk factors' values in the long-term cohort studies may have resulted in a distortion of the expected effect of a baseline risk factor. For example, if subjects with a very high initial cholesterol value are most likely to lower this value during their follow-up, then the estimated relationship between the initial cholesterol level and CHD risks may become non-linear (Abrahamowicz et al, 1997). Specifically, the risk may increase gradually over the range of low and moderate value, but it will not increase with a further increase of very high initial cholesterol levels that, later on, will be reduced to more moderate level.

The above discussion suggests that the issue of changes in the risk factor value is related with that of possible changes in the effect of a given risk factor over time. However, handling of each of the two issues requires a different methodological approach. Changes in risk factor value can be analyzed using time-dependent covariates in the classic Cox model that imposes the proportional hazard assumption. By contrast, to represent a change in the effect of a risk factor over time, no needs to relax the PH assumption and allow for non-proportional hazards (Hess, 1994). In the next section I present a brief review of recent statistical models that incorporate non-proportional hazards i.e. allow for the change of the predictor's effect over time.

2.6 Flexible Generalizations of the Cox Model

During the last decade, several flexible generalizations of the conventional regression models have been proposed in the statistical literature. More recently, they have been demonstrated to yield new insights into complex epidemiological and clinical data, including recent non-parametric models for survival analysis.

Throughout the 1990's, several new flexible statistical models that generalize the conventional proportional hazard model by incorporating the time-varying hazard ratio have been proposed (Zucker and Karr, 1990; Gray, 1992; Hastie and Tibshirani, 1993; Hess, 1994; Kooperberg et al, 1995; Verweij and Houwelingen, 1995; Abrahamowicz et al, 1996), and used for analyzing the real-life clinical and epidemiological data (Esdaile et al, 1994; Côté et al 1995; Lewis et al, 1997; Rachet et al, 1998). The main objective of these models is to allow partial likelihood estimation while accounting for the violation of the PH assumption by one or more covariates.

These models rely on flexible non-parametric modeling techniques to represent hazard ratio as a relatively unconstrained function of time. For example, Verweij and Houwelingen (1995) smoothed local hazard ratios by introducing a penalty for the first difference, while Gray (1992) and Kooperberg, Stone, and Truong (1995) used low-order regression splines (step functions and breaking lines). Abrahamowicz, MacKenzie and Esdaile (1996) employed more flexible quadratic splines with a low number of knots. Hess (1994) worked with a natural cubic spline. Zucker and Karr (1990) and Hastie and Tibshirani (1993) relied on smoothing splines.

The advantage of the regression spline model developed by Abrahamowicz et al (1996) is that it retains modeling flexibility with accurate statistical inference. The method was evaluated in simulations that showed that it can prevent inflation of type I error rate, when testing the PH hypothesis, and at the same time, provide an unbiased estimate of the HR function even in small samples. By contrast, some other models do not allow for accurate hypothesis testing (Kooperberg et al, 1995). The non-parametric regression spline model can be defined as following:

$$\lambda(t|\underline{x}) = \lambda_0(t)\, exp[\Sigma\beta_i(t)x_i] \qquad \textbf{[2.9]}$$

In 2.9, the $\underline{x} = (x_1,\ldots,x_k)$ is a vector of k covariates, $\lambda_0(t)$ is an unspecified baseline hazard function corresponding to $\underline{x}=0$, and $\beta_i(t)$ is the logarithm of the hazard ratio at time t corresponding to a unit increase in covariate x_i. The constant log hazard ratios β_i in the PH model are replaced by estimable function of time $\beta_i(t)$ in 2.9. The shape of the estimated function represents the pattern of changes in the predictor's effect over time. This shape is usually presented graphically, together with a pointwise 95% confidence interval that allows to assess the precision of the estimated hazard ratio at a given point during follow-up (Abrahamowicz et al, 1996).

The greater flexibility requires more degrees-of-freedom (d.f.), which selected depending on sample size and expected complexity of the HR functions to be estimated. The number of d.f. can be selected either a priori or a posteriori. Usually, for large samples, a 5 d.f. model, consisting of three quadratic polynomial pieces, is recommended (Abrahamowicz et al, 1996). The regression spline model has been used to reassess

predictive ability of baseline measurements in cancer (Rachet, et al, 1998; Quantin et al, 1999) and cardiovascular epidemiology (Côté et al, 1995; Lewis et al, 1997). However, little has been done to date in the area of flexible modeling of time-dependent effects of CHD risk factors.

Chapter 3

Methods

3.1 Objectives

The overall objective of this thesis is to enhance our understanding of the independent effects of within-subject changes in modifiable risk factors on coronary heart disease mortality. To achieve this, some methodological issues that might have affected the results of previous analyses must be addressed. Specifically, this research is motivated by the expectation that a proper definition of time-dependent covariates, which represent various aspects of change in predictor values over time, and their inclusion in the Cox (1972) proportional hazards model will offer more valid and more pertinent estimates of the impact of risk factor modifications than most of the previously published estimates.

As indicated by the review of the relevant literature in previous chapters, the explanation for this improvement is that the published results are mostly based on statistical models that have failed to represent the temporal relationship between recent changes in the risk factor value and subsequent outcomes. In fact, most conclusions regarding the expected effect of CHD risk factors modifications are based on between-

subject, rather than within-subject, analyses. Moreover, those few studies that considered data on longitudinal within-subjects changes did not represent these changes by time-dependent covariates, therefore restricting the ability to assess dynamic associations between changes occurring up to a given time, and outcomes that occur immediately afterwards.

For example, Pederson et al. (1998) estimated the impact of risk factor changes during the first part of follow-up on CHD events observed in the second part of follow-up, and thus ignored more recent changes, that might have had strong effects on subsequent outcomes. Therefore, this thesis will aim at overcoming the limitations of previous analyses, in order to provide new insights about the impact of changes in the modifiable risk factors on coronary heart disease mortality.

To achieve this, I will address the following methodological and substantive issues to re-analyze the data of the well-known Framingham heart study:

1. Comparison of the predictive ability of baseline versus updated value of a risk factor;
2. Development of a new approach for representing within-subject changes in risk factors by time-dependent covariates in the Cox model, for both continuous and binary predictors;
3. Incorporation of the information on intermediate non-fatal events observed during follow-up in the analyses discussed in 2 above, and assessment of the potential risk of confounding due to a failure to account for these events;

4. Assessment of the changes over time in the predictive ability of a baseline risk factor, based on flexible modeling of the time-varying hazard ratio;
5. Re-analysis of the Framingham study data using methods outlined in points 1-4 above, with a particular focus on the effects of changes in total serum cholesterol and smoking status.

3.2 Data Source

The Framingham heart study, which began in 1948, was administered by the National Heart Institute (NHI). About 10,000 of Framingham's 28,000 residents fell within the study's eligible age range of 30 to 62 years (the age group shown to precede the age group at risk of developing heart disease). From this group, 5,209 men (2336) and women (2873) agreed to physical exams every two years for the next 20 years. Of these, 5,127 were free of coronary heart disease at the start of the study (Friedman et al, 1967; Margolis et al, 1974 and Higgins, 1984). The objective of the Framingham Heart Study was to identify the common factors or characteristics that contribute to CHD by following its development over a long period of time in a large group of participants who had not yet developed overt symptoms of CHD or suffered a heart attack or stroke. The original participants underwent detailed physical examinations every two years, included an electrocardiogram, a chest X-ray, and various laboratory tests (Holland, 1990, NHLBI, 1999).

Fifty years have passed since the beginning of the Framingham heart study. Much of what is known today about the risk factors for cardiovascular disease - cigarette smoking, high blood pressure, high blood cholesterol, obesity, and diabetes — is at least partly based on the results from this study. Today, about 75 percent of the original Framingham study population has died (NHLBI, 1999). The most common cause of death for the participants was cardiovascular disease. Due to its great contributions to the understanding of cardiovascular disease, the Framingham Heart Study is now considered one of the most important epidemiological studies in medical history.

In this thesis, data on all 5209 participants of the original Framingham study will be re-analyzed. This analysis will include the data from visits 0 to 16, or up to 30 years of participant follow-up time.

3.3 Data Management and Definition of Study Variables

In contrast to some other well-known large prospective studies of CHD, which were limited to baseline risk factor evaluation, the design of the Framingham heart study involved repeated measurements of the common risk factors every second year for 30 years. As a result, up to 16 measurements are available for individual participants (Kannell et al, 1988). The availability of measurements repeated over a long period is essential for our purpose of revising the effect of longitudinal changes in risk factors values and separating these effects from the cross-sectional effects of baseline risk factors. However, to operate repeated measures, it is necessary to address the issue of missing data.

3.3.1 Missing data

According to the original protocol, each subject had to be evaluated biennially during 30 years of follow-up, and several risk factors such as serum cholesterol, blood pressure, smoking, were examined at each follow-up visit. Thus, there should be 16 values per subject for each risk factor measured. As is to be expected, some subjects have incomplete data. In order to avoid eliminating these subjects from the repeated-measures analyses, the missing data must be replaced.

There are many ways to handle the missing data (Brand et al, 1994; Brick et al, 1996; Fayers et al, 1998). Some of the most common strategies for replacing missing data would not be appropriate and/or effective in the context of this study. First, pairwise correlations between risk factors at a specific visit (e.g. cholesterol at visit 7 Vs systolic blood pressure at visit 7) were low. Thus, replacing a missing value by an estimate based on regressing the relevant risk factor on all other non-missing risk factor values would produce very noisy estimates. Moreover, regression-based replacements would be impossible in the frequent cases when all risk factor values for a given participant were missing for a given visit, presumably because the subject failed to come for this specific visit. Another popular method is to replace a missing value of a quantitative independent variable by its sample mean, calculated from all subjects with non-missing values. The main disadvantage of this approach is that it ignores the information on the values observed in the same subject at other visits. Thus, even if a given subject's serum cholesterol is systematically 1.5 to 2 standard deviations above the mean of the values for

all subjects, the missing value(s) would be replaced by the mean. This approach would create very large values for the difference between subsequent cholesterol values for some subjects, creating potentially influential data points.

To avoid the undesirable influence of imputed values, a more conservative carry-forward approach has been employed for this thesis. The essence of the carry-forward approach is that a missing value is replaced by the closest-in-time, earlier, non-missing value for the same risk factor, for the same subject. For example, if a subject has missing cholesterol values at visits 5 and 6, both will be replaced by this subject's cholesterol measurement from visit 4. In the rare instances where the initial value was missing, it was replaced by the first available measurement (typically visit 2 or visit 3). The approach is conservative in the sense that it reduces the within-subject variability of the risk factor values over time, as it implies no changes during the periods where data are missing. On the other hand, it may occasionally result in an over-estimation of changes from the last visit, if the imputed previous value actually represents the value observed several years earlier. Software for this data rearrangement has been written in the C programming language (C++ program design, 1997).

3.3.2 Identification of outcomes and censoring

The primary study outcome was death from CHD. Death was defined as fatal, definite or probable (history and ECG) CHD event, which included myocardial infraction, coronary insufficiency, and angina pectoris. Because the exact date of death

was not available in the original Framingham database (Feinleib, 1985), time-to-event was defined as time until the first visit at which the subject was declared dead, and the cause of death was recorded as definite or probable CHD in this study. Computer simulations using a dataset that was similar in size and structure to the Framingham data indicated that the resulting loss of precision had practically no impact on the results from a Cox model (Abrahamowicz, 1998, personal communication).

For the purpose of this study, only those subjects who were dead from CHD were considered as having the event. All other subjects are censored at (i) the time of non-CHD death, (ii) the time of loss to follow-up or (iii) visit 16, whichever of the three endpoints occurred first. Among the 5209 subjects, 29 subjects had values missing for each of the variables. Of the remaining 5180 subjects, 4582 subjects were censored.

3.4 Construction of Time-dependent Variables for Cholesterol and Systolic Blood Pressure

Total serum cholesterol is an important component of cell membranes and is vital to the structure and function of all cells in our body. However, cholesterol is a predominant substance in antherosclerotic plaques, which may develop in arteries and impede the flow of blood, and thus contribute to the development of CHD. Many studies have already shown that high serum cholesterol is the main risk factor for CHD (Manninen et al, 1988; Posner et al, 1991; Menotti, 1992).

For example, in the Scandinavian Simvastatin Survival Study (4S) (Pederson et al, 1998), it was found that each additional 1% reduction in total cholesterol, corresponded to a reduction of CHD risk by 1.9% (P=0.00005). A percent decrease in low density lipoprotein cholesterol (LDL-C) and Non-high density lipoprotein cholesterol (HDL-C), would reduce risk by 1.7% (P<0.00001). These results demonstrate that lower serum cholesterol is strongly related to lower CHD risk. However, these estimates are based on a comparison of subjects with different post-intervention cholesterol values and it is unclear to what extent they may be extrapolated to predict the effect of longitudinal within-subject changes.

In this thesis, I used the Cox proportional hazards model to assess the impact of serum cholesterol, either at baseline (initial) or updated, as well as the effects of subsequent changes from the baseline value, or changes from the last visit's values, on the risks of CHD mortality.

Assume repeated measures of serum cholesterol value are represented by variables X (j), j = 0, 2, 4, ...30, with j indicating the number of years elapsed since the beginning of the study. I defined X_2 as the baseline variable (at 2 year visit). Because most subjects, detailed measurements of risk factors were first carried out at visit 2. Then, the following time-dependent variables were constructed:

- Current cholesterol value (most recent value)

$$X(t) = X_j \qquad j \leq t < j+2$$

where it denotes follow-up time in years.

- Absolute change from the baseline value

$$B(t) = X(t) - X_2 \qquad \text{for } t > 2$$

If $t \leq 2$, $B(t) = 0$.

- Absolute change from the last value

$$D(t) = X(t) - X_{j-2} \qquad j-2 < t \leq j$$

If $t \leq 2$; $D(t) = 0$

- Relative change from baseline value

$$B^*(t) = \frac{B(t)}{X_2}$$

- Relative change from last value

$$D^*(t) = \frac{D(t)}{X_{j-2}}$$

In order to help analyze the effect of these variables, I also constructed another new time-dependent variable: CHDNEW. I defined this variable as a binary variable. If a person did not have evidence of CHD other than death until a given visit, I assigned a variable of 0 at this visit, and all previous visits. When a person had the first evidence of CHD at visit j, I defined CHDNEW as 1 at this visit and the following visit ($j \geq 1$).

To calculate the latter five time-dependent variables, I use a SAS program (SAS Institute). The same procedures and formulas were used to construct time-dependent variables that were related to changes in systolic blood pressure.

3.5 Construction of Time-dependent Variables for Smoking

Cigarette smoking is a powerful risk factor for CHD that probably predisposes the smoker to CHD in several ways (Jajich et al, 1984; Mason et al, 1985; Seltzer, 1991) and is especially dangerous in-patients with advanced coronary atherosclerosis (Tresch et al, 1996; Rigotti et al, 1996).

