# Simulating the early detection and intervention of vascular disease in the Caerphilly cohort

**By**

## Venkat Krishna Chaitanya Timmaraju

A Thesis Submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy at the Cardiff University

September 2007

UMI Number: U584230

UMI

Dissertation Publishing

ProQuest

Abstract of the thesis entitled

# "Simulating the early detection and intervention of vascular disease in the Caerphilly cohort"

## Submitted by

## Venkat Krishna Chaitanya Timmaraju

## For the degree of doctor of philosophy
## At the Cardiff University
## in September 2007

**Introduction:** The purpose of the project is to simulate the effect of hypothetical intervention on risk of vascular disease in the Caerphilly cohort. The cohort comprises a total population sample of 2959 men aged 45-59 years at the recruitment who has been followed up for 20 years. During that time there has been particular emphasis on assessing exposure to vascular risk factors and assessing vascular related outcomes.

**Aim:** The aim of the thesis is to estimate the effects at population level of public health interventions to change the levels of modifiable risk factors for the vascular disease.

**Methods:** Various statistical techniques such as logistic, fractional polynomial and Cox's proportional hazards models along with various parametric models were used to analyse the data. New risk prediction models were estimated and compared with the existing models in the literature. Various standard simulation techniques were used to simulate hypothetical data using Caerphilly data parameters. Hypothetical interventions were carried out on these generated samples to assess the public health impact.

**Results:** Multivariate analysis suggested that the combined effect of psychological variables measured in the study were significantly associated with the increased risk of MI. New risk prediction models constructed using the Caerphilly study data showed that they were significantly different from the standard available models from the literature. Simulation results suggested that there could be a reduction MI events by 25-30% and stroke events by 50-55% using plausible intervention scenarios available from the literature review.

**Conclusion:** A hypothetical intervention to modify psychological factors showed a higher reduction in MI events. Therefore, plausible interventions to modify psychological factors should be commissioned along with the standard biological and behavioural interventions.

# Acknowledgements

I would like to take this opportunity to express my gratitude to my supervisors, Prof. Frank Dunstan and Dr. John Gallacher, for their tremendous support, continuous guidance and stimulation throughout the entire period of my PhD study. Their ideas, experience, criticism and encouragement are indispensable to this thesis. I would also like to thank Prof. Peter Elwood for all the advice, help and encouragement he as given me when we were sharing the same office.

I would like to thank my family and friends for their support and advice when I was facing hard times. I am thankful to my parents for their constant prayers for my successful life and my sister for supporting me at all times.

**APPENDIX 1:**
**Specimen Layout for Thesis Summary and Declaration/Statements page to be included in a Thesis**

**DECLARATION**

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ................................................................. (candidate)     Date
.............28/0.2/.08..

**STATEMENT 1**

This thesis is being submitted in partial fulfillment of the requirements for the degree of
................................(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed  ................................................................. (candidate)          Date
.............2f/o.2/.o.8..

**STATEMENT 2**

This thesis is the result of my own independent work/investigation, except where otherwise stated.
Other sources are acknowledged by explicit references.

Signed  ................................................................. (candidate)          Date
.............2.8/.o.2/.o.f..

**STATEMENT 3**

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed  ................................................................. (candidate)          Date
.............28/.o2/.o.f..

**STATEMENT 4  - BAR ON ACCESS APPROVED**

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access approved by the Graduate Development Committee**.

Signed  ................................................................. (candidate)          Date
.................................

# Index

# CHAPTER 1

## 1.1 Introduction to the Problem

Vascular Disease (VAD) is the single most common cause of death of both men and women in the UK, with over 235,000 deaths a year, and the national death rate from VAD is amongst the highest in the western world [1]. According to the World Health Organization, VAD causes 12 million deaths in the world each year [2] and yet it is claimed that the disease is largely preventable by healthy lifestyles and effective management of high blood pressure and cholesterol [2]. Vascular disease is not a single disease, with about half of the UK deaths being from coronary heart disease and a quarter from stroke, and the aetiologies of these disease are different and the effects of preventive strategies therefore need to be evaluated separately for these diseases.

The Framingham study was one of the first to study risk factors for VAD [3]. The term "risk factor" is used to denote a factor, whether biological, behavioural or psychosocial, which modifies the risk of developing VAD [3]. Initially raised total cholesterol, smoking and raised blood pressure were linked to the increasing incidence of VAD. Nearly 45 years later, more than 300 other associations have been found between biochemical, clinical, social and demographic variables and the development of VAD. The major risk factors that have been shown consistently to be associated with increased heart disease include gender, age, heredity, smoking, high blood pressure, high cholesterol, obesity and physical inactivity. Psychosocial factors such as emotions, stress and social context have also been implicated in VAD.

Some risk factors such as gender, age and genetic status are not modifiable while others, such as psychosocial factors, smoking, blood pressure, cholesterol, obesity, lack of exercise and diet can, in principle at least, be modified. Therefore preventive strategies need to be targeted at these. The question of which would be the best to target remains an open one. Two issues arise. The first is whether there are simple interventions which can be directed at modifying a risk factor. For example high blood pressure can be lowered in most people by the use of anti-hypertensive drugs, while strategies to persuade people to give up, or not take up, smoking are somewhat

less effective. Secondly the effect of an intervention on the level of the risk factor, and ultimately on the risk of vascular disease, needs to be evaluated.

## 1.2 Introduction to the major risk factors

The major modifiable risk factors which have been identified to date can be divided into three broad categories, biological, behavioural and psychosocial. An overview of each group will be given; particular ones will be discussed in more detail later in the thesis.

### 1.2.1. Biological Risk Factors

The earliest work on risk factors concentrated on biological ones. For example it has been shown that that the risk of VAD increases as blood pressure (systolic or diastolic) increases [4].

Early work implicated total cholesterol as a risk factor but later work suggests that the relationships between lipids and vascular disease is more complex. There is substantial evidence showing the independent associations of both low-density lipoprotein cholesterol (LDL-C) and high-density lipoprotein cholesterol (HDL-C) with the risk of IHD. Results from several epidemiological studies have shown the incidence of coronary artery disease to be positively correlated with LDL-C levels and negatively correlated with HDL-C levels [5, 6]. The Framingham study has reported that there was no connection found between the levels of cholesterol and incidence of stroke [7] but some studies have shown different results [8, 9].

### 1.2.2. Behavioural Risk Factors

Research has shown that several behavioural risk factors such as smoking, alcohol consumption, exercise and diet are related to VAD. These may have direct effects but may also have indirect ones in that they have an effect on the levels of the biological factors discussed above.

There is a lot of evidence indicating that cigarette smoking is a risk factor for VAD [10]. People who smoke at least two packs of cigarettes per day have a two- to three fold greater risk of VAD [11]. The risk for VAD also increases with greater depth of inhalation and with increasing years of smoking, although people who stop smoking

early enough eventually reduce their risk of VAD to a level approaching that of non-smokers [12]. Several studies suggest that exposure to environmental tobacco smoke ("passive smoking") also increases the risk of developing heart disease [13, 14].

There is evidence that moderate alcohol consumption protects against the risk of CHD [15]. On the other hand, it has been suggested that a heavy drinking problem is associated with an increased risk of CHD and other cardiovascular disease [16, 17]. These facts suggest that the association between the risk of CHD and alcohol consumption is a $U$- or, more precisely a $J$-shaped function [18].

Obesity is associated with high rates of VAD deaths, especially sudden death among men [19]. The high death rate might occur largely as a consequence of the influence of being overweight on blood pressure, blood lipid levels, and the onset of diabetes [19]. A report from the Framingham Study indicates that obesity is also an independent risk factor for VAD [20].

### 1.2.3. Psychosocial Risk Factors

In this thesis there will be a strong focus on the influence of psychosocial factors on VAD. The idea that emotions, feeling and social contexts may contribute to the cause of heart disease is not new [21]. For centuries, it has been thought that there exists an association between CHD and the emotions [22]. Emotionally-related reactions that have been linked to heart disease include anxiety, depression, anger and stress [23]. It is beyond the scope of this thesis to describe the complex cognitive, emotional and biological relationships that occur during various affective states but a brief description of each is given below.

Anxiety is part of the emotional response to threat. Anxiety is characterised by feelings of fear, worry, nervousness and panic. The idea that a link may exist between anxiety and the heart has been around for as long as the history of medicine has been documented [24]. However, despite a widespread public perception that anxiety is a significant risk factor for VAD, there are numerous conceptual and methodological difficulties in studying whether a relationship between anxiety and CHD exists. These include the development of theoretically rigorous definitions of anxiety, the development of valid instruments to assess anxiety and the availability of longitudinal

3

data. It is only very recently, with advances in methodology, that possible associations between certain types of anxiety and CHD have been demonstrated [24].

Depression is an extreme form of sadness. Sadness is an emotional response to loss. It is concerned with how we emotionally detach from aspirations and relationships that are either surrendered or taken. When the detachment does not occur and the duration and intensity of the sadness becomes chronic and extreme, depression is the likely result. Depression is characterised by loss of motivation, inactivity and disrupted sleep and eating patterns [24].

Anger is a complex psychological construct. A close scrutiny shows that major components of anger include hostility and an emphasis on cynical, suspicious, and damaging beliefs held toward others [25]. In other terms, hostility is considered to cause increased exposure to anger-related processes by increasing the frequency, intensity, or duration of anger episodes [25].

Stress has been defined in many ways. From a psychological perspective a widely accepted definition is that of Lazarus who suggests stress is the result of two cognitive appraisal processes [26]. Primary appraisal is where the demands of a situation are assessed and secondary appraisal is where the availability of resources to meet the demands is assessed. If the resources are insufficient to meet the demands the person feels stressed. From an epidemiological perspective, however, these definitions are unhelpful as they are difficult to operationalise. Consequently less precise definitions have been used. A convenient operationalisation of stress that is widely used in epidemiology is Type A behaviour. [10]. This refers to a cluster of personality characteristics such as impatience, competitiveness, and hostility. A person with Type A behaviour tends to engage in a perpetual attempt to achieve as much as possible in the least time even though their goals were often unrealistic. Rosenman and Friedman [27] provided the first prospective evidence that this behavioural pattern was a risk factor for coronary heart disease.

### 1.2.4 Multiple risk factors

As has been hinted at above, listing the risk factors as three separate groups is an over-simplification of the problem as the factors are inter-related. It could be that a psychological factor has a direct effect on the risk of vascular disease or that it affects a biological variable which then affects the risk. Alternatively it could change behaviour which might act directly on risk or again may change a biological risk factor. Causal pathways are likely to be extremely complex and this thesis will not explore them. The data needed to do this will be very detailed and complex and are not available for these purposes. While the joint effects of multiple risk factors will be considered, there will be no attempt to build a causal model.