New time-dependent variables were also calculated to reflect the subjects' current smoking status, as well as the longitudinal changes in this status during the follow-up period. These variables were then included in a multivariable Cox model to test whether changes in smoking status were associated with the risks of death from CHD. The operational definitions of these newly constructed binary time-dependent variables are showed below:

- Current smoking status:

S (t)=1 if smoker at visit $j \leq t < j+2$

S (t)=0 otherwise

- Indicator of a new smoker:

$$\text{Snew}(t) = \begin{cases} 0 & \text{Otherwise} \\ 1 & \text{Subject did not smoke at baseline (t=2) and started to smoke later, i.e. S(2)=0 and S(t)=1} \end{cases}$$

- Indicator of smoking cessation

$$\text{Sstop}(t) = \begin{cases} 0 & \text{Otherwise} \\ 1 & \text{Subject smoked at baseline and stopped smoking later, i.e. S(2)=1 and S (t)=0.} \end{cases}$$

For example, if a person who smoked at baseline (visit 2) stopped smoking at the 6th visit, I defined the time-dependent smoking cessation variable Sstop(t) = 1 at the 6th visit and all the following visits. The same method was used to define the time-dependent variable indicating a new smoker that takes the value Snew(t)=1 at the visit the person began to smoke (for a person who did not smoke at baseline) and all the following visits. Table 3.1 illustrates the construction of time-dependent variables for smoking. Subject A is a person who did not smoke at the baseline visit (visit year=2) and began to smoke at the 3rd visit (visit year=6). For this subject, all the values of Sstop(t) equal to 0 and the

values of S(t) and Snew(t) are equal to 1 at the 3rd visit and the following visits. Subject B is a person who smoked at baseline visit and stopped smoking at the 4th visit (visit year=8). Accordingly, all the values of Snew(t) equal to 0, the values of Sstop(t) are to equal 1 for the 4th visit, and the following visits, and the values of S(t) are 1 until year 6 and 0 thereafter.

3.6. Models of the Effects of Changes in Modifiable Risk Factors

3.6.1. General modeling strategy

All multivariable survival analyses relied on the Cox (1972) PH regression model or on its non-parametric generalization, in which a time-dependent hazard ratio is modeled by regression splines (Abrahamowicz et al, 1996). In all survival analyses, CHD death was used as the endpoint event. Subjects who did not die of CHD were censored at the earliest of the three occurrences: (i) non-CHD death, (ii) loss to follow-up, or (iii) alive at visit 16, corresponding to the end of the 30 years follow-up period. All models included baseline values of the following common CHD risk factors: age, gender, body mass index (BMI) and glucose intolerance (yes/no). In addition, at least one variable related to each of the other three CHD risk factors: total serum cholesterol, systolic blood pressure and smoking status was included.

Table 3.1. Methods for the Construction of Time-dependent Variables for Smoking

Subject	Visit Year	Baseline Smoking Status	Current Smoking Status S (t)	StopSmoker C (t)	NewSmoker N (t)
A	2	0	0	0	0
	4		0	0	0
	6		1	0	1
	8		1	0	1
	10		1	0	1
	12		1	0	1
	14		1	0	1
B	2	1	1	0	0
	4		1	0	0
	6		1	0	0
	8		0	1	0
	10		0	1	0
	12		0	1	0
	14		0	1	0
	16		0	1	0
	18		0	1	0

The main focus of the analyses reported in this thesis is on these three latter risk factors, each of which can be modified through various interventions. To evaluate to what extent changes in these modifiable risk factors are associated with the risks of CHD mortality, additional analyses were carried out. For these analyses, the effects of interest were represented by various time-dependent variables, which were defined in sections 3.3 and 3.4. To facilitate interpretation of the results, most of these analyses attempted to assess the impact of changes in just one of the three modifiable risk factors, while adjusting for the baseline values of all other common CHD risk factors, including initial values of the two other modifiable factors. When three time-dependent variables, each representing changes in one of the three modifiable variables (cholesterol, SBP, smoking) were included in the same model, the result did not change materially.

3.6.2. Comparing prognostic ability of initial and updated values of modifiable risk factors

For each of the three modifiable risk factors, two alternative multivariable Cox models were estimated. In one model, a risk factor was represented by a fixed-in-time variable corresponding to its initial value. In the second model, the initial value was replaced by a time-dependent variable that represented the updated, or the most recent, value of this factor. As both models included exactly the same set of other independent variables, this modeling strategy allowed a head-to-head comparison of the prognostic ability of initial versus updated measurements of the risk factor. This allowance provided an opportunity to assess to what extent the Framingham data corroborate one of the two alternative conjectures, according to which the current level of cardiovascular risk in an

individual is mostly determined by, (i) earlier (mid-life) values, as postulated by Benfante et al. (1994); or (ii) recent, updated values, as suggested by Cupples (1988) and Abrahamowicz et al (1997), respectively. Because both variables use 1 degree of freedom to represent a given risk factor, their predictive ability may be compared directly by an appropriate goodness-of-fit statistics for the corresponding multivariable models. Specifically, I used the log likelihood of the data under a given model as the primary criterion for these comparisons. Moreover, as both variables use the same measurement unit, the resulting hazard ratio estimates can be directly contrasted to establish which of the two has a stronger independent impact on CHD mortality, after having adjusted for all other common risk factors.

To further assess the prognostic utility of the baseline values of selected modifiable risk factors, and in particular, total serum cholesterol and smoking, I have tested if their predictive ability changes over time. Adjusted hazard ratio was employed as a measure of the variable's ability to predict CHD deaths (Lewis et al, 1997). To achieve this, I used the regression spline generalization of the Cox model to estimate time-dependent hazard ratios (Abrahamowicz et al, 1996) for each of the three risk factors, while adjusting for the baseline values of all other predictors. Following recommendations by Abrahamowicz et al (1996), the time-dependent effect was modeled using 5 degrees of freedom (df), so that statistical significance of the changes in the effect of a given baseline variable was tested using a 4-df likelihood ratio test (LRT). The significant result of this test, at a 0.05 significance level, was interpreted as evidence that the predictive ability of the risk factor does change during the follow-up (Quantin et al, 1999). In that case, the shape of the estimated hazard ratio function allows for interpretation of the direction of these changes (Lewis et al, 1997). However, if -

according to the LRT test - the changes are not statistically significant, the graph of the hazard ratio function does not provide reliable information on the pattern of changes and any apparent changes in the estimated function should be considered as likely due to over-fitting bias (Quantin et al, 1999).

3.6.3. Assessing prognostic ability of time-dependent variables representing different aspects of changes in risk factor

To evaluate the associations between various measures of changes in a risk factor, I relied on time-dependent covariates, which were defined in sections 3.3 and 3.4. In order to avoid bias, a change variable had to be adjusted for another measure of the same risk factor, representing its constant value that was taken at a given point during follow-up. Otherwise, the results may be biased. For example, this bias may occur because of the regression to the mean phenomenon, according to which subjects with very high initial risk factor values are more likely to decrease their values over time, while very low initial values have a tendency to increase over time. Thus, if the initial value is not taken into account, then subjects with increasing cholesterol values may be found to have lower risks than those with decreasing cholesterol, just because their increases and decreases will act as markers for low and high initial levels, respectively. To avoid such a confounding bias and to dissociate the effects of baseline cholesterol levels from the impact of subsequent changes in cholesterol levels, each of the four change variables was adjusted for the baseline value in a separate model. This adjustment allowed me to assess to what extent the recent change from either baseline or previous risk factor value helped predict CHD deaths among subjects with the same initial

value. Thus, I investigated whether there was a systematic impact of change that was independent from the baseline value.

A similar modeling strategy was used to adjust the effects of different change variables for the updated - rather than baseline - value of the same risk factor. However, the interpretation of the results of a model that contains a time-dependent updated measure of variables such as cholesterol, in addition to a time-dependent measure of its change, is less straightforward than a model with the baseline and change variables. Inclusion of the variable representing the change from the previous measurement, together with the current value, allowed me to assess if - among subjects with the same current value - those who recently increased their cholesterol levels had different risks from those who stayed at the same level for a longer period. Interpretation of the estimate for the change-from-baseline variable is more complex when adjusted for the current value. This interpretation has to take into account that - among subjects with the same current value - there is a perfect negative correlation between the amount of change from the baseline (current - baseline) and the actual baseline value. Thus, a more positive value for the change from baseline, when adjusted for the current value, becomes in fact a proxy for a lower baseline value. Therefore, in chapter 4, when interpreting the results of models with two time-dependent covariates (current value and change from the baseline), I will take the above relationship into account.

Chapter 4

Results

The main objective of this study is to compare the predictive ability of the CHD risk factors measured at the baseline with newly constructed time-dependent variables representing current value and/or different measures of change. I focus on the effects of two modifiable risk factors: total serum cholesterol and smoking.

In this chapter I report on the results of several multivariable Cox models that were estimated and compared to draw conclusions regarding predictive ability of relevant variables. In all analyses, the outcome was time to CHD mortality. The additional baseline covariates taken into account were diastolic blood pressure (DBP), age, body mass index (BMI), sex and a binary indicator of glucose intolerance. In some post hoc analyses, an additional time-dependent binary covariate, representing the occurrence of a post-baseline CHD event, was introduced to further explore the mechanisms involved in the longitudinal evaluation of CHD and its risk factors.

4.1 Effects of Total Serum Cholesterol

The association of total serum cholesterol levels with CHD mortality is well established in many studies, such as Framingham heart study (Gordon et al, 1977) and Honolulu heart study (Stemmermann et al, 1991;). In this study, I will compare the baseline cholesterol values with the new time-dependent variables with respect to their ability to predict the CHD mortality.

4.1.1 Summary of distributions of time-dependent variables for cholesterol

Table 4.1.1 shows the descriptive statistics for different cholesterol variables, including baseline value, current value, absolute and relative changes from baseline value, as well as absolute and relative change from current value. To ensure comparability of results across different follow-up visits, all results in Table 4.1.1 are restricted to a subset of subjects who have been followed for 30 years, i.e. until the last visit. For time-dependent variables, their distributions are shown only at selected visits, specifically at 2,6,12,24,26 years from the baseline visit. For each variable, the following sample statistics are reported: the arithmetic mean of the all values for a given visit, their median, standard deviation (St Dev), and the 25^{th} and 75^{th} percentiles (Q_1, Q3). Results in Table 4.1.1 indicate that the general trend in changes of cholesterol over time was non-monotone; the mean value increased initially and decreased in the later phase of follow-

Table 4.1.1 Descriptive Statistics for the Distributions of Different Cholesterol Variables at Selected Follow-up Visits

Variable	Years from first visit	mean	median	Std Dev	Q1-Q3	Min	Max
Baseline (mg/dL)	0	227.7	224	45.0	196, 255	96	568
Current (mg/dL)	2	229.8	225	44.9	199, 255	110	600
	6	240.3	237	46.0	208, 267	107	696
	12	245.7	243	46.3	213, 274	113	614
	24	232.6	229	45.1	202, 260	96	538
	26	231.7	228	45.4	201, 259	84	538
Absolute Change from baseline (mg/dL)	2	2.1	0	27.7	-10, 16	-230	146
	6	12.6	10	33.1	-4, 31	-288	337
	12	18.0	16	38.4	-2, 41	-254	255
	24	4.9	3	43.9	-21, 31	-270	227
	26	3.9	3	44.5	-21, 31.5	-264	227
Absolute Change from last (mg/dL)	2	2.1	0	27.7	-10, 16	-230	146
	6	3.8	0	26.1	-8, 16	-141	372
	12	-3.6	0	28.0	-16, 6	-738	356
	24	-0.91	0	22.6	-4, 2	-346	132
	26	-0.9	0	20.2	-1, 0	-141	150

variable	Years from first visit	mean	median	Std Dev	Q1-Q3	Min	Max
Relative change from baseline	2	0.017	0	0.12	-0.05, 0.07	0.9	-0.5
	6	0.067	0.046	0.15	-0.02, 0.15	1.2	-0.5
	12	0.095	0.074	0.18	0, 0.20	1.3	-0.5
	24	0.040	0.014	0.20	-0.09, 0.15	1.1	-0.6
	26	0.036	0.011	0.20	-0.09, 0.15	1.3	-0.6
Relative Change from last	2	0.017	0	0.12	-0.05, 0.07	0.9	0
	6	0.022	0	0.11	-0.03, 0.07	1.1	-0.5
	12	-0.009	0	0.10	-0.06, 0.03	0.4	-0.7
	24	0.001	0	0.10	-0.02, 0.01	1.1	-0.6
	26	-0.001	0	0.09	0, 0	0.9	-0.5

up. However, there were considerable individual differences in the direction and magnitude of change. Comparison of Q_1 and median values for absolute change from the baseline shows that, at each subsequent visits, more than 25% of subjects had values lower than their baseline values while at least 50% of subjects had values higher than the baseline. Standard deviation of both absolute and relative changes from the baseline increased monotonically with increasing follow-up time, indicating that the initial cholesterol measurement becomes an increasingly inaccurate indicator of the current value. The result of the Pearson correlation test showed that the correlation coefficients between baseline (1^{st} visit) cholesterol value with 2^{nd} visit value is 0.81, 0.74 with 4^{th} visit value, 0.65 with 7^{th} visit value, 0.5 with 13^{th} visit value, 0.47 with 14^{th} visit value. The coefficients is decreasing with the increase of years from 1^{st} visit, this trend can further confirm that baseline cholesterol value become less and less representative of the current cholesterol value with the follow-up visits.

On the other hand, most changes are rather small, with standard deviation of changes not exceeding 20% of the mean baseline value. This may have some implications for the statistical power of testing the effect of changes in total cholesterol on CHD mortality. As expected, changes from the last measurement (typically 2 years before) are systematically smaller than changes from the baseline.

4.1.2 Comparing predictive ability of baseline cholesterol and current cholesterol variable

Table 4.1.2 compares the goodness-of-fit for two models, corresponding to baseline and current cholesterol variable, respectively. The Likelihood Ratio Test (LRT), Wald and Score are used to test whether all independent variables in a given model improve prediction of CHD deaths significantly better than the null model without covariates. These statistics are also useful for comparing fit of different models: higher values of Wald and Score, and lower value of LRT indicate a better-fitting model.

Table 4.1.2 indicates that both model with baseline and model with current cholesterol are very significant using any of these three tests (p-value approximately 0.0001). We also can see that the value of -2 LOG L of the baseline model (12361.820) is considerably lower than that of the current model (12397.216), which indicates that baseline cholesterol is a much more powerful predictor than the current cholesterol value. This is also confirmed by substantially higher values of Score and Wald tests for the model with baseline cholesterol.

Table 4.1.3 shows the detailed results of Cox models with, respectively, baseline and current cholesterol. From Table 4.1.3, we can see that all of the common cardiovascular risk factors: cholesterol, diastolic blood pressure (DBP), smoking, glucose intolerance, BMI, age and sex are highly significant in both models (p-value<<0.05).

Table 4.1.2. Comparing Models with Baseline versus Current Cholesterol Values: Testing Global Null Hypothesis

Cholesterol variable	Model's Fit (-2LogL)	Criterion	Model Chi-Square
Baseline	**12361.820**	LRT*	745.995 with 7 DF (p=0.0001)
Current	**12397.216**		710.598 with 7 DF (p=0.0001)
Baseline	.	Score	725.416 with 7 DF (p=0.0001)
Current	.		704.103 with 7 DF (p=0.0001)
Baseline	.	Wald	660.088 with 7 DF (p=0.0001)
Current	.		643.547 with 7 DF (p=0.0001)

* LRT = Likelihood Ratio Test

Table 4.1.3. Results of Two Multivariable Regression Models: Baseline versus Current Cholesterol Value

Variables	B/C	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
Total serum	Baseline(a)	0.0073	1.007	(1.006, 1.009)	71.7	**0.0001**
cholesterol (mg/dl)	Current (b)	0.0048	1.005	(1.003, 1.006)	34.07	**0.0001**
DBP	Baseline	0.0329	1.033	(1.027, 1.040)	97.7	**0.0001**
	Current	0.0346	1.035	(1.028, 1.042)	109.9	**0.0001**
Smoking	Baseline	0.3373	1.401	(1.160, 1.693)	12.2	**0.0005**
(current/no-smoker)	Current	0.3389	1.403	(1.163, 1.693)	12.5	**0.0004**
Diabetes	Baseline	0.8246	2.281	(1.741, 2.989)	35.8	**0.0001**
(Yes/No)	Current	0.8217	2.274	(1.736, 2.980)	35.5	**0.0001**
BMI (kg/m^2)	Baseline	0.0309	1.031	(1.010, 1.053)	8.7	**0.0031**
	Current	0.0313	1.032	(1.011, 1.053)	9.3	**0.0023**
Age (years)	Baseline	0.0850	1.089	(1.077, 1.101)	240.1	**0.0001**
	Current	0.0895	1.094	(1.082, 1.105)	267.0	**0.0001**
Sex	Baseline	-1.080	0.340	(0.281, 0.411)	122.3	**0.0001**
(F/M)	Current	-1.085	0.338	(0.279, 0.409)	122.2	**0.0001**

(a) Baseline denotes Cox model with baseline total serum cholesterol variable

(b) Current denotes Cox model with current (updated) total serum cholesterol variable

Comparison of the first two rows in Table 4.1.3 shows that the adjusted impact of 1mg/dL increase in baseline cholesterol on CHD mortality is higher than the same increase in current cholesterol (HR=1.007 Vs HR=1.005). The fact that the corresponding 95% confidence intervals barely overlap suggests this difference is rather robust and systematic. The much higher value of the Wald statistics (71.68364>34.03990) further confirms that baseline cholesterol value is a more powerful predictor than current cholesterol value in predicting the CHD mortality. The fact that the hazard ratios for all covariates are almost the same in the two models suggests the difference between the estimates for the two cholesterol variables is unlikely to be confounded by other risk factors.

4.1.3 Assessing the impact of within-subject changes in total serum cholesterol

To dissociate the cross-sectional effects of baseline risk factors from the effects of longitudinal within-subject changes in these values, I have constructed the new time-dependent variables that represent, respectively, change from last value, change from baseline value, as well as the corresponding relative change measurements (see section 3.3).

From Table 4.1.4, we can see that all of these new constructed time-dependent covariates do not have statistically significant association with CHD mortality (all p-values>>0.05) when adjusted for *baseline* cholesterol model. By contrast, after adjusting

for any of the change variables, the baseline cholesterol value still shows highly significant associations with CHD mortality (all p-values<0.0001). In addition, the estimated hazard ratios for change from the last value (1.002) and change from the baseline (1.000) are much smaller than the hazard ratio for baseline cholesterol (1.007) and their 95% confidence intervals do not overlap at all. This indicates that the non-significance of the effects of change variables is not simply due to low statistical power, because of limited variation. Thus, introducing the variables representing changes in total serum cholesterol does not enhance the predictive ability of CHD mortality over and above prediction based on baseline value, while baseline cholesterol remains a very significant predictor even when adjusted for various time-dependent measures of change.

The results of the four Cox regression models that included *current* cholesterol value and one of time-dependent variables representing changes in cholesterol, are shown in Table 4.1.5. Both variables representing changes from the last value are statistically non-significant (p>0.05). In addition, the HR for change for the last value is very close to 1.0. Thus, information on the recent change in cholesterol does not enhances the prediction of CHD deaths, once the updated value is available. However, p-values in the last column of Table 4.1.5 show that both time-dependent variables representing change from the baseline (absolute and relative, respectively) are statistically very significant when adjusted for current cholesterol. The hazard ratios are 0.993 and 0.205 respectively, that would suggest that a reduction of total serum cholesterol from the baseline is associated with an increase in the risk of CHD death.

Table 4.1.4. Estimated Effects of Time-dependent Variables Representing Changes in Cholesterol, Adjusted for BASELINE Cholesterol Value

Model	Variables	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
1	Baseline SCL(mg/dL)	0.0074	1.007	(1.006, 1.009)	72. 8	0.0001
	Absolute change from last (mg/dL)	0.0021	1.002	(0.999, 1.005)	1.7	**0.1984**
2	Baseline SCL(mg/dL)	0.0074	1.007	(1.006, 1.009)	62.1	0.0001
	Absolute change from baseline (mg/dL)	0.0003	1.000	(0.998, 1.003)	0.1	**0.7824**
3	Baseline SCL(mg/dL)	0.0074	1.007	(1.006, 1.009)	71.7	0.0001
	Relative change from last	0.0940	1.099	(0.486, 2.483)	0.1	**0.8213**
4	Baseline SCL(mg/dL)	0.0076	1.008	(1.006, 1.009)	65.8	0.0001
	Relative change from baseline	0.1980	1.219	(0.694, 2.142)	0.5	**0.4911**

- All results are from the multivariable models that include diastolic blood pressure, smoking, body mass index, age, sex and diabetes.

Table 4.1.5. Results of Cox Regression with Time-dependent Variables Representing Changes in Cholesterol Value, Adjusted for CURRENT Cholesterol Value

Model	Variables	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
5	Current SCL(mg/dL)	0.005096	1.005	(1.003, 1.007)	32.19945	0.0001
	Absolute change from last (mg/dL)	-0.001421	0.999	(0.995, 1.002)	0.72023	**0.3961**
6	Current SCL(mg/dL)	0.007447	1.007	(1.006, 1.009)	62.11315	0.0001
	Absolute change from baseline (mg/dL)	-0.007126	0.993	(0.991, .0995)	37.08331	**0.0001**
7	Current SCL(mg/dL)	0.005335	1.005	(1.004, 1.007)	23.97972	0.0001
	Relative change from last	-0.803440	0.448	(0.188, 1.068)	3.28364	**0.0700**
8	Current SCL(mg/dL)	0.007142	1.007	(1.005, 1.009)	57.65264	0.0001
	Relative change from baseline	-1.586027	0.205	(0.113, 0.371)	27.33906	**0.0001**

- All results are from the multivariable models that include diastolic blood pressure, smoking, body mass index, age, sex and diabetes.

At a first glance, these statistically significant associations seem contradictory with a conventional wisdom regarding the impact of cholesterol. However, a more in-depth interpretation of these results indicates that these apparently negative associations are, in fact, due to an artifact of the statistical modeling. To understand this methodologically interesting phenomenon, one should first take into account that Table 4.1.2 showed convincingly that after having adjusted for baseline values of other common CHD risk factors, baseline cholesterol is a much more powerful predictor of CHD mortality than current cholesterol. Now, in a model that includes two time-dependent variables: the current cholesterol and the difference between current and baseline, among subjects who have the same current value, those that had higher baseline will automatically have lower change from baseline (change=current-baseline). Therefore, the change-from-baseline variable, when adjusted for the current value, is in fact a negative-slope linear-function of baseline cholesterol. This leads to the final conclusion that the significant negative effect of the change-from-baseline represents, in fact, a disguised positive effect of baseline cholesterol, when adjusted for its current value. Among individuals with same current cholesterol, those who had higher baseline, have higher risk. Because in this model, higher baseline corresponds to lower change from baseline, this produces a spurious negative correlation between change-from-baseline and CHD risk.

4.1.4 Revised analyses: adjusting the effects of time-dependent cholesterol variables for the newly incident non-fatal CHD

In all analyses reported in section 4.1.2 and 4.1.3, the impact of total serum cholesterol was adjusted only for those relevant patients' characteristics that were available at the time of the baseline visit (visit 2). This is consistent with the conventional regression modeling of time-to-event data in which the effect of "exposure" (risk factor, treatment etc.) is typically not adjusted for any intermediate outcome or change in the subject's health status that may occur after time 0. However, in the view of results presented above, it seems possible that this conventional approach could induce some confounding bias in estimates of time-dependent variables. Specifically, there is a risk for such a bias if some changes in cholesterol level of individual subjects during the follow-up are simulated by an earlier event indicating a deterioration in their health. Assume that a recent change in health status increases the individual's risk of CHD death and s/he is aware of this. Then it is conceivable that the individual will attempt to reduce his/her risk by lowering some modifiable risk factors. If, in this situation, the original event is not taken into account in the analysis, then the subsequent reduction in his/her cholesterol level becomes, at least partly, a marker for a precious health deterioration and may appear to be associated with increased risks of CHD death. These considerations suggest that it may be worthwhile to repeat analyses reported in section 4.1.2 and 4.1.3 while adjusting for incidence of non-fatal CHD event(s) that might have occurred during the follow-up. To achieve this, I introduced in the regression model an additional binary time-dependent

covariate indicating whether an individual has had a non-fatal CHD event at any time in the past (in section 3.3 I describe in detail how this variable was constructed).

Table 4.1.6 has the same structure as Table 4.1.2 but gives opposite results after having adjusted for the incident CHD events, the model with current cholesterol predicts CHD mortality better than the model with baseline cholesterol as indicated by higher values of all three global test statistics. Comparing the Table 4.1.2 and 4.1.6, we see also that adding the new time-dependent variable improves considerably the model's fit as the –2 LOG L values in the models with CHDNEW are dramatically smaller than in the models without it (10346.4<12361.2, 10333.0<12397.2). This is also confirmed by substantially higher values of Score and Wald tests for the models with CHDNEW. As expected, a non-fatal CHD event is a very powerful predictor of a subsequent CHD death.

Table 4.1.7 further confirms that after adjusting for incident non-fatal CHD events, updated cholesterol becomes a stronger predictor of CHD death than the baseline cholesterol. This is indicated by higher values of both estimated hazard ratio and Wald statistics for the time-dependent variable representing current serum cholesterol level. Comparison of results in Table 4.1.3 and 4.1.7 shows that the estimated effect of baseline cholesterol is reduced after having adjusted for CHDNEW (HR of 1.004 Vs HR of 1.007) and the same is observed for most other baseline measurements of risk factors. The reason is that some of the impact of baseline risk factors on CHD death are mediated through the occurrence of non-fatal CHD events.

Table 4.1.6. Comparing Models with Baseline versus Current Cholesterol with Adjustment for Incidence of CHD during Follow-up: Testing Global Null Hypothesis

Cholesterol variable	Model's Fit (-2LogL)	Criterion	Model Chi-Square
Baseline	**10346.426**	LRT*	2761.388 with 8 DF (p=0.0001)
Current	**10333.002**		2774.812 with 8 DF (p=0.0001)
Baseline	.	Score	5914.186 with 8 DF (p=0.0001)
Current	.		5918.417 with 8 DF (p=0.0001)
Baseline	.	Wald	1101.393 with 8 DF (p=0.0001)
Current	.		1114.689 with 8 DF (p=0.0001)

* LRT = Likelihood Ratio Test

Table 4.1.7. Results of Two Multivariable Regression Models: Baseline versus Current Cholesterol Value with Adjustment for Incidence of CHD during Follow-up

Variables	Model	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
Total serum	Baseline(a)	0.0037	1.004	(1.002, 1.005)	17.7	0.0001
cholesterol (mg/dl)	Current(b)	0.0046	1.005	(1.003, 1.006)	32.9	0.0001
CHDNEW	Baseline	4.7919	120.530	(85.35,170.2)	740.3	0.0001
	Current	4.8190	123.835	(87.75,174.8)	752.1	0.0001
DBP	Baseline	0.0193	1.019	(1.013, 1.026)	35.7	0.0001
	Current	0.0197	1.020	(1.013, 1.026)	37.7	0.0001
Smoking	Baseline	0.3053	1.357	(1.112, 1.655)	9.1	0.0026
(current/no-smoker)	Current	0.3135	1.368	(1.122, 1.669)	9.6	0.0020
Diabetes	Baseline	0.7782	2.177	(1.658, 2.859)	31.3	0.0001
(Yes/No)	Current	0.7801	2.182	(1.661, 2.866)	31.4	0.0001
BMI (kg/m^2)	Baseline	-0.0012	0.999	(0.997, 1.021)	0.01	**0.9100**
	Current	0.0023	1.002	(0.981, 1.024)	0.04	**0.8375**
Age (years)	Baseline	0.0534	1.055	(1.043, 1.066)	90.8	0.0001
	Current	0.0570	1.059	(1.047, 1.070)	102.2	0.0001
Sex	Baseline	-0.6556	0.519	(0.424, 0.635)	40.6	0.0001
(F/M)	Current	-0.7353	0.479	(0.390, 0.589)	49.2	0.0001

(a) Baseline denotes Cox model with baseline total serum cholesterol variable

(b) Current denotes Cox model with current (updated) total serum cholesterol variable

From Table 4.1.7, we see also that after adding the new time-dependent variable-CHDNEW, BMI becomes completely non-significant (p-value ≈1), and whereas all other risk factors still have very significant associations with CHD mortality. This result is consistent with other studies that also reported non-significance of the linear effect of BMI. For example, Abrahamowicz et al. (1997) used conventional multiple logistic regression to analyze the CHD mortality in the LRC follow-up study and found that there is no evidence of the independent effect of BMI on the risk of CHD death and the adjusted odds ratio is very close to 1.0 when the relationship between BMI and logodds is assumed to be linear on the logit scale. However, non-parametric Generalized additive model analyses suggested BMI may have a non-monotonic association with CHD mortality, with risk increasing in both left and right tails of BMI distribution (Abrahamowicz, et al, 1997).

Table 4.1.8 and 4.1.9 re-evaluate the role of various measures of change in serum cholesterol when adjusted for the incident non-fatal CHD (CHDNEW) and, respectively, baseline and current cholesterol value. Comparison of Table 4.1.4 and 4.1.8 shows that the inclusion of the new time-dependent covariate CHDNEW changes conclusions regarding the impact of changes from baseline cholesterol. In fact, whereas both measures of changes from baseline were not significant in Table 4.1.4, they are both highly significant in Table 4.1.8. The latter indicates that among individual with the same current value of CHDNEW and the same baseline cholesterol, the amount of increase from baseline is strongly associated with the risk of CHD death. The hazard ratio for 1mg/dL increase from baseline is, indeed, quite similar to that for 1 mg/dL difference in

the baseline value (1.004 Vs 1.005). The amount of change from the last cholesterol measurement, typically taken 2 years earlier, is also a significant predictor of subsequent CHD death, when adjusted for baseline value while the relative change for the last value is marginally non-significant ($p<0.09$).