## 1.3 The aim of the project

The aim of this project is to assess the effects on the risk of vascular disease of interventions to change the levels of modifiable risk factors. The gold standard for this would be to conduct randomised controlled trials but these will be lengthy and expensive, Instead the method used here is that of simulation. This requires a risk model, linking the levels of risk factors to the risk of different types of vascular disease. Using this model it is possible to simulate the effects of interventions on the risk factors, not just on the risk factors themselves but also on the risk of disease, and to estimate the benefits of such interventions. There will be particular emphasis on psychosocial factors but biological and behavioural ones will be considered too.

In order to do this risk models are needed. There are many in existence [28] but there is evidence that they do not transfer automatically from one population to another and so the approach adopted here is to build one for a specific population, based on a detailed cohort study, and to direct the simulation at the population from which this was drawn. For these purposes the cohort used was that of the Caerphilly Prospective Study (CaPS) [29], a well-regarded study which has run for nearly 30 years. This cohort consists of a total population sample of 2959 men aged 45-59 at recruitment. In addition to data on standard vascular risk factors such as lipids, hormones and lifestyle (diet, smoking exercise), a broad range of psycho-social risk factor data

(anxiety, anger and depression) are also available. Vascular related outcome variables analysed are myocardial infarction (MI) and stroke.

Although the Caerphilly cohort is well suited for this purpose, as the data come from a single, narrowly defined cohort, the estimates of heath impact will be limited to the male population of Caerphilly in the first instance and South Wales more generally. Nevertheless, the principle of demonstrating a health impact of psychosocial factors over and above that of standard risk factors is of wider significance.

To achieve this aim, a detailed analysis of the Caerphilly data is required to provide an explanatory model. This analysis will examine the effects of the risk factors on the vascular related outcomes, using a variety of techniques including survival analysis, allowing for measurement error where ever possible. These effects can then be incorporated into a simulation model, which simulates the distributions of these risk factors in the population. Interventions to modify the values of these risk factors can be simulated by generating changes in the levels of the risk factors consistent with the effect of the intervention. This in turn generates new risks of VAD and by comparing those with the risk in the absence of the intervention, the effect on the incidence of VAD can be estimated; this is the estimated effect of the intervention. By comparing the results of different interventions it is possible to see which are the most effective.

## 1.4 Thesis structure

The aim of the thesis is to estimate the effects at population level of public health interventions to change the levels of modifiable risk factors for vascular disease. Within that aim there are several specific objectives:

- To construct risk models linking individual risk factors with the risk of MI and stroke and to compare these with those found from literature reviews
- To construct multivariate models linking the risk of vascular disease with multiple risk factors.
- To compare risk models constructed from the CaPS data with others in the literature.
- To review the literature on interventions aimed at modifying risk factors

6

- To simulate interventions to assess their public health impact on vascular disease.

The Caerphilly study is described in more detail in Chapter 2, with summaries of the variables which were measured. Statistical methods are described and illustrated in the Chapter 3, giving more emphasis to less familiar models such as fractional polynomial regression. Methods for estimating regression dilution bias are discussed, along with techniques required for the simulations later in the thesis.

Chapters 4, 5 and 6 deal in turn with the three groups of risk factors. For each factor the literature on the evidence of associations with vascular disease is reviewed briefly, a detailed analysis of the CaPS data is presented, and the results are discussed in the light of the literature review.

Chapter 7 deals with an analysis of multiple risk factors, to assess which ones contribute independently, and the role of the various psychological factors is explored in depth. The survival analysis used to this point is based on the semi-parametric Cox's proportional hazards model. This is ideal for assessing the role of the risk factor but less useful for estimating an actual risk for an individual. In Chapter 8 parametric survival analysis is discussed and a risk model is produced, which is then compared with others in the literature.

Chapter 9 deals with the simulation of the effects of interventions on the risk factors, following reviews of the literature for each relevant risk factor. The impact of plausible interventions is assessed by comparing the number of events likely to be saved in a population such that in Caerphilly.

Finally Chapter 10 gives an overview of the results and addresses areas for future work, as well as highlighting limitations of this work.

# CHAPTER 2

## 2.1 The Caerphilly Cohort

The Caerphilly Prospective Study comprises an integrated research programme based on Caerphilly, a small town in South Wales (Population 45,000). All the classic epidemiological research techniques have been used including cross sectional surveys, case-control comparisons, prospective studies and, in separate population samples, intervention studies including randomised controlled dietary trials into the effects of diet on vascular events and of aspirin on cognitive decline. The study began in 1979 and was designed to provide comparable data with a sister study [30] begun one year earlier in Speedwell, Bristol.

There were 187 research papers published on the Caerphilly study up to June 2005. Around 90 of them were related to heart disease, six on cognitive function, four on Type A behaviour, 25 on diet, eight on psycho-social factors, fourteen on blood samples collected. The remaining publications were on clinical data.

The core of the Caerphilly project is a community-based prospective study into the determinants of cardiovascular disease and cognitive function [31]. The population for the Caerphilly Study was all men who were resident in the town of Caerphilly and five outlying villages. The original cohort of 2512 men aged 45 to 59, drawn from the electoral register, was recruited between 1979 and 1983. Since then, the participants have been re-examined at roughly 5-year intervals. The men who are the subject of this report are those seen at the first re-examination between 1984 and 1988, when they were aged 49 to 65 years. Men of the same age who had moved into the geographically defined area since the original recruitment were also considered eligible. In effect, the cohort was redefined at the first re-examination (phase 2). A total sample, therefore, of 2959 men were identified. Of these, 2398 men were seen at phase 2. The second phase of the cohort is considered as the baseline for the thesis.

In common with other similar studies the main aims of the study included examination of known risk factors for CVD, identification of new risk factors, and

estimating the association of risk of CVD with the risk factors found. Some of the original aims of the work were.

1) The examination of High Density Lipoprotein (HDL) cholesterol as a risk factor for CVD.

2) The examination of certain thrombosis-related tests and CVD.

3) A detailed explanation of dietary factors of possible relevance to CVD.

4) The examination of several hormones (oestradiol, testosterone, and cortisol) in relation to CVD.

Other aims introduced at later stages included the examination of associations between psychosocial factors and the risk of CVD (phase 2) along with the effect of vascular risk factors on cognitive performance and cerebrovascular disease (phase 3).

A small random sample of women was studied in the Caerphilly study to obtain evidence on male-female differences in the various risk factors and determinants. This has the same upper age range to the main male cohort, though because of interest in the possible effects of the menopause the lower age limit was 40 years. A preliminary cross-sectional analysis of these women into dietary and other determinants of lipoproteins has been reported [32, 33].

## 2.2 Data collected

The first phase of the cohort started in 1979 and ended in 1983. Subsequent phases followed in 1984, 1989, 1993. The 5$^{th}$ phase started in 2002 and was completed in 2004, but is not included in this thesis. A variety of data was collected to estimate the risk of cardiovascular disease and cognitive function.

### 2.2.1 CVD data

Data regarding CVD was collected in all phases of the study. The data collected were London School of Hygiene and Tropical Medicine (LSHTM) chest pain questionnaire, 12 lead ECG, family history of IHD, all CVD events between Phases I and IV such as death due to vascular event, myocardial infarction (MI), ECG

myocardial infarction, and history of stroke, stroke events and Transient Ischemic Attack (TIA).

MI can be defined as the death of the heart muscle due to insufficient blood supply, usually due to a clot obstructing blood flow and characterised especially by chest pain. The narrowing of the coronary arteries may be sufficient to prevent adequate blood supply to the heart muscle; this is usually caused by atherosclerosis, and may progress to the point where the heart muscle is damaged due to lack of blood supply. This means that the myocardial infarction is the cause of coronary heart disease. MI and stroke are used in the analyses as the outcome variables.

There were 537 men with prevalent IHD at the start of the 2$^{nd}$ phase of whom 180 were asymptomatic. If these 537 men are included in the analysis, there is a possibility that effects of the disease in terms of behavioural changes and therapy will affect the risk factors and so there might be a case of reverse causality. It was decided to exclude men with prevalent IHD despite the adverse effect of this strategy on sample size.

As in all large studies the dataset is not complete for all variables. For the purpose of the univariate analysis, all subjects with appropriate data were used to maximise the efficiency of the results. Therefore, univariate models may different in the number of subjects between outcomes. For the purpose of multivariate analysis, only men with complete data were included.

As the aim of the project is to model the risk of MI or stroke for men without previous heart disease, the first MI or stroke was considered as the most appropriate outcome variable.

Since most of the coronary-related behavioural variables are also risk factors for stroke, men with symptomatic or asymptomatic IHD are excluded for the analysis of stroke.

## 2.2.2 Cognitive function data

The data regarding cognitive function were collected in phase three, four and five. The information collected concerned novel problem solving (AH4), four-choice serial reaction time (SRT), the Cambridge Cognitive Examination (CAMCOG), which includes the Mini Mental State Examination (MMSE), The Rivermead Behavioural Memory Test and the National Adult Reading Test (NART). The General Health Questionnaire (GHQ30) was also collected to allow adjustment for mood at time of cognitive testing.

The second group of data is the set of possible risk factors. They may be divided into four different categories, namely 'life style and demographic', clinical, psychosocial and blood related factors. The data collected on these factors are listed below.

## 2.2.3 Life Style and demographic factors

There is a large amount of data regarding participants' life style. The data collected were marital status, social class and father's social class, employment status, smoking habit, diet, alcohol consumption, leisure and work activity, sleep pattern and sleep apnoea, and education.

## 2.2.4 Clinical data

Clinical data included medical history, prescribed and over-the-counter (OTC) medications, height and weight, birth weight, skin folds, lung function: Forced expiratory volume (FEV$_1$) and Forced Vital Capacity (FVC), resting blood pressure (random zero sphygmomanometer), London School of Hygiene and Tropical Medicine (LSHTM) chest pain and intermittent claudication questionnaire, hearing: audiometry at four frequencies (phase 2), self report noise exposure and noise sensitivity, strokes and TIA questionnaires, audiometry at eight frequencies (phase 4), Doppler ultrasound carotid flow recording.

## 2.2.5 Psychosocial factors

The psychosocial factors measured in Phase 2 were social support, job satisfaction, Type A Behaviour (Jenkins, Bortner and Framingham questionnaires), Health Attitude Inventory (HAI), Neuroticism and Trait Anxiety scales, GHQ30.

## 2.2.6 Blood Tests

The final group of data are derived from a participant's blood: including total White Cell Count, fibrinogen, plasma viscosity, anti-thrombin III, Lipids, cholesterol, VLDL, LDL, HDL, $HDL_2$, $HDL_3$, triglycerides, testosterone, oestradiol, cortisol, insulin, blood glucose, routine thyroid, liver and kidney function tests. There were other tests in the subsequent phases and many were repeated in all phases.