Furthermore, the estimated effects of changes from the baseline and from the last value are very similar, suggesting that the difference in the corresponding p-value is mostly due to lower variance of the distribution of the latter change, that is measured over a period of only about 2 years (see Table 4.1.1). Thus, the results in Table 4.1.8 show that both long-term increases (from baseline) and short-term increases in serum cholesterol are associated with significantly higher risk of CHD mortality, once baseline value and current CHD status (symptomatic Vs asymptomatic) are taken into account. By contrast, Table 4.1.9 shows that there are no significant associations between various change variables after having adjusted for the current cholesterol value, in addition to adjustment for CHDNEW. Taking into account the interpretation of the change from baseline in the model with current value (see section 4.1.3), it is important to note that this change is not significant in Table 4.1.9 and the estimated effect is very small (log hazard ratio is 5 times smaller than for current cholesterol: 0.0010 Vs 0.0050). Thus, once current CHD status and current cholesterol value are taken into account, the initial cholesterol value is of little relevance for predicting CHD mortality.

Table 4.1.8. Estimated Effects of Time-dependent Variables Representing Changes in Cholesterol, Adjusted for BASELINE Cholesterol with Adjustment for Incidence of CHD During Follow-up:

Model	Variables	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
1	Baseline SCL(mg/dL)	0.0037	1.004	(1.002, 1.005)	17.9	0.0001
	Absolute change from last (mg/dL)	0.0043	1.004	(1.001, 1.007)	7.4	0.0064
2	Baseline SCL(mg/dL)	0.0050	1.005	(1.003, 1.007)	29.8	0.0001
	Absolute change from baseline (mg/dL)	0.0040	1.004	(1.002, 1.006)	14.6	0.0001
3	Baseline SCL(mg/dL)	0.0037	1.004	(1.002, 1.005)	18.0	0.0001
	Relative change from last	0.6989	2.012	(0.915, 4.420)	3.0	0.0819
4	Baseline SCL(mg/dL)	0.0051	1.005	(1.003, 1.007)	30.8	0.0001
	Relative change from baseline	1.0455	2.845	(1.741, 4.648)	17.4	0.0001

- All results are from the multivariable models that include diastolic blood pressure, smoking, body mass index, age, sex and diabetes.

Table 4.1.9. Results of Cox Regression with Time-dependent Variables Representing Changes in Cholesterol Value, Adjusted for CURRENT Cholesterol Value with Adjustment for Incidence of CHD During Follow-up:

Model	Variables	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
5	Current SCL(mg/dL)	0.004345	1.004	(1.003, 1.006)	26.13193	0.0001
	Absolute change from last (mg/dL)	0.001606	1.002	(0.998, 1.005)	0.98454	**0.3211**
6	Current SCL(mg/dL)	0.005026	1.005	(1.003, 1.007)	29.83177	0.0001
	Absolute change from baseline (mg/dL)	-0.001013	0.999	(0.997, 1.001)	0.81026	**0.3680**
7	Current SCL(mg/dL)	0.004600	1.005	(1.003, 1.006)	29.61420	0.0001
	Relative change from last	0.027293	1.028	(0.453, 2.333)	0.00426	**0.9480**
8	Current SCL(mg/dL)	0.004800	1.005	(1.003, 1.007)	27.59953	0.0001
	Relative change from baseline	-0.113156	0.893	(0.524, 1.521)	0.17359	**0.6769**

- All results are from the multivariable models that include diastolic blood pressure, smoking, body mass index, age, sex and diabetes.

4.2 Effects of Smoking

4.2.1 Distributions of time-dependent variables for smoking

The distributions of time-dependent variables representing various aspects of smoking at selected visits are summarized in Table 4.2.1. The first column of Table 4.2.1 shows that the percentage of current smokers in the Framingham dataset gradually decreased with increasing follow-up duration. Two other columns show the cumulative proportion of study participants who either stopped or started smoking at any earlier visit. The cumulative nature of these statistics explains why these two proportions steadily increase over time. It should be noticed that the percentages shown in Table 4.2.1 correspond to proportion of subjects who have the value of 1 for the variables, Sstop and Snew, respectively, which are discussed in section 3.5. The small proportion of subjects with Snew=1 suggests that the power of the tests of the significance of its effect will be low. By contrast, a considerably larger proportion of subjects stopped smoking during the follow-up. This trend is consistent with general trends in North American societies, and explains why the proportion of current smokers decreases.

4.2.2 Comparing the predictive ability of baseline smoking and current smoking status

The goodness-of-fit was compared for models corresponding to baseline and current smoking status (Tables 4.2.2, 4.2.3). The difference between these two tables is that Table 4.2.3 contains the time-dependent variable CHDNEW, which identifies participants who had a non-fatal CHD event earlier during follow-up, and Table 4.2.2 does not. Table

Table 4.2.1 Proportions of Current Smokers, New Smokers and Ex-smokers among the Participants of the Framingham Study at Selected Visits

Years since baseline visit	current smokers (%)	baseline smokers who stopped smoking (%) (a) *	baseline non-smokers who started smoking (%) (b) *
0	57.5%	N/A	N/A
6	54.6%	7.6%	3.9%
12	51.1%	12.8%	5.2%
18	44.2%	19.9%	5.6%
24	38.3%	28.3%	5.9%
28	36.9%	29.9%	6.2%

(*) Cumulative percent, i.e. percent of subjects who stopped (a) or started (b) smoking at any previous visit.

Table 4.2.2. Comparing Models with Baseline versus Current Smoking Status: Testing the Global Null Hypothesis

Smoking Status	Model's Fit (-2LogL)	Criterion	Model Chi-Square
Baseline	**12361.820**	LRT*	745.995 with 7 DF (p=0.0001)
Current	**12371.332**		736.482 with 7 DF (p=0.0001)
Baseline	.	Score	725.416 with 7 DF (p=0.0001)
Current	.		715.070 with 7 DF (p=0.0001)
Baseline	.	Wald	660.088 with 7 DF (p=0.0001)
Current	.		648.574 with 7 DF (p=0.0001)

* LRT = Likelihood Ratio Test

Table 4.2.3. Comparing Models with Baseline versus Current Smoking Status, with adjustment for incidence of CHD during follow-up: Testing the Global Null Hypothesis

Smoking Status	Model's Fit (-2LogL)	Criterion	Model Chi-Square
Baseline	**10346.426**	LRT*	2761.388 with 8 DF (p=0.0001)
Current	**10312.683**		2795.131 with 8 DF (p=0.0001)
Baseline	.	Score	5914.186 with 8 DF (p=0.0001)
Current	.		5955.105 with 8 DF (p=0.0001)
Baseline	.	Wald	1101.393 with 8 DF (p=0.0001)
Current	.		1141.825 with 8 DF (p=0.0001)

* LRT = Likelihood Ratio Test

4.2.2 indicates that both the model with baseline smoking values and the model with current smoking values are very significant using any of these three tests (p-value approximately 0.0001). We also can see that the value of -2 LOG L of the baseline model (12361.820) is considerably lower than that of the current model (12371.332). This result suggests that baseline smoking status is a more powerful predictor than the current smoking status. This suggestion is confirmed by substantially higher values of Score and Wald tests for the model with baseline smoking values included. By contrast, Table 4.2.3 shows that current smoking becomes a much more powerful predictor after adjusting for the occurrence of non-fatal CHD events during follow-up. The reason why adjusting for incident CHD events changes the relative predictive ability of updated versus baseline risk factor values are discussed in section 4.1.4 in the context of cholesterol. The fact that the same pattern of results is found for smoking indicates the general nature of this phenomenon and points to its importance from both the methodological and the practical perspective.

Tables 4.2.4 and 4.2.5 compare the detailed results of the Cox models using baseline smoking indicators with the results of the models using current smoking indicators. Table 4.2.5 presents results adjusted for CHDNEW, while the results in table 4.2.4 are not adjusted. We see that most common cardiovascular risk factors, such as smoking, cholesterol, diastolic blood pressure, glucose intolerance, age and sex, are

Table 4.2.4. Results of Two Multivariable Regression Models: Baseline versus Current Smoking Status

Variables	B/C	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
Smoking	Baseline(a)	0.3373	1.401	(1.160, 1.693)	12.2	**0.0005**
(current/no-smoker)	Current (b)	0.1535	1.166	(0.908, 1.387)	0.1	**0.0831**
Total serum	Baseline	0.0073	1.007	(1.006, 1.009)	71.7	**0.0001**
cholesterol (mg/dL)	Current	0.0073	1.007	(1.006, 1.009)	71.0	**0.0001**
DBP	Baseline	0.0329	1.033	(1.027, 1.040)	97.6	**0.0001**
	Current	0.0324	1.033	(1.026, 1.040)	94.9	**0.0001**
Diabetes	Baseline	0.8246	2.281	(1.741, 2.989)	35.8	**0.0001**
(Yes/No)	Current	0.8338	2.302	(1.757, 3.016)	36.6	**0.0001**
BMI (kg/m^2)	Baseline	0.0309	1.031	(1.010, 1.053)	8.7	**0.0031**
	Current	0.0291	1.029	(1.009, 1.051)	7.7	**0.0054**
Age (years)	Baseline	0.0851	1.089	(1.077,1.101)	240.1	**0.0001**
	Current	0.0837	1.087	(1.076, 1.099)	233.5	**0.0001**
Sex	Baseline	-1.0795	0.340	(0.281, 0.411)	122.3	**0.0001**
(F/M)	Current	-1.1691	0.311	(0.258, 0.374)	154.5	**0.0001**

(a) Baseline denotes Cox model with baseline smoking status

(b) Current denotes Cox model with current (updated) smoking status

Table 4.2.5. Results of Two Multivariable Regression Models: Baseline versus Current Smoking Status, with Adjustment for Incidence of Coronary Heart Disease during Follow-up

Variables	B/C	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
Smoking	Baseline(a)	0.3053	1.357	(1.112, 1.655)	9.1	**0.0026**
(current/no-smoker)	Current (b)	0.5803	1.787	(1.503, 2.124)	43.3	**0.0001**
CHDNEW	Baseline	4.7919	120.530	(85.35, 170.2)	740.3	**0.0001**
	Current	4.8692	130.221	(92.13,184.06)	760.6	**0.0001**
Total serum	Baseline	0.0037	1.004	(1.002, 1.005)	17.7	**0.0001**
cholesterol (mg/dL)	Current	0.0036	1.004	(1.002, 1.005)	17.9	**0.0001**
DBP	Baseline	0.0193	1.019	(1.013, 1.026)	35.7	**0.0001**
	Current	0.0182	1.018	(1.012, 1.025)	31.6	**0.0001**
Diabetes	Baseline	0.7782	2.177	(1.658, 2.859)	31.3	**0.0001**
(Yes/No)	Current	0.7764	2.174	(1.654, 2.856)	31.1	**0.0001**
BMI (kg/m^2)	Baseline	-0.0012	0.999	(0.977, 1.021)	0.01	**0.9100**
	Current	0.0009	1.001	(0.980, 1.023)	0.00	**0.9320**
Age (years)	Baseline	0.0534	1.055	(1.043,1.066)	90.8	**0.0001**
	Current	0.0539	1.055	(1.044, 1.067)	94.4	**0.0001**
Sex	Baseline	-0.6556	0.519	(0.424, 0.635)	40.6	**0.0001**
(F/M)	Current	-0.6235	0.536	(0.445, 0.646)	42.7	**0.0001**

(a) Baseline denotes Cox model with baseline smoking status

(b) Current denotes Cox model with current (updated) smoking status

highly significant in all models. The only exception is BMI, which becomes completely non-significant (p-value >0.9) after adjustment for the incidence of non-fatal CHD events (Table 4.2.5). The fact that, in both Tables 4.2.4 and 4.2.5, the hazard ratios for all covariates are almost the same suggests that the differences between the estimates for the two smoking status variables are unlikely to be confounded by other risk factors. Comparison of the first two rows in Table 4.2.4 shows that baseline smoking has a higher impact on CHD mortality than current smoking (HR=1.401 Vs HR=1.166). In fact, current smoking has a marginally non-significant effect (p=0.0831 in Table 4.2.4). However, Table 4.2.5 indicates that after adding the new time-dependent variable CHDNEW, the estimated hazard ratio for current smoking (1.787) becomes much larger than that for baseline smoking (1.357). A substantial increase in the estimated effect of current smoking, observed after having adjusted for CHDNEW, demonstrates that its effect in Table 4.2.4 is confounded due to a failure to take into account changes in subjects' health status that occur during follow-up. A likely mechanism is that many subjects stop smoking after having a non-fatal CHD event that acts as a "warning signal".

4.2.3 Assessing predictive ability of time-dependent variables representing changes in smoking status

Tables 4.2.6 and 4.2.7 summarize the estimated effects of time-dependent variables related to the start of smoking (Snew) and smoking cessation (Sstop) during the follow-up. The upper part of each Table (models 1-3) presents estimates adjusted for the baseline smoking status, whereas the lower part presents that adjusted for the current

Table 4.2.6. Estimated effects of time-dependent variables representing changes in smoking status, adjusted for Baseline or Current smoking status:

Model	Variables	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
1	Baseline smoking	0.3025	1.353	(1.106, 1.656)	8.6	0.0033
	Sstop	0.1111	1.118	(0.898, 1.390)	1.0	**0.3187**
2	Baseline smoking	0.3827	1.466	(1.205, 1.785)	14.6	0.0001
	Snew	0.3970	1.487	(0.978, 2.261)	3.4	0.0633
3	Baseline smoking	0.3480	1.416	(1.149, 1.745)	10.7	0.0011
	Sstop	0.1107	1.117	(0.898, 1.390)	1.0	**0.3207**
	Snew	0.3965	1.487	(0.978, 2.260)	3.4	0.0636
4	Current smoking	0.2555	1.291	(1.073, 1.554)	7.3	0.0069
	Sstop	0.3480	1.416	(1.136, 1.766)	9.5	0.0020
5	Current smoking	0.1505	1.162	(0.977, 1.383)	2.9	0.0894
	Snew	0.1555	1.168	(0.782, 1.746)	0.6	**0.4478**
6	Current smoking	0.2536	1.289	(1.071, 1.551)	7.2	0.0073
	Sstop	0.3566	1.429	(1.145, 1.783)	10.0	0.0016
	Snew	0.2053	1.228	(0.820, 1.838)	1.0	**0.3184**

- All results are from the multivariable models that include total serum cholesterol, diastolic blood pressure, body mass index, age, sex and diabetes.