## 2.3 Methods used in data collection

At each examination, men were invited to attend an afternoon or evening clinic, where a detailed medical and lifestyle history was obtained. At the first re-examination, this included smoking habit and alcohol consumption, social class and employment status, the GHQ30 [34], three questions on social support taken from the Whitehall study [35], the Minnesota leisure exercise questionnaire [36], and a dietary questionnaire [37]. The LSHTM chest pain questionnaire [38] was also used. Along with a full 12-lead ECG, body weight and blood pressure were measured. The men were then invited to come back, fasting, to an early morning clinic, where a blood sample was taken. Cardiovascular risk factors measured included baseline blood pressure, fibrinogen, and white cell count.

## 2.3.1 CVD events

This information is taken from the general questionnaire that has questions regarding medical history, smoking habits etc. For those who died before the follow-up, the information about the date and reason of death was updated from the Office of

National Statistics (ONS).

## 2.3.2 Lifestyle and demographic data

The data collected were marital status, social class and father's social class, employment status, smoking habit, diet questionnaire on food frequency and alcohol consumption.

The question regarding marital status had five options i.e., married, single, widower, divorced, or separated. There were two smoking questions of which one was intended to discover whether the men smoked or not and, if they did, how many cigars and cigarettes every day. Another question was for those who quit smoking, asking them the length of the time since they quit smoking with four options i.e., less than a year, 1-4 years, 5-9 years or more than 10 years.

Diet questions asked about consumption of bread, type of breakfast cereals, fresh fruits, vegetables, fat, milk products, eggs, meat, sugars, and drinks. Information about alcohol consumption was also part of the diet questionnaire. Information was collected regarding the number of drinks and the type of drink the men had per week. It was then recoded into the amount of alcohol consumed in ml/week. Questions were asked about the consumption levels of different types of alcohols such as beers, wine, and spirits etc. For example they asked how many pints (568ml) of beer they had consumed. This consumption was converted into ml of alcohol and summed over different types of consumption. For example, beer contains an average of 5% alcohol. Those who quote that they have consumed 2 pints of beer have consumed $2 \times 5\% \times 570\text{ml} = 57\text{ml}$ of pure alcohol. The final alcohol consumption score was recoded into ml of alcohol per week.

Physical activity information was collected using three questions. The first question asked men whether they were physically active, occasionally active or physically inactive when they were at work. The next question was about the transport they had to go to work. If they travelled to work by cycle or by walking, the distance travelled was recorded. The final question enquired whether the men had any exercise in their leisure hours and if so, the number of minutes they spent doing the exercise.

## 2.3.3 Psychological data

### 2.3.3.1 Trait anxiety

Psychological measurements were taken in the $2^{nd}$ phase of the cohort. Only the General Health Questionnaire (GHQ) measurement was repeated in the subsequent phases of the cohort. Trait Anxiety was measured using the trait scale of the Spielberger State-Trait Anxiety Inventory (STAI) [39]. It consists of 20 statements that ask people to describe how they generally feel. An example of a statement is given below.

|  | Column 1 | Column 2 | Column 3 | Column 4 |
|---|---|---|---|---|
| I lack self-confidence. | Almost never | Sometimes | Often | Almost always |
| STAI score | 1 | 2 | 3 | 4 |

The range of possible scores for STAI varies from a minimum score of 20 to a maximum score of 80. Subjects respond to each A-Trait item by rating themselves on a four-point scale. The validity of this questionnaire is discussed elsewhere [39].

### 2.3.3.2 Anger

Eleven items from the 12 item Framingham study were used to assess anger. These items assess the expression of anger "when really angry or annoyed" [40]. The $12^{th}$ item was omitted in error. Answers for each item were recorded using a three-point Likert response format of "very likely," "somewhat likely," and "not too likely. The anger scales assessed ways of expressing anger such as keeping it to oneself (anger-in), taking it out on others (anger-out) or talking with a friend or relative (anger-discuss) or expressing anger symptoms such as worry or headache (Anger symptoms). The responses were scored as 2, 1, or 0, respectively, and the scales were calculated by summing these response scores. The reliability of this questionnaire was also discussed in the Framingham study [40]. A sample question is as follows

14

| When really angry or annoyed..do you...? | Column 1 | Column 2 | Column 3 |
|---|---|---|---|
| Feel weak | Very likely | Somewhat likely | No too likely |
| Anger score | 2 | 1 | 0 |

## 2.3.3.3 GHQ

The 30-item GHQ measures psychological distress. The long form of the GHQ has 140 questions. This questionnaire was shortened to 60 questions using factor analysis. It was further reduced to 30 questions [34]. Despite its title, GHQ was designed to assess mental health, not "general health" and is a measure of psychological distress including depression. The GHQ was based on features that identify psychiatric cases from general population samples. It is not concerned with differences between psychiatric patients. Each item consists of a question asking whether the respondent has recently experienced a particular symptom or item of behaviour on a scale ranging from 'less than usual' to 'much more than usual'. A sample question is shown below.

| | Column 1 | Column 2 | Column 3 | Column 4 |
|---|---|---|---|---|
| Have you recently been feeling sad and gloomy? | Less so than usual | No more than usual | Rather more than usual | Much more than usual |
| GHQ SCORE | 0 | 1 | 2 | 3 |

The validation of this questionnaire was carried out by several studies such as Split Half Reliability Study [41] amd Validity Study in General Practice Setting [42]. A depression score using this questionnaire was calculated by recoding the responses to first two options of each question to zero and the remaining two responses to one and then adding up all those scores. Respondents scoring 0 to 4 on this new scale were considered non-cases, and those scoring $\geq 5$ were considered possible cases of psychological distress. The GHQ has been validated for depression in the Caerphilly cohort [43].

### 2.3.3.4 Type A behaviour
### 2.3.3.4.1 Jenkins Activity Survey

Type A behaviour was assessed using the Jenkins Activity survey (JAS), Bortner and Framingham Type A scales. The full JAS has four different scales measuring Type A, Speed and Impatience, Job involvement, and Hard Driving and Competitive scales. The Type A scale of JAS has 21 questions [44] and reflects items from the full scale that are most strongly associated with heart disease. The JAS Type A is a self-administered questionnaire and asks about speed and impulsivity, job involvement, and aggressive behaviour. Ten out of twenty one questions have four response options whereas another 10 have three response options. There is only one question that has two response options. The JAS uses a weighted scoring system with item weights being derived from the Western Collaborative Group Study [44].

### 2.3.3.4.2 Bortner type A

The Bortner Type A scale has 14 questions and uses bi-polar analogue response scales [45]. Each question has a response scale that is 2.5 cm in length. Subjects were asked to draw a vertical line at the point where they feel they stand. Each of the statements reflects the way a person behaves in his everyday life. For example, if the subject were generally on time for appointments, he would draw a line toward the right hand side of the line. If he were usually casual about appointments, he would draw a line that will be toward the left hand side of the line. An example question is given below:

Casual about appointments ———————————————— Never late

Each point was converted in to a value between 0 and 25 which is the score for that item. Item scores are summed to give the Type A score. The score ranges from 0 to 350.

### 2.3.3.4.3 Framingham type A
The Framingham Type A scale is based on an interviewer-administered questionnaire and has ten questions [40]. This questionnaire measures emotional liability,

ambitiousness, and not-easygoing scales. Because of the variety of the item responses throughout the questionnaire, i.e., yes-no or multiple choice responses, scales were scored by summing the responses to each question. The number of questions divided these scores. Thus, equal weight was given to each question in the scale. This Type A score ranges from 0 to 1, 1 meaning complete presence of the trait. An example question is as follows

| Question<br>Insert the stem here? | | | | |
|---|---|---|---|---|
| Usually pressed for time | Very well | Fairly well | Somewhat | Not at all |

## 2.3.3.5 Health attitudes

Heart disease is a condition with a multi-factorial aetiology of both behavioural and biological risk factors. Psychological theory proposes that the determinants of behaviour include attitudes towards cardiovascular risk factors[46]. This theory proposes that attitudes might predict coronary-related behaviour and hence coronary heart disease. In the Caerphilly heart disease, attitudes towards coronary-related behaviours were measured using the Health Attitude Inventory (HAI) The description, development and application of HAI are discussed in detail elsewhere [47].

Development of the HAI was based on Fishbein and Ajzen's theory of reasoned action (TRA). There are three components of the TRA.

1) Behavioral intention
2) Attitude
3) Subjective norm

TRA suggests that a person's behavioral intentions depend on their attitudes about the behaviour and subjective norms. If a person intends to do a behaviour then it is likely that the person will do it. Furthermore a person's intentions are themselves guided by two things: the person's attitude towards the behaviour and the subjective norm. In

other words, "the person's perception that most people who are important to him or her think he should or should not perform the behaviour in question" (Azjen and Fishbein, 1975 [48]).

Seven aspects of coronary related behaviour were measured in the Caerphilly cohort. Attitudes towards exercise, dairy produce, fried food and smoking had five beliefs and values questions each, attitudes towards vegetables, wholemeal bread had four value and belief questions, and finally attitude towards stress had six beliefs and value questions. There were seven normative belief questions for each attitude and a single question to measure the motivation to comply for each corresponding normative belief. There were all together 76 questions measuring seven different attitudes towards coronary related behaviours.

Scoring of the HAI has been designed to reflect the meaning behind the judgement made by the subjects and to be relatively simple for the researcher to calculate the score. Scoring is explained briefly here since it is detailed elsewhere [47].

Response to each value question is multiplied to the corresponding belief question and similarly the response to the corresponding normative question of the attitude is multiplied to the response of motivation to comply. All these scores are adding up to the total score of each attitude. Hence, there are seven totals corresponding to each attitude and these scores are used to measure the risk of MI and stroke. A brief description of how the attitudes are measured is discussed in the chapter 6.

### 2.3.4 Clinical data

Clinical information such as weight was noted in at every phase of the cohort. Height was measured for each man when he joined the study. Blood pressure was measured with a random zero sphygmomanometer to get accurate measurements. Blood samples were collected at every phase.

18

## 2.4 Data used in the Caerphilly analysis

Outcomes used in this analysis are first MI and first stroke (fatal or non-fatal). Survival times were calculated as the time from January 1$^{st}$ 1983 until an event occurred. Follow-up ceased on December 31$^{st}$ 2000. Those who dropped out during the study, or reached the end of 2000 without an event, had survival times censored at the end of the follow-up. Due to exclusions and omissions the maximum period time of follow-up used in the analysis was 16.1 years. Stroke and MI survival times were calculated separately.

Of the 2,398 who were eligible for inclusion and were seen at baseline, all were successfully followed for CHD/ MI status either by examination or through medical records. Of the 2398 men, 537 were excluded from the analysis due to prevalent IHD at baseline and 3 were excluded due to refusal for follow-up. There were 61 subjects with their follow-up updated after December 2000. However, these subjects were included in the analysis after their follow-up date was recoded back to the end of 2000. Events which occurred after the follow-up dates (i.e., 31$^{st}$ December 2000) were not recognised. Details of individual missing values of each variable are presented at the time of the analysis. The table showing the measurements that were made during the four phases of the Caerphilly cohort is given in the appendix 1.