Table 4.2.7. Estimated Effects of Time-dependent Variables Representing Changes in Smoking Status, Adjusted for Baseline and Current Smoking Status with Adjustment for Incidence of Coronary Heart Disease During Follow-up (CHDNEW)

Model	Variables	Parameter Estimate	Hazard Ratio	95% CI	Wald χ^2_w	P-value
1	Baseline smoking	0.5876	1.800	(1.457, 2.222)	29.8	0.0001
	Sstop	-0.7122	0.491	(0.394, 0.611)	40.3	0.0001
2	Baseline smoking	0.2963	1.345	(1.093, 1.655)	7.8	0.0051
	Snew	-0.0623	0.940	(0.618, 1.429)	0.1	0.7707
3	Baseline smoking	0.5782	1.783	(1.432, 2.220)	26.8	0.0001
	Sstop	-0.7123	0.490	(0.394, 0.611)	40.4	0.0001
	Snew	-0.0650	0.937	(0.616, 1.425)	0.1	0.7610
4	Current smoking	0.4932	1.638	(1.352, 1.983)	25.4	0.0001
	Sstop	-0.2410	0.786	(0.626, 0.987)	4.3	0.0385
5	Current smoking	0.5914	1.806	(1.519, 2.149)	44.6	0.0001
	Snew	-0.3171	0.728	(0.487, 1.090)	2.4	0.1233
6	Current smoking	0.4976	1.645	(1.357, 1.994)	25.7	0.0001
	Sstop	-0.2596	0.771	(0.613, 0.970)	4.9	0.0263
	Snew	-0.3548	0.701	(0.468, 1.050)	3.0	0.0852

- All results are from the multivariable models that include total serum cholesterol, diastolic blood pressure, body mass index, age, sex and diabetes.

status. The difference between the two tables is that the results in Table 4.2.7 are additionally adjusted for the incidence of non-fatal CHD events during follow-up (CHDNEW).

Table 4.2.6 shows that the estimated effect of smoking cessation on CHD mortality changes substantially depending on whether it is adjusted for baseline or current status. Surprisingly, once baseline smoking status is taken into account, smoking cessation during the follow-up does not appear to be significantly associated with the hazard of CHD mortality (models 1 and 3). By contrast, among current non-smokers, those who stopped smoking during the follow-up are at significantly higher risks than those who did not smoke at baseline (models 4 and 6). It is not entirely clear how this finding should be interpreted. One possibility would be that it demonstrates that negative impact of previous exposure to tobacco persists even several years after smoking cessation. An alternative interpretation concerns confounding by incident non-fatal CHD events. By the argument similar to that discussed in section 4.2.2, one may expect that smoking cessation will be more frequent among participants who had a "warning signal" (i.e. a non-fatal CHD event). Since these participants will be at considerably higher risk of CHD death, a failure to account for these events will likely bias the estimate of the "Sstop" variable away from the null. The same argument will explain why this variable appears to be non-significant in models that adjust for the baseline smoking status. A similar phenomenon may partly explain why the "Snew" variable is almost statistically significant and has estimated hazard ratios similar to those for baseline status in the models that

adjust for the baseline (models 2 and 3 in Table 4.2.6). While some of this effect may represent an increase in risks due to newly started smoking habit. it may be largely due to the fact that subjects who had a non-fatal CHD event will be quite unlikely to start smoking thereafter.

The above discussion of results presented in Table 4.2.6 indicates that their interpretation is difficult because of the failure to account for the incidence of non-fatal CHD events during the follow-up. The estimated effects of the time-dependent smoking-related variables presented in Table 4.2.7 are adjusted for these events. This adjustment is represented by the time-dependent variable CHDNEW. Comparisons with the results in Table 4.2.6 show several important differences between the two.

The first difference between the result is the estimated effects of both baseline and current smoking status, adjusted for different aspects of changes in this status, become much stronger after incident CHD events are taken into account.

Secondly, a very different picture emerges with respect to the effects of changes in the smoking status. Among subjects who were smokers at the baseline visit, smoking cessation is associated with a 50% reduction of risks of CHD death (hazard ratios of about 0.49 for Sstop variable in models 1 and 3 in Table 4.2.7). In fact, the estimated risk for those who stopped smoking during the follow-up (1.8 * 0.491= 0.88 for model 1) seems to be slightly lower than for subjects who

never smoked during the study period. This result is further corroborated by models 4 and 6 which show a significant risk reduction associated with smoking cessation, at an earlier visit but after the baseline, among subjects who currently do not smoke (p-values of 0.0385 and 0.0263, respectively, for Sstop variable). Such an association may be partly due to residual confounding by some other, simultaneous in time changes in the subjects' CHD risk profile that are not accounted for in our analyses. It is possible that participants who decided to stop smoking changed other aspects of their lifestyle, diet etc. at the same time. Together, these changes may have resulted in reduction of the risk beyond the reduction due to smoking cessation. If so, in the models that fail to account for these additional changes in relevant risk factors, their effects will be attributed to the simultaneous smoking cessation.

Finally, Table 4.2.7 shows that after having adjusted for incident non-fatal CHD events, beginning of smoking during the follow-up (represented by the Snew variable) has virtually no effect on the risks of CHD death among subjects who were non-smokers at baseline. The point estimate for Snew variable in models 2 and 3 are so close to 1.0 that it is clear that the non-significance of its effect is not merely due to low statistical power. The reasons for the lack of difference in risks between subjects who started smoking later during the follow-up and those who never smoked are not obvious and require further analyses. By contrast, the effect of Snew becomes almost significant in the models 5 and 6 that adjust for the current rather than baseline smoking status. These results suggest that, among

current smokers, those who started smoking only after the baseline visit may have lower risks of CHD mortality. Again, this finding may be interpreted in two different ways and results presented here do not allow discriminating between the two alternatives. One possibility is that the apparently protective effect of Snew reflects the fact that a person who decided to start smoking during the follow-up is very unlikely to have experienced any problems related to CHD, and so the start of the smoking acts as a marker of very low CHD risk. Another possibility is that those who start smoking later in life smoke less on average, in terms of the number of cigarettes per day, than long-time smokers. Since in my analyses the information on smoking is limited to binary variables, smoking intensity could not be taken into account. Finally, it is possible that those who have smoked for a long time experience cumulative effects of damage due to long term exposure. However, the last explanation does not seem to fit well with the findings that smoking cessation brings the CHD risks back to the level observed in non-smokers within a few years.

4.3 Assessing Changes Over Time in the Predictive Ability of Baseline Values of Selected Coronary Heart Disease risk factors

The Flexible regression spline model (Abrahamowicz et al, 1996) was employed to test the proportional hazards (PH) assumption for baseline measurements of each of three modifiable CHD risk factors: total serum cholesterol, diastolic blood pressure (DBP) and smoking status. Due to the

limitations of the currently available version of the C program that implements this method, all flexible analyses were restricted to the 2311 male participants of the Framingham study. These participants accounted for 45% of the total study sample. There were 391 CHD deaths in this sub-population. The PH hypothesis was tested separately for each of the three risk factors. In each analysis, its effect was adjusted for all other common risk factors, which were represented by constant-in-time hazard ratios. The likelihood ratio test (LRT) on 4 degrees of freedom (df) was used to compare the fit of the conventional PH model with the model in which the effect of a given risk factor was allowed to vary over time, and it was represented by a 5-df quadratic spline. These LRT tests failed to reject the PH hypothesis for any of the three risk factors, with p-values of 0.7842, 0.4907 and 0.1504 for cholesterol, DBP and smoking respectively. These high p-values provide strong support for the null hypothesis that the effects of baseline measurements of each of the three risk factors on CHD mortality remain stable during 30 years of follow-up.

Figures 1, 2 and 3 represent estimated hazard ratio functions for the three risk factors, and their corresponding 95% pointwise confidence intervals (Abrahamowicz et al 1996). The solid curve in each figure shows how the point estimate of the log hazard ratio varies with increasing follow-up time. The dotted curves around the point estimate correspond to the 95% confidence intervals and allow the assessment of the precision of the estimate at a given point in time. The confidence intervals are particularly wide in the first few years because few deaths occurred in the initial phase of follow-up. The fact that the hazard ratio functions

representing the estimated effects of total serum cholesterol and DBP are very flat further confirms that baseline values of both risk factors maintain their predictive ability over 30 years of follow-up. Figure 3 suggests that the predictive ability of baseline smoking status may tend to decrease with increasing follow-up time but the estimated change is too small to reach statistical significance. Accordingly, the graph shows that the constant function, consistent with the PH assumption, would fit within the 95% confidence intervals.

Figure 1: Time-Dependence of the Effect of Baseline Total Cholesterol

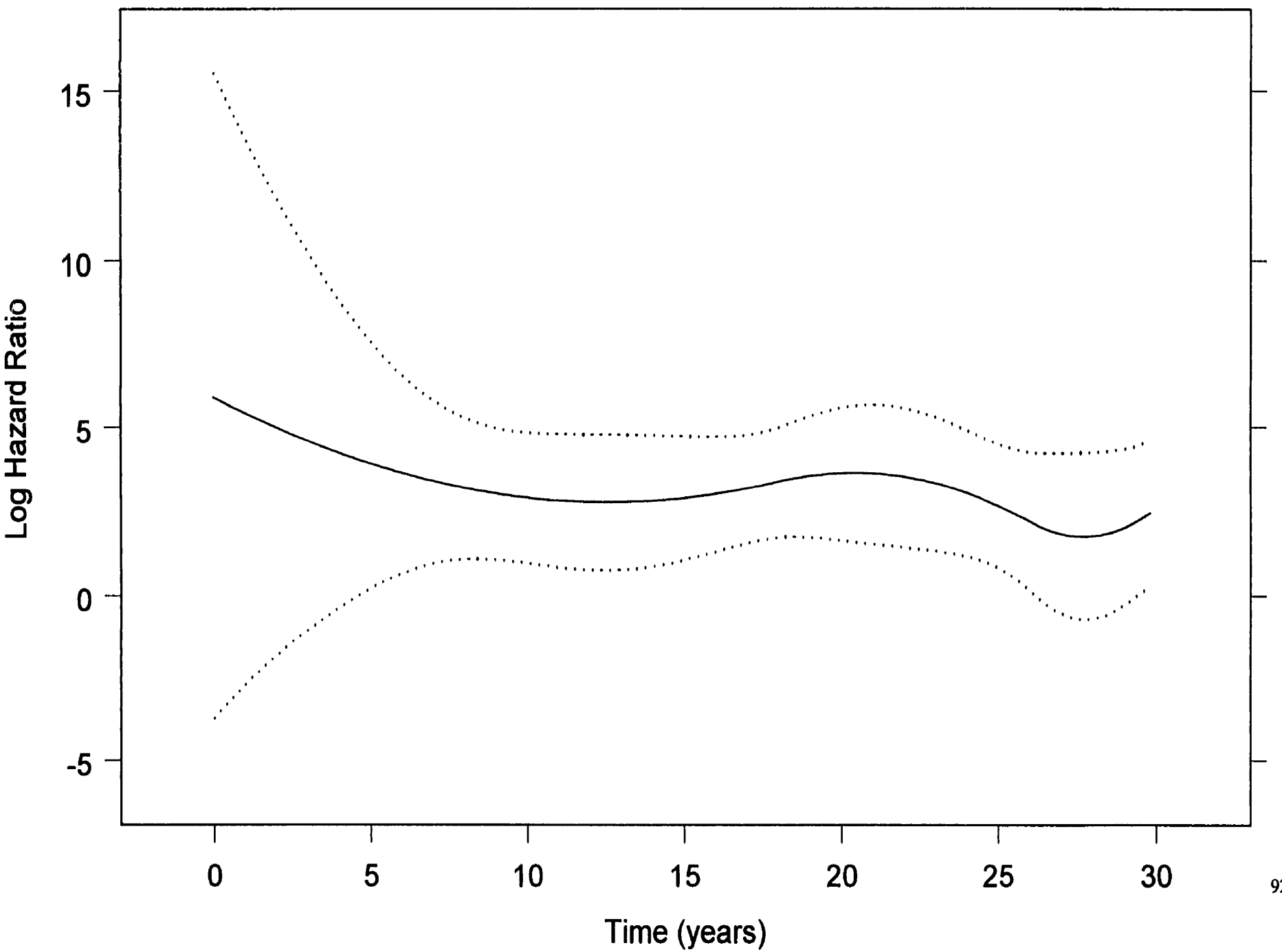


Figure 2: Time-Dependence of the Effect of Baseline Diastolic Blood Pressure

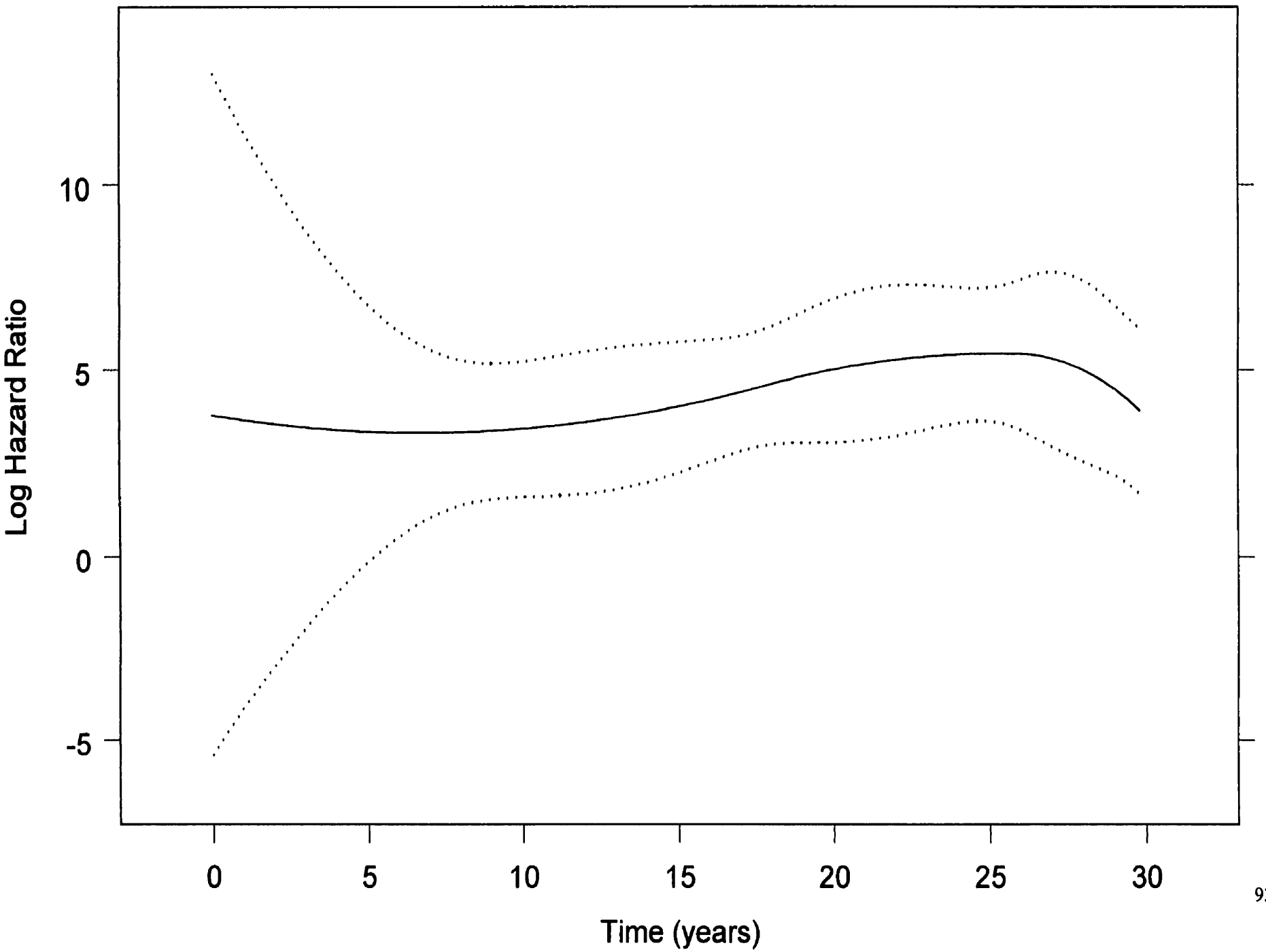
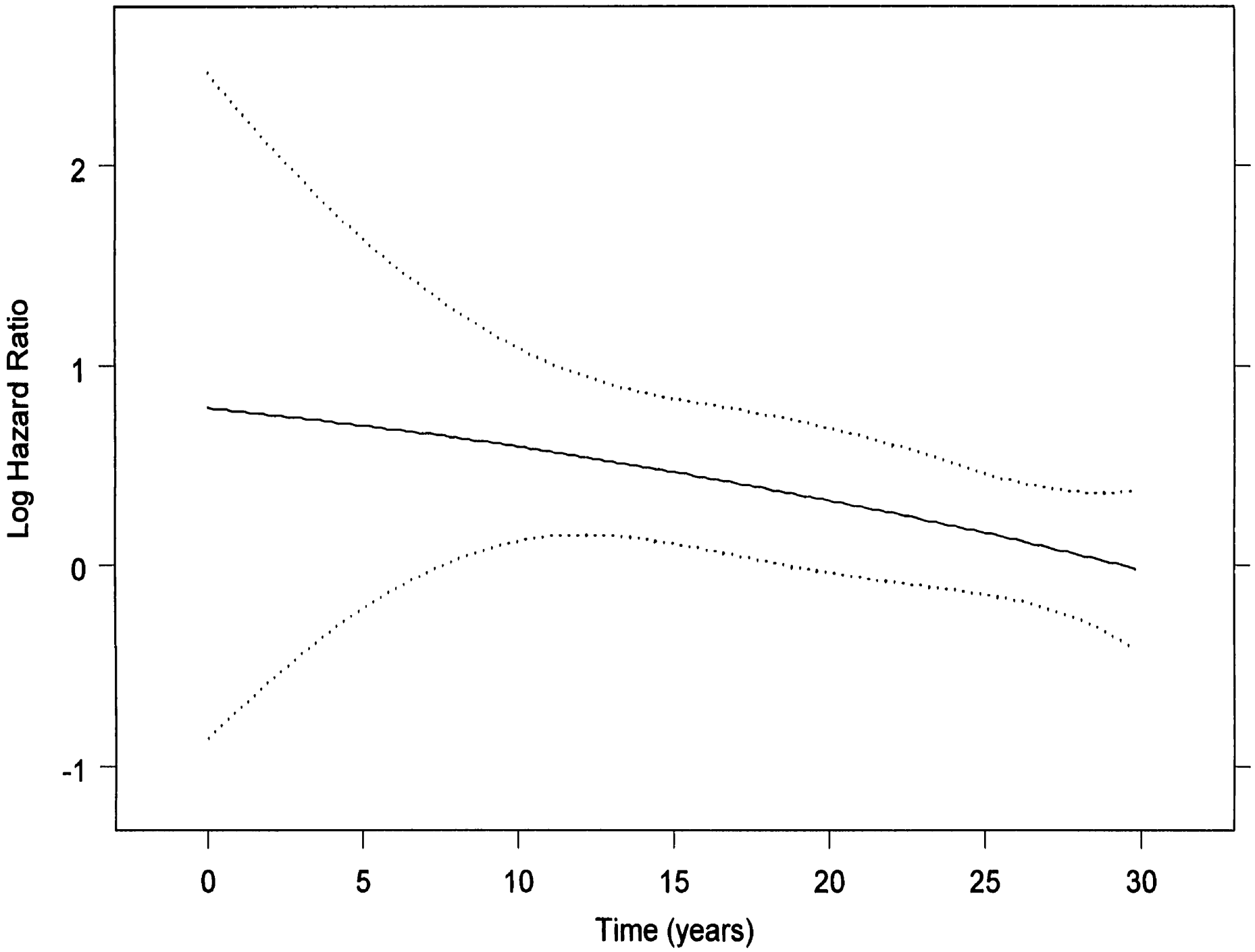


Figure 3: Time-Dependence of the Effect of Baseline Smoking Status



Chapter 5

Discussion

In this thesis, I have adopted a new approach for the analysis of the impact of changes in risk factors on the coronary heart disease (CHD) mortality. The approach relies on the inclusion of time-dependent covariates in the Cox model. These covariates represent various aspects of longitudinal changes in risk factor values. This representation allows for an assessment of the role of within-subject variation in coronary risk profiles, and thus provides more direct estimates of the expected effects of various interventions aimed at risk factor modification than the results in previous analyses that relied on between-subjects comparisons. However, the previous chapter shows to what extent the results of these time-dependent analyses depend on the detailed specification of the regression model. This poses new challenges for a data analyst, both with respect to choice of the appropriate model, and for interpretation of results. To address these challenges, Chapters 3 and 4 contain some methodological comments that may help to guide researchers in their future attempts to evaluate the impact of changes in risk factors for CHD or other diseases. In addition to contributing to the methodology of complex modeling of within-subject variation in risk factors, application of these new models to Framingham data yielded new insights into the role of two major modifiable CHD risk factors: total serum cholesterol and smoking status.

In what follows, I briefly summarize the main findings of this study, separating methodological issues from results of more substantive interest. However, when discussing how the results of this thesis relate to previously published analyses of the effects of changes in CHD risk factors, I discuss the possibility that the differences may be at least partly due to differences in respective modeling approaches. I then conclude the thesis with a discussion of limitations of my analyses, and with some recommendations for future research on the modeling the impact of CHD risk factors modifications.

The methodological contributions to this thesis fall into three categories. First, I have developed operational definitions of several time-dependent variables, which are related to various aspects of changes occurring in a risk factor value during the follow-up period, and I have specified the appropriate regression models to assess their effects. Using total serum cholesterol as an example of a modifiable risk factor measured on a continuous scale, I have constructed time-dependent variables representing absolute as well as relative change either from the baseline value or from the most recent value. As discussed below, I have then demonstrated that the interpretation of results depends substantially on what specific measure of change is employed and what other variable(s) it is adjusted for in a multivariable regression model. A binary risk factor, such as smoking status, requires a somewhat different approach and I have constructed two new time-dependent variables that identify subjects who, respectively, stopped smoking and started smoking at an earlier follow-up visit but after the baseline evaluation. Again, I have shown that interpretation of the estimated effects of these new variables depends on

what other variables are included in the model and suggested some models that may offer useful insights as to the impact of smoking cessation.

The second methodological issue addressed in this thesis concerns comparison of the predictive ability of the baseline value with that of an updated value of a risk factor. This issue is of practical importance, as there is some controversy in the epidemiological literature on CHD regarding which of the two values has stronger impact on the risks of subsequent mortality and morbidity (Benfante et al, 1994; Cupples et al, 1988). Due to near-collinearity of the baseline and updated values of the same risk factor, simultaneous inclusion of variables representing both values in the same model would produce unstable estimates. To avoid these problems, I propose to compare the goodness of fit of two separate models that include either the baseline or the updated variable, and demonstrate the feasibility of this approach in the context of modeling the effects of serum cholesterol and smoking on CHD mortality in the Framingham Heart Study. Furthermore, I employ a recently developed regression spline generalization of the conventional Cox model (Abrahamowicz et al 1996) to assess the stability over time of the prognostic utility of the baseline values of modifiable CHD risk factors.

Whereas the two issues discussed above correspond to the original objectives of this thesis, arguably the most interesting and challenging methodological problem was in fact identified only a posteriori, based on an unexpected pattern of results of some analyses involving newly constructed time-dependent variables. The fact that, contrary to expectations, post-baseline changes in both serum cholesterol values and smoking status

appeared to have no effect on subsequent CHD mortality suggested that these results may be partly confounded by some subjects' characteristics that have not been taken into account. To investigate this possibility, a series of additional, post hoc analyses, was carried out with a new time-dependent binary variable, which identified subjects who had non-fatal CHD events earlier during the follow-up but after the baseline visit. The reason for the inclusion of this additional variable in the multivariable models was partly based on the findings of the Honolulu Heart Study. This study reported that during 25 years of follow-up, the greatest changes in risk factors occurred in those participants who had developed various clinical manifestations of cardiovascular disease (Benfante et al, 1994).

Adjusting for incident CHD events dramatically changed many results related to the effects of various cholesterol- and smoking-related variables and, in particular, revealed statistically significant benefits of lowering total serum cholesterol and of smoking cessation. These results were unlike those from original analyses where the effects of both variables were definitely non-significant after the baseline cholesterol and smoking status were taken into account. Indeed, a closer examination of the relevant data indicated that the CHDNEW variable, representing previous incidence of non-fatal CHD events during the follow-up period, met statistical criteria for a confounder of the effects of time-dependent changes in cholesterol and/or smoking status. First, this variable was obviously strongly associated with the outcome, as subjects who had a non-fatal CHD event during the study period had much higher risks of CHD deaths than those who had the same baseline risk profile but remained free of CHD disease. Second, CHDNEW was

also associated with the time-dependent variables representing changes in risk factors; subjects who had had a non-fatal CHD event were more likely to attempt to decrease their cholesterol level and/or stop smoking afterwards than subjects who did not have such a "warning signal".

The above discussion suggests that the analyses reported in this thesis provide an interesting and compelling example of confounding occurring at the level of time-dependent covariates in survival analysis; the time-dependent variable CHDNEW acts as a confounder for time-dependent measures of changes in serum cholesterol and smoking status. The results of these analyses provide convincing evidence that such a confounding may result in an important alternative for the conclusions. This alternation is of particular relevance given that it calls for adjusting for relevant time-dependent covariate(s), even if such covariates may appear to represent "intermediate outcomes" i.e. events that lay on a pathway leading from the baseline risk factors to the final outcome. Whereas adjusting for intermediate outcomes seems to go against one of the fundamental principles of deciding what variables should be included as potential confounders in multivariable models, the very principle needs some qualification in the specific case of simultaneous modeling of the effects of time-dependent and baseline risk factors. Specifically, CHDNEW is on a pathway between baseline cholesterol and/or smoking and CHD death, but it is not on a pathway between changes in these risk factors that occur after a non-fatal event, perhaps partly because of such an event. In other words, adjusting an updated value of a time-dependent change variable for an intermediate outcome that occurred after the baseline but before the updated measurement respects the temporal sequence of

relevant events. At the time the updated risk factor value was obtained, the non-fatal CHD event has already been observed and the change might have been partly caused or provoked by this event. Thus, adjusting for the time-dependent CHDNEW variable is appropriate and, indeed, necessary if the main focus of the analysis is on the estimation and significance testing of the effects of time-dependent variables representing longitudinal within-subject changes in relevant risk factors. Identification of methodologically interesting and practically important phenomenon of confounding between different time-dependent variables may help other researchers, interested in the impact of within-subject changes in different risk factors, treatments, and environmental exposures on a variety of health outcomes.

In addition to addressing the above methodological issues, this study provides new insights into some substantive aspects of the epidemiology of coronary heart disease. Whereas different authors made various conjectures regarding the relative importance of baseline versus updated CHD risk factors, there is little empirical evidence related to these conjectures. Specifically, Benfante et al. (1994) argued that among men who began follow-up in their mid-life, initial values of risk factors may be more important for the assessment of current CHD risks than updated, recently measured values. The reason for such a conjecture was the authors' belief that artheriosclerotic changes provoked by high mid-life risk factor values would be unlikely to be reversed by risk profile modification later in one's life. If correct, this conjecture might have major impact on the evaluation of expected effectiveness and cost-effectiveness of various, often expensive, interventions aimed at risk factor modifications, and suggests that such modifications might have little

effect of the actual CHD risks, especially among older subjects. By contrast, Cupples et al. (1988) and Abrahamowicz et al. (1997) have suggested, independently of each other, that with increasing follow-up duration the baseline risk factor values will gradually become less important for predicting CHD mortality than the updated measures of the same factors. However, none of these contradictory conjectures was supported by empirical data. Results reported in this thesis provide some relevant evidence, though the picture is not completely clear.

Firstly, flexible regression spline analyses clearly showed that baseline measurements of serum cholesterol and blood pressure maintain their predictive ability at a constant level during almost 30 years of follow-up time. While the impact of baseline smoking may have a tendency to decrease with increasing follow-up, the resulting changes in its prognostic ability are too small to reach statistical significance. However, the fact that initial values of risk factors maintain significant associations with CHD mortality risks even several decades after these measurements were taken does not necessarily indicate that up-dating these initial values is of no prognostic utility (Lewis et al, 1997).

Direct comparisons of the prognostic ability of the baseline and updated values of the risk factors of interest revealed the strong dependence of the conclusions on the modeling strategy. If all other variables included in the multivariable model were limited to baseline measurements, then baseline cholesterol (or smoking status) turned out to

predict CHD deaths better than the updated values of the same variables. Thus, conventional analyses seem to corroborate the Benfante et al's (1994) conjecture.

However, the updated risk factor values have stronger impact on the CHD mortality than the corresponding baseline values once the incidence of non-fatal CHD events during the follow-up has been represented by, and adjusted for an appropriate time-dependent covariate. Thus, among subjects who, at a given point in time have the same CHD disease status (symptomatic or not) and who had the same baseline values of all other risk factors, the updated serum cholesterol value is a more powerful predictor of subsequent CHD fatality than the baseline cholesterol value. The same applies to smoking status. Accordingly, while prediction based on the baseline measurements is quite robust, it is less powerful than prediction based on updated risk factor values, once current CHD status is taken into account.

The above result contradicts at least partly Benfante et al's (1994) conjecture and provides empirical support to guidelines that call for risk factor modifications even for individuals of advanced age. Indeed, after having adjusted for the previous incidence of non-fatal CHD, as well as for baseline cholesterol, the amount of change from the baseline level of serum cholesterol became a significant predictor of subsequent CHD mortality, with a bigger decrease being associated with lower risks.