# Chapter 3

## 3.1 Statistical methods

This chapter aims to provide the background to the models that are used in subsequent chapters. A brief description is given of standard, widely-used models, followed by a more detailed explanation of more complex models. Linear regression, logistic regression and Cox's proportional hazard model are explained briefly while fractional polynomial regression and methods for estimating regression dilution bias are explained in more detail.

## 3.2 The Linear Regression model

Linear regression is a technique used for modelling the relationship between a response variable and one or more predictor variables. It provides a way of predicting the value of the response variable from the predictor variables. It is assumed that the relationship between a continuous response variable $Y$ and a single explanatory variable $x$ is given by the following model

$$Y_i = \alpha + \beta x_i + \varepsilon_i \qquad i = 1,..., n \qquad (3.1)$$

where $n$ is the sample size, $i$ denotes the $i^{th}$ subject and $\varepsilon_i$ is the measurement error or the random variation in $Y_i$, assumed to follow a normal distribution with zero mean and constant variance. This method can be extended to multiple explanatory variables of the response variable $Y$. The equation is as follows

$$Yi = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} + \varepsilon_i \qquad i = 1,..., n \qquad (3.2)$$

for every subject $i$, $x_{1i},..., x_{pi}$ are the explanatory variables of the outcome variable $Y$.

20

## 3.3 The Logistic Regression model

Linear distribution assumes the outcome variable Y is normally distributed. But more often in the medical research, the outcome variable is binary; 0 representing no disease and 1 for disease. When using linear regression to predict a binary variable of 0 or 1, the predicted values may be out of the range of 0 to 1. Linear regression does not restrict the range of the predicted values and by having a binary variable as a dependant variable in the linear regression analysis, the normality assumption is violated. As the probability of the disease varies between subjects due to a variety of different factors, $p_i$ is used to denote the probability that person $i$ has the disease. Representing these probability values on the real line would avoid the problem of the predicted values to lie outside (0, 1) as the probability can only be between 0 and 1. Various transformations can be used to show these probabilities on a real line. However, 'logit' transformation is most commonly used. A 'logit' transformation for relating the probability $p$ to an explanatory variable $x$ is therefore given by

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta x \tag{3.3}$$

Or equivalently

$$p = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \tag{3.4}$$

This method can be extended to multiple explanatory variables $x_1, \ldots, x_n$

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \tag{3.5}$$

$\beta_i$ can be interpreted as the log odds ratio corresponding to the change of one unit in $x_i$. In some cases, reporting a one unit change is not very informative. For example, suppose an independent variable $x_i$ ranges from 1 to 200. The change in odds for a unit increase in $x_i$ is likely to be very small. Therefore, reporting, say, the effect of a 10 unit increase makes it easier to interpret the importance of the explanatory

variable. To assess the effect of a 10 unit increase, the corresponding $\beta_i$ is multiplied by 10 and then the exponential of that value is taken as the 10 unit change in the odds.

## 3.4 Fractional polynomial regression

The relationship between a response variable and one or more continuous covariates is often nonlinear. Representing the trend of the data in single or multiple regression models is often done by using polynomial functions of the covariates. Lower order polynomials offer a limited family of shapes and higher order polynomials may fit poorly at extremes of the covariates and are also often unstable.

Fractional polynomial regression uses an extended family of curves whose power terms are restricted to a small pre-defined set of integer and non-integer values. The powers of the conventional polynomials form the subset of the family of fractional powers [49].

A polynomial of degree $m$ may be written as

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_m x^m \tag{3.6}$$

whereas a fractional polynomial (FP) of degree $m$ has $m$ integer and/or fractional powers $p_1 < ... < p_m$

$$Y = \beta_0 + \beta_1 x^{(p_1)} + \beta_2 x^{(p_2)} + ... + \beta_m x^{(p_m)} \tag{3.7}$$

where for a power $p$

$$x^{(p)} = \begin{cases} x^p & \text{if } p \neq 0 \\ \log x & \text{if } p = 0 \end{cases} \tag{3.8}$$

$x$ is assumed positive here; if $x$ can take negative values then we can transform it with a additive constant to ensure positivity. An FP of first degree ($m = 1$) involves a single power or log transformation of $x$. It can be easily written as

$$Y = \beta_0 + \beta_1 x^{(p_1)} + \beta_2 x^{(p_1)} \log x \tag{3.9}$$

For $m \geq 2$ and $p_1 = p_2 = ... = p_m$, expression (3.9) may be generalized to

$$Y = \beta_0 + \beta_1 x^{(p_1)} + \sum_{j=2}^{m} \beta_j x^{(p_1)} (\log x)^{j-1} \tag{3.10}$$

For arbitrary powers $p_1 \leq p_2 \leq ... \leq p_m$, let $H_0(X) = 1$, $p_0 = 0$ and combining definition (3.6) with (3.10) we obtain an extended definition

$$Y = \sum_{j=0}^{m} \beta_j H_j(X) \tag{3.11}$$

where $H_1(x) = x^{(p_1)}$ and for $j = 2, ..., m$,

$$H_j(x) = \begin{cases} x^{(p_j)} & \text{if } p_j \neq p_{j-1} \\ H_{j-1}(x) \log x & \text{if } p_j = p_{j-1} \end{cases}$$

Suppose that the interest is to find the best relationship between $X$ and $Y$. the following plot shows the flexibility of the fractional polynomial models with different powers.