This finding may contribute new insights into the biological and behavioral mechanisms by which changes in cholesterol can influence CHD. It is in an apparent

contrast to the results of a limited number of published analyses that investigated the role of within-subject changes in serum cholesterol. For example, the authors of the Finnish Cohorts of the Seven Countries Study (Pekkanen et al, 1994) concluded that greater than average declines in the serum cholesterol level were associated with increased mortality from CHD, especially among the older participants of the cohort. A similar conclusion was reached by Anderson et al. (1987), who re-analyzed data on 30 years of follow-up from the Framingham Study and found that a decrease in the serum cholesterol level was strongly associated with increased cardiovascular mortality. The authors of these two studies suggested two possible explanations of their findings. First, a spontaneously occurring decrease in serum cholesterol could by itself cause increased CHD mortality. Alternatively, the decrease in serum cholesterol level and the associated high CHD risk could both be caused by one or several other factors. Thus, the first explanation postulates an unknown cause-effect mechanism, while the other attributes the apparent association of a decline in serum cholesterol with increased CHD mortality to confounding by unknown, or un-accounted for, factors, that may influence both.

The results of this thesis seem to corroborate the latter explanation, by pointing out to an observable phenomenon that may be responsible for such a confounding. The estimated effect of changes in serum cholesterol between the baseline level and the level observed in a given point in time t, varies very substantially depending on whether it is adjusted or not for the incidence of non-fatal CHD that occurred before time t. Thus, from a statistical perspective, a time-dependent binary indicator of previous post-baseline non-fatal CHD event(s) acts in fact as a confounder for the association between time-

dependent changes in serum cholesterol and subsequent CHD fatality. This confounding can be explained by two mechanisms, one of which concerns mostly biological and other behavioral aspects of CHD. First, non-fatal CHD event is obviously a very powerful predictor of a subsequent CHD death. Second, one may postulate that a non-fatal CHD event acts likely as a "warning signal" and, therefore, motivates at least some subjects to undertake some preventive measures, one of which may be to lower the values of modifiable risk factors. This phenomenon is consistent with the report by Benfante et al. (1994), who found that among the participants of the Honolulu Heart Study, the biggest changes in risk factors occurred in those individuals who developed CHD at some point during the follow-up.

Thus, the above discussion suggests a new explanation of the apparent association between decreasing serum cholesterol levels and increased risks of CHD death. Subjects who developed CHD at some point during the study will be more likely to both undertake attempts to lower their cholesterol, and die of CHD later on. To avoid the resulting confounding, it is essential to accurately represent the temporal sequence of relevant changes in the subjects' health, i.e. to take into account that non-fatal CHD precedes subsequent reduction of serum cholesterol. The approach proposed in this thesis, based on properly constructed time-dependent covariates, enables such representation and, therefore, is recommended for further studies of the impact of changes in risk factors on CHD mortality.

The time-dependent analyses reported in the section 4.2 confirmed previous findings regarding the benefits of smoking cessation. Shaten et al. (1991) showed that among nonsmokers, rates of CHD death for former smokers and never smokers were similar (2.51 and 2.21 deaths per 1,000 person-years, respectively). Former smokers who quit within 12 months of randomization had higher rates of CHD death than those who quit 12 or more months before randomization (3.34 and 2.27 deaths per thousand person-years, respectively). Other studies show that smoking cessation results in a dramatic reduction in the risk of mortality from both coronary heart disease and stroke, and would increase population-wide life expectancy by about a year and the life expectancy of a smoker by several years. (Lakier, 1992; Tsevat, 1992).

Some limitations of the present study have to be taken into account when interpreting its results. First, while most common CHD risk factors were taken into account, the high-density lipids (HDL) cholesterol values were not included in the multivariable analyses. The reason for this was that in the Framingham study HDL levels were measured only once, several years after the baseline visit, so that no data on within-subject changes in this important risk factor were available. Secondly, I considered only a subset of a potentially large set of various time-dependent variables representing different aspects of longitudinal changes in risk factors. Another limitation is that, in order to facilitate the interpretation of analyses involving time-dependent variables, no interactions between risk factors were considered. Further research should investigate such questions as whether occurrence of a non-fatal CHD event changes the impact of blood pressure or cholesterol or CHD mortality. In this thesis, I analyzed both absolute

and relative measures of change from either the baseline or the last measurement of a relevant variable. Yet, other variables of potential interest might represent some cumulative aspects of changes occurring over a longer time interval. One such variable may be the estimated within-subject slope from the linear model regressing, for example, the current cholesterol value over time-since-baseline. Another possibility would be to construct a cumulative measure of reduction in cholesterol based on the idea of area-under-the-curve. Yet another variable of interest could be defined as a weighted version of the area-under-the-curve, with weights reflecting the regency of the measurements. One reason that these more complex measures have not been employed in this thesis is that their interpretation may pose additional challenges. Therefore, it was deemed more appropriate to restrict present analyses to simpler measures with the expectation that their results will guide the researchers about how to design and interpret more complex analyses in future.

In summary, the approach proposed in this thesis, based on time-dependent measures of various aspects of within-subject changes in modifiable risk factors, may yield new insights into the role of different preventive interventions but, at the same time, poses new analytical challenges. This opens new venues for future research on the epidemiology of coronary heart disease. Results reported in this thesis should be replicated in a similar analysis of an independent dataset. Then, more complex time-dependent analyses should be undertaken. These future analyses should include new measures that will more accurately represent cumulative effects of long-term patterns of change, as well as investigate possible temporal lags in these effects. Finally,

simultaneous changes in several risk factors, with possible adjustment for intermediate changes in the subjects' health status may be necessary to provide a more comprehensive outlook of complex mechanisms linking risk factors with final clinical outcomes. In this thesis I have made an attempt to make a step in this direction.

Reference:

Abbott RD. Sharp DS. Burchfiel CM. Curb JD. Rodriguez BL. Hakim AA. Yano K. Cross-sectional and longitudinal changes in total and high-density-lipoprotein cholesterol levels over a 20-year period in elderly men: the Honolulu Heart Program. *Annals of Epidemiology.* 1997;7(6): 417-24.

Abrahamowicz Michal, Todd MacKenzie, and John M Esdaile. Time-dependent hazard ratio: Modeling and Hypothesis testing with application in Lupus Nephritis. *Journal of the American Statistical Association.* 1996; 91: 436.

Abrahamowicz Michal, Roxane du Berger, and Steven A. Grover. Flexible Modeling of the Effects of Serum Cholesterol on Coronary Heart Disease Mortality. *American Journal of Epidemiology.* 1997; 145: 714-29.

Altman DG, De Stavola BL, Love SB and Stepniewska KA. Review of survival analyses published in cancer journals. *British Journal of Cancer.* 1995; 72(2): 511-8.

Anderson KM, Castelli WP, Levy D. Cholesterol and mortality, 30 years of follow-up from the Framingham Study. *Journal of American Medical Association.* 1987; 268: 2045-9.

Anderson KM, Wilson PWF, Odell PM, Kannel WB. An Updated Coronary Risk Profile. A Statement for Health Professionals. *Circulation.*1991; 83:356-362.

Austin MA. Plasma triglyceride and coronary heart disease. *Arteriosclerosis, Thrombosis & Vascular Biology.* 1991; 11:2-14.

Benfante R. Reed D. Is elevated serum cholesterol level a risk factor for coronary heart disease in the elderly? *Journal of American Medical Association.* 1990; 263(3): 393-6.

Benfante R, Hwanj LJ, Masaki K, Curb JD. To what extent do cardiovascular risk factor values measured in elderly men represent their midlife value measures 25 years earlier? A preliminary

report and commentary from the Honolulu Heart Program. *American Journal of Epidemiology*.1994; 140 (3): 206-216.

Brand J, Van BS, Van ME, Timmers T, Gelsema E. Multiple imputation as a missing data machine. *Proceedings - the Annual Symposium on computer applications in Medical care*. 1994; 303-306.

Brick JM, Kalton G. Handling missing data in survey research. *Statistical methods in medical research*. 1996; 5(3): 215-38.

Burchfiel CM, Laws A, Benfante R, Goldberg RJ, Hwang L, Chiu D, Rodriguez B, Curb JD, Sharp DS. Combined effects of HDL cholesterol, triglyceride, and total cholesterol concentrations on 18-year risk of atherosclerotic disease. *Circulation*. 1995; 92: 1430-1436.

Carlson LA, Danielson M, Ekberg I, Klintemar B, Rosenhamer G. Reduction of myocardial reinfarction by the combined treatment with clofibrate and nicotinic acid. *Atherosclerosis*.1977; 28: 81-86.

Chen Z; Peto R; Collins R; MacMahon S; Liu J; Li W. Serum Cholesterol concentration and coronary heart disease in population with low cholesterol concentration. *British Medical Journal*. 1991; 303:276-282.

Committee of Principal Investigators, W.H.O. Clofibrate Trial. A cooperative trial in the primary prevention of ischaemic heart disease using clofibrate, report. *British Heart Journal*. 1978; 40: 1049-1118.

Coronary Drug Project Research Group. Clofibrate and niacin in coronary heart disease. *Journal of the American Medical Association*.1975; 231: 360-81.

Côté R, Battista RN, Abrahamowicz M. Langlois Y, Bourque F, Mackey A. Lack of effect of aspirin in asymptotic patients with carotid bruits and substantial carotid narrowing. The Asymptotic Cervical Bruit Study Group. *Annals of Internal Medicine*. 1995; 123 (3): 649-655.

Cox, DR. Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society*. 1972; 34: 187-220.

Cox, DR. Partial Likelihood. *Biometrika*, 1975; 62: 269-276.

Cox, DR, and Oakes, D. (1984), *Analysis of the Survival Data*, London: Chapman and Hall.

CPRRC: Cardiac Prevention and Rehabilitation Research Center.

Crowley J and M Hu. Covariance analysis of heart transplant data. *Journal of American Statistical Association*. 1977; 72: 27-36.

Culp K. Flanigan M. Lowrie EG. Lew N. Zimmerman B. Modeling mortality risk in hemodialysis patients using laboratory values as time-dependent covariates. *American Journal of Kidney Diseases*. 1996; 28(5): 741-6.

Cupples LA, D'Agostino RB, Anderson K, Kennel W. Comparison of baseline and repeated measure covariate techniques in the Framingham Heart Study. *Statistics in Medicine*. 1988: 7: 205-218.

C++ program design: an introduction to programming and object-oriented design: Publisher: Chicago: Irwin, 1997.

Dahlen GH. Lp(a) lipoprotein in cardiovascular disease. *Atherosclerosis*. 1994; 108:111-126.

Dawber, TR. The Framingham Study: the Epidemiology of atherosclerotic Disease, *Harvard University Press*, Cambridge, Massachusetts, 1980.

D'Agostino, R.B. and Kannel, W.B. Epidemiological background and design: the Framingham Study, *in Proceedings of the American Statistical Association Sesquicentennial Invited Paper Sessions. American Statistical Association*, Alexandria, Virginia, 1989; 707-718.

D'Agostino RB. Lee ML. Belanger AJ. Cupples LA. Anderson K. Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Statistics in Medicine*. 1990; 9(12): 1501-15.

Efron, B. Logistic Regression, Survival Analysis and the Kaplan-Meier Curve. *Journal of the American Statistical Association*.1988; 83: 414-425.

Esdaile JM, Abrahamowicz M, MacKenzie T, Hayslett JP, Kashgarian M. The time-dependence of long-term prediction in lupus nephritis. *Arthritis and Rheumatism*.1994; 37(3): 359-368.

Fact Book: Fiscal Year 1988. Bethesda, Md. *National Heart, Lung and Blood Institute*.

Feinleib M. The Framingham Study: sample selection, follow-up, and methods of analyses. *National Cancer Institute Monographs*. 1985; 67:59-64.

Frayers PM, Curran D, Machin D. Incomplete quality of life data in randomized trials: missing items. *Statistics in Medicine*. 1998; 17(5-7): 679-96.

Freund KM. D'Agostino RB, Belanger AJ, Kannel WB, Stokes J. Predictors of smoking cessation: the Framingham Study. *American Journal of Epidemiology*. 1992; 135(9):957-64.

Frick MH, Elo E, Haapa K, Heinonen OP, Heinsalmi P, Helo P, Huttunen JK, Kaitaniemi P, Koskinen P, Manninen V et al. Helsinki Heart Study: primary prevention trial with gemfibrozil in middle-aged men with dyslipidemia: safety of treatment, changes in risk factors, and incidence of coronary heart disease. *New England Journal of Medicine*.1987; 317: 1237-1245.

Friedman GD. Kannel WB. Dawber TR. McNamara PM. An evaluation of follow-up methods e: the Framingham Heart Study. *American Journal of Public Health & the Nations Health*. 1967; 57(6): 1015-24.

Gordon DJ. Cholesterol lowering and total mortality. In: Rifkind BM, ed. *Lowing Cholesterol in High Risk Individuals and Populations*. New York, NY: Marcel Dekker; 1995: 33-47.

Gould AL, Rossouw JE, Santanello NC, Heyse JF, Furberg CD. Cholesterol reduction yields clinical benefit: impact of statin trials. *Circulation*. 1998; 97:946-952.

Gray RJ. Flexible Methods for Analyzing Survival Data Using Splines, With Applications to Breast Cancer Prognosis. *Journal of the American Statistical Association*. 1992; 87:942-951.

Grohn YT. Ducrocq V. Hertl JA. Modeling the effect of a disease on culling: an illustration of the use of time-dependent covariates for survival analysis. *Journal of Dairy Science*. 1997; 80(8): 1755-66.

Grover SA. Gray-Donald K. Joseph L. Abrahamowicz M. Coupal L. Life expectancy following dietary modification or smoking cessation. Estimating the benefits of a prudent lifestyle. *Archives of Internal Medicine*. 1994; 154(15): 1697-704.

Hamilton VH, Racicot FE, Zowall H, Coupal L, Grover SA. The cost-effectiveness of HMG-CoA reductase inhibitors to prevent coronary heart disease. Estimating the benefits of increasing HDL-C. *Journal of the American Medical Association*. 1995:273(13): 1032-8.

Hanley J. *Personal communication based on the Science Citation Index*. 1989.

Harold L. Kennedy. The Importance of Randomized Clinical Trials and Evidence-Based Medicine: A Clinician's Perspective. *Clinical Cardiology*. 1999; 22: 6–12.

Hastie T.J., Tibshirani R. *Generalized Additive Models*. 1990. New York: Chapman and Hall.

Hastie, T.J., Tibshirani, R.J. Varying-Coefficient Models (with discussion), *Journal of the Royal Statistical Society*. 1993; 55: 757-796.

Hess, KR. Assessing time-by-covariate interactions in proportional hazard regression models using cubic spline function, *Statistics in Medicine*. 1994; 13:1045-62.

Higgins MW. The Framingham Heart Study: review of epidemiological design and data limitations and prospects. *Progress in Clinical & Biological Research*. 1984; 147: 51-64.

Holland L. The Framingham studies: a family affair. *The Scientist.* 1990; 4(21).

Hu, FB, Stampfer, MJ, Rimm, E, Ascherio A. Rosner, BA, Spiegelman D, Willett, WC. Dietary fat and coronary heart disease: A comparison of approaches for adjusting for total energy intake and modeling repeated dietary measurements. *American Journal of Epidemiology.* 1999; 149(6): 531-540.

Hubert HB, Holford TR, Kannel WB. Clinical characteristics and cigarette smoking in relation to prognosis of angina pectoris in Framingham. *American Journal of Epidemiology.*1982; 115:231-242.

Hubert HB, Feinleib M, McNamara PM, Castelli WP. Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation.* 1983; 67:968-977.

Illingworth DR, Erkelens DW, Keller U, Thompson G, Tikkanen MJ. Defined daily doses in relation to hypolipidaemic efficacy of lovastatin, pravastatin, and simvastatin. *Lancet.* 1994; 343:1554-55.

Jacobs DR Jr., Mebane IL, Bangdiwala SI, Criqui MH, Tyroler HA. High-density lipoprotein cholesterol as a predictor of cardiovascular disease mortality in men and women: the follow-up study of the Lipid Research Clinics Prevalence Study. *American Journal of Epidemiology.* 1990; 131(1): 32-47.

Jajich CL.Ostfeld AM. Freeman DH Jr. Smoking and coronary heart disease mortality in the elderly. *Journal of American Medical Association.* 1984; 252(20):2831-4.

Kagan A. McGee DL. Yano K. Rhoads GG. Nomura A. Serum cholesterol and mortality in a Japanese-American population: the Honolulu Heart program. *American Journal of Epidemiology.* 1981; 114(1): 11-20.

Kahn, H.A. and Dawber, T.R. The development of coronary heart disease in relation to sequential biennial measures of cholesterol in the Framingham Study. *Journal of Chronic Diseases.* 1966; 19: 611-620.

Kalbfleisch, JD and MacKay RJ. On constant-sum models for censored survival data. *Biometrika*, 1979; 66: 87-90.

Kalbfleisch, JD and Prentice RL. John Wiley & Sons (1980). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Inc.

Kannel WB, Castelli WP, Gordon T et al.: Serum cholesterol, lipoproteins, and the risk of coronary heart disease: the Framingham Study *Annals of Internal Medicine*. 1971, 74: 1-12.

Kannel WB, Wilson P, Blair SN. Epidemiological assessment of the role of physical activity and fitness in development of cardiovascular disease. *American Heart Journal*. 1985; 109:876-885.

Kannel WB, Wolf PA, Castelli WP, D'Agostino RB. Fibrinogen and risk of cardiovascular disease: the Framingham Study. *Journal of the American Medical Association*. 1987; 258:1183-1186.

Kannel WB, Wolf PA, Garrioson RJ (Eds.). The Framingham Study: An Epidemiological investigation of Cardiovascular Disease, Section 34, Some Risk Factors Related to the Annual Incidence of Cardiovascular Disease and Death Using Pooled Reported Biennial Measurements, 30 Year Follow-up (Springfield, IL: *US Department of Commerce, National Technical Information Service*). 1988.

Kawachi I. Colditz GA. Stampfer MJ. Willett WC. Manson JE. Rosner B. Speizer FE. Hennekens CH. Smoking cessation and time course of decreased risks of coronary heart disease in middle-aged women. *Archives of Internal Medicine*. 1994;154(2): 169-75.

Kooperberg, C., Stone, C. J., and Truong, Y.K. Hazard Regression. *Journal of the American Statistical Association*. 1995; 90(429): 78-94.

Lakier JB. Smoking and cardiovascular disease. *American Journal of Medicine*. 1992; 93(1A): 15.

Lamarche B, Tchernof A, Moorjani S, Cantin B, Dagenais GR, Lupien PJ, Despres JP. Small, dense low-density lipoprotein phenotype as a predictor of the risk of ischemic heart disease in men: prospective results from the Quebec Cardiovascular Study. *Circulation*. 1996; 94:69-75.

Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *British Medical Journal.* 1994; 308: 367-372.

Lewis RF, Abrahamowicz M, Côté R. Battista RN. Predictive Power of Duplex Ultrasonography in Asymptomatic Carotid Disease. *Annals of Internal Medicine.*1997; 127: 13-20.

Manninen V, Elo MO, Frick MH, Haapa K, Heinonen OP, Heinsalmi P, Helo P, Huttunen JK, Kaitaniemi P, Koskinen P, Mäenpää H, Mälkönen M, Mänttäri M, Norola S, Pasternack A, Pikkarainen J, Romo M, Sjöblom T, Nikkilä EA. Lipid alterations and decline in the incidence of coronary heart disease in the Helsinki Heart Study. *Journal of the American Medical Association.* 1988; 260:641-651.

Margolis JR, Gillum RF, Feinleib M, Brasch RC, Fabsitz RR. Community surveillance for coronary heart disease: the Framingham Cardiovascular Disease Survey. Methods and preliminary results. *American Journal of Epidemiology.* 1974; 100(6): 425-36.

Mason RJ. Buist AS. Fisher EB. Merchant JA. Samet JM. Welsh CH. Cigarette smoking and health. *American Review of Respiratory Disease.* 1985; 132(5): 1133-6.

Mänttäri O, Elo O, Frick MH, Haapa K, Heinonen OP, Heinsalmi P, Helo P, Huttunen JK, Kaitaniemi P, Koskinen P et al, The Helsinki Heart Study: basic design and randomization procedure. *European Heart Journal.*1987; 8:Suppl.1: 1-29.

Menotti A. The relationship of total serum cholesterol to coronary heart disease older men. The Italian rural areas of the Seven Countries Study. *Annals of Epidemiology.* 1992; 2(1-2): 107-11.

Merkel C, Morabito A, Sacerdoti D, Bolognesi M, Angeli P, Gatta A. Updating prognosis of cirrhosis by Cox's regression model using Child-Pugh score and aminopyrine breath test as time-dependent covariates. *Italian Journal of Gastroenterology & Hepatology.*1998; 30(3): 276-82.
Morris G. Russian Lipid Research Clinics Follow-Up Study. *Circulation.* 1994; 89(3):1447.

Myers RH, Kiely DK, Cupples LA, Kannel WB. Parental history is an independent risk factor for coronary artery disease: the Framingham Study. *American Heart Journal.*1990; 120:963-969.

NHLBI: National Heart, Lung, and Blood Institute.

Nieto FJ. Cardiovascular disease and risk factor epidemiology: a look back at the epidemic of the 20th Century. *American Journal of Public Health*. 1999; 89(3): 292-294.

Oldham PD. A note on the analysis of repeated measurements of the same subjects. *Journal of Chronic Diseases*. 1962; 15:969-77.

Pekkanen J. Nissinen A. Vartiainen E. Salonen JT. Punsar S. Karvonen MJ. Changes in serum cholesterol level and mortality: a 30-year follow-up. The Finnish cohorts of the seven countries study. *American Journal of Epidemiology*. 1994; 139(2): 155-65.

Pedersen TR; Olsson AG; Færgeman O; Kjekshus J; Wedel H; Berg K; Wilhelmsen L; Haghfelt T; Thorgeirsson G; Pyörälä K; Miettinen T; Christophersen B; Tobert JA; Musliner TA; Cook TJ. Lipoprotein Changes and Reduction in the Incidence of Major Coronary Heart Disease Events in the Scandinavian Simvastatin Survival Study (4S). *Circulation*. 1998; 97:1453-1460.

Posner BM, Cobb JL, Belanger AJ, Cupples, LA, D'Agostino RB, Stokes J III. Dietary Lipid Predictors and Coronary Heart Disease in Men. The Framingham Study. *Archives of Internal Medicine*. 1991; 151(6): 1181-7.

Quantin C. Abrahamowicz M, Moreau T, Bartlett G, MacKenzie T, Tazi M.A. Lalonde L, Faivre J. Variation over time of the effects of prognostic factors in a population based study of colon cancer: comparison of statistical models.

Rachet B, Sasco AJ, Abrahamowicz M and Benyamine D. Prognostic factor for mortality in nasopharyngeal cancer: accounting for time-dependence of relative risks. *International Journal of Epidemiology*. 1998; 27: 772-780.

Reed D. Yano K. Kagan A. Lipids and lipoproteins as predictors of coronary heart disease, stroke, and cancer in the Honolulu Heart Program. *American Journal of Medicine*. 1986; 80(5): 871-8.

Rigotti NA. Pasternak RC. Cigarette smoking and coronary heart disease: risks and management. *Cardiology Clinics*. 1996; 14(1): 51-68.

Rosenberg L, Kaufman DW, Helmrich SP, Shapiro S. The risk of myocardial infarction after quitting smoking in men under 55 years of age. *New England Journal of Medicine*. 1985; 313:1511-1514.

SAS Institute, Inc. *SAS Language: Reference, Version 6*. SAS Institute, Inc., Cary, NC, 1990.

Schatzkin A. Cupples LA. Heeren T. Morelock S. Kannel WB. Sudden death in the Framingham Heart Study. Differences in incidence and risk factors by sex and coronary disease status. *American Journal of Epidemiology*. 1984; 120(6): 888-99.

Scott M. Grundy; Gary J. Balady; Michael H. Criqui; Gerald Fletcher; Philip Greenland; Loren F. Hiratzka; Nancy Houston-Miller; Penny Kris-Etherton; Harlan M. Krumholz; John LaRosa; Ira S. Ockene; Thomas A. Pearson; James Reed; Reginald Washington; Sidney C. Smith. Jr.; Prevention of Coronary Heart Disease: Guidance From Framingham. *Circulation*. 1998; 97:1876-1887.

Seltzer CC. Framingham studies data and "established wisdom" about cigarette smoking and coronary heart disease. *Journal of Clinical Epidemiology*. 1989; 42(8): 743-50.

Shaten BJ. Kuller LH. Neaton JD. Association between baseline risk factors, cigarette smoking, and CHD mortality after 10.5 years. MRFIT Research Group. *Preventive Medicine*. 1991; 20(5):655-9.

Shea S, Cook EF, Kannel WB, and Goldman L. Treatment of hypertension and its effect on cardiovascular risk factors: data from the Framingham Heart Study. *Circulation*.1985; 71:22-30.

Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, MacFarlane PW, McKillop JH, Packard CJ, for the West of Scotland Coronary Prevention Study Group. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *New England Journal of Medicine*. 1995; 333:1301-1307.

Shurtleff D. Section 30: 'Some characteristics related to the incidence of cardiovascular disease and death: Framingham Study 18-year follow-up', in The Framingham study: An epidemiological investigation of cardiovascular disease, edited by Kannel, W.B. and Gordon, T., *U.S. Government Printing Office* (DHEW publication no. (NIH) 74-599), 1974.

Stamler J, Wentworth D, Neaton JD. Is the relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? Findings in 356,222 primary screenees of the Multiple Risk Factor Intervention Trials (MRFIT). *Journal of the American Medical Association.* 1986; 256: 2823-2828.

Stemmermann GN. Chyou PH. Kagan A. Nomura AM. Yano K. Serum cholesterol and mortality among Japanese-American men. The Honolulu (Hawaii) Heart Program. *Archives of Internal Medicine.* 1991; 151(5): 969-72.

Strong JP, Richards ML. Cigarette smoking and atherosclerosis in autopsied men. *Atherosclerosis.*1976; 23:451-476.

Tamblyn R. Abrahamowicz M. Brailovsky C. Grand'Maison P. Lescop J. Norcini J. Girard N. Haggerty J. Association between licensing examination scores and resource use and quality of care in primary care practice. *Journal of American Medical Association.* 1998; 280(11): 989-96.

The Lipid Research Clinics Program. The Coronary Primary Prevention Trial: design and implementation. *Journal of Chronic Diseases.* 1979; 32: 609-631.

The Lipid Research Clinics Program. The Lipid Research Clinics Coronary Primary Prevention Trial results, I. Reduction in incidence of coronary hearts disease. *Journal of the American Medical Association.*1984; 251-364.

The Lipid Research Clinical Program. The Lipid Research Clinics Coronary Primary Prevention Trial results, II: the relationship of reduction in incidence of coronary heart disease to cholesterol lowering. *Journal of the American Medical Association.* 1984; 251:365-374.

The Multiple Risk Factor Intervention Trial (MRFIT) Research Group. Multiple Risk Factor Intervention Trial: risk factor changes and mortality results. *Journal of the American Medical Association.* 1982; 248: 1465-1477.

The Scandinavian Simvastatin Survival Study Group. Randomized trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet.* 1994; 334: 1383-89.

The West of Scotland Coronary Prevention Study Group. A coronary primary prevention study of Scottish men aged 45-64 Years: Trial Design. *Journal of Clinical Epidemiology.*1992; 45(8): 849-860.

The West of Scotland Coronary Prevention Study Group. Influence of Pravastatin and Plasma Lipids on Clinical Events in the West of Scotland Coronary Prevention Study (WOSCOPS). *Circulation.* 1998; 97: 1440-1445.

Todd PA, Goa KL. Simvastatin, a review of its pharmacological properties and therapeutic potential in hypercholesterolemia. *Drugs.* 1990; 40: 583-607.

Tresch DD. Aronow WS. Smoking and coronary artery disease. *Clinics in Geriatric Medicine.* 1996; 12(1): 23-32.

Trombold JC. Moellering RC Jr. Kagan A. Epidemiological aspects of coronary heart disease and cerebrovascular disease: the Honolulu Heart Program. *Hawaii Medical Journal.* 1966; 25(3): 231-4.

Tsevat J, Weinstein MC, Williams LW, Tosteson AN, Goldman L. Expected gains in life expectancy from various coronary heart disease risk factor modifications. *Circulation.* 1991; 83(4): 1194-201.

Tsevat J. Impact and cost-effectiveness of smoking interventions. *American Journal of Medicine.* 1992; 93(1A): 43S-47S.

Tsoukas CM, Bernard NF, Abrahamowicz M, Strawczynski H, Growe G, Card RT, Gold P. Effect of splenectomy on slowing human immunodeficiency virus disease progression. *Archives of Surgery.* 1998; 133:25-31.

Verweij, J.M., and Houwelingen, H.C. Time-Dependent Effects of Fixed Covariates in Cox Regression. *Biometrics*. 1995; 51: 1550-1556.

Weinstein MC, Stason WB. Cost-effectiveness of interventions to prevent or treat coronary heart disease. *Annual Review of Public Health*. 1985; 6:41-63.

World Health Organization (WHO). *World Health Statistical Quarterly*. 1993; 46(2) & 1995; 48(3/4)

Wu, M. and Ware, J.H. On the use of repeated measurements in regression analysis with dichotomous response. *Biometrics*. 1979; 35: 513-521.

Yano K, Reed DM, McGee DL. Ten-year incidence of coronary heart disease in the Honolulu Heart Program. Relationship to biologic and lifestyle characteristics. *American Journal of Epidemiology*. 1984; 119: 653-66.

Zucker D, Karr AF. Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *Annals of Statistical*. 1990; 18:329-353.