**Figure 3.1: non linear function and linear fit to the data**

model with powers 1.5, -2 and 2 —— model with powers 1, 0.5 and 2
model with powers 1 and -0.5
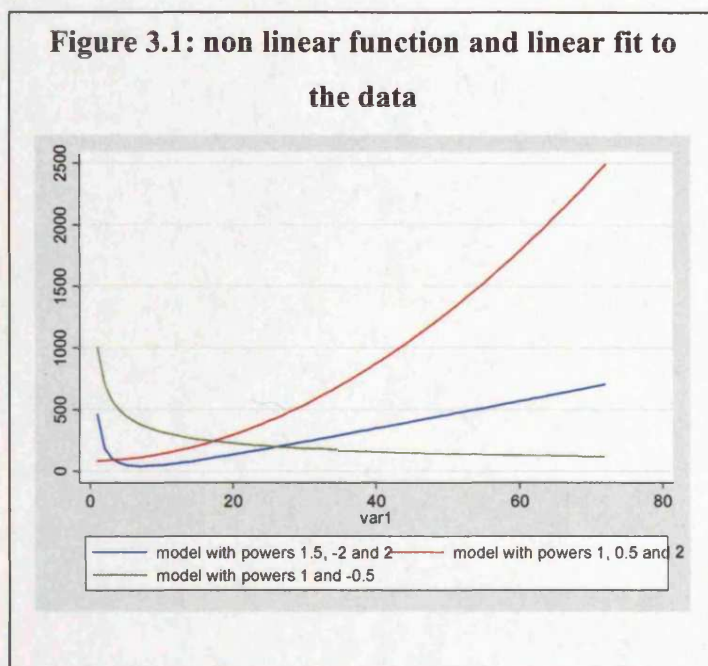
The models that are shown in the above plot are more flexible than a standard linear regression because different combination of powers used by the fractional polynomial

23

model. One of the reasons for using fractional polynomials is that it is often argued that polynomials may fit the data well but give odd behaviour towards the extremes of the $X$ range. Fitting a linear model for that kind of data sets may not yield good results. There is a need for a better regression models like fractional polynomial models that can fit these non-linear functions with more accuracy.

A Linear model is a special case of a fractional polynomial model (when degree = 1). Therefore, a fractional polynomial cannot be worse than a linear model and will be better unless the data are perfectly linear. The improvement of the fractional polynomial model over a linear model can be tiny and may not be worth the extra complexity. There needs to be some criteria specified for choosing the best model. The following section explains one way of choosing a best model.

### 3.4.1 Choosing the best model.

The deviance of a model is defined as -2 times its maximised log likelihood. Models are fitted for every combination of fractional powers in the predefined set (STATA uses these powers as pre-defined (-2, -1, -.5, 0, .5, 1, 2, 3)) for a given model and the deviance of each model is calculated. The gain for the fractional polynomial model is defined as the deviance for a straight line (i.e., for the model $\beta_0 + \beta x$) minus the deviance for the fractional polynomial model; thus it is a measure of the improvement in fit from using the more complex model. The larger the gain, the greater the nonlinearity in the relationship between the outcome variable and $x$. Significance is assessed by comparing the gain with a $\chi^2$ test with a degrees of freedom equal to the number of powers used in the fractional polynomial at 10% significance level.

## 3.5 Non Parametric smoothing (Lowess smoothing)

Regression models assume a particular type of relationship between the outcome variable and the covariates. It would be useful to explore the data first to assess what sort of model would be useful. So a non-parametric approach might be better as an exploratory tool.

The name "Lowess" is derived from the term "locally weighted scatter plot smooth," using locally weighted linear regression to smooth data.

It is of interest to look at the relationship between $X$ and $Y$. suppose we have the data $\{(x_i, y_i), i = 1,....n\}$ with $x_i \leq x_i+1, i = 1,....., n-1$.

To estimate a smoothed value of $y_i$, we take a window central of $x_i$, that is a collection of $\{x_i\}$ close to $x_i$. A weighted regression is applied to the $\{(x_i, y_i)\}$ in the window with weights decreasing with the distance from $x_i$. The regression model is used to predict the value of $y_i$ corresponding to $x_i$; this is the smoothed values of $y_i^s$. The smoothing process is considered local because each smoothed value is determined by neighboring data points defined within the window.

The subset of $X$ used in calculating $y_i^s$ is indices $i_- = \max(1, i - k)$ through $i_+ = \min(i + k, N)$ where $k = [(N * \text{bandwidth} - 0.5)/2]$.

where the weights for each of the observations between $j = i_-,......., i_+$ is given by

$$w_j = \left\{ 1 - \left( \frac{|x_j - x_i|}{\Delta} \right)^3 \right\}^3 \qquad (3.11)$$

where $\Delta = 1.0001 \times \max (x_{i_+} - x_i, x_i - x_{i_-})$.

Bandwidth determines the percentage of the data used within the subset of $X$. If the bandwidth is 0.4, then the model uses only 40% of the data with in the subset of $X$. therefore, the bandwidth of 1 implies using all the data within the subset of $X$ to estimate the corresponding value $y_i$. The optimal bandwidth is dependent on the variable and dataset, and is determined through examination of smoothed profiles plotted against unsmoothed ones.

Using this model gives a good idea about the relationship between two variables. The relationship is essentially non-parametric if the window is fairly narrow and the smoothed values are data driven not model driven. Polynomial smoothing methods and other parametric smoothing methods smooth the data globally which means that what happens at the one end of the plot while smoothing can affect the other end. By using Lowess smoothing, one can restrict the bandwidth and control the smoothing locally. As the bandwidth increases, the smoothness of the relationship increases. The following plot shows the Lowess smoothing plot for the relationship between the SBP and total cholesterol with different bandwidths.



**Figure 3.2: Lowess Smoothing**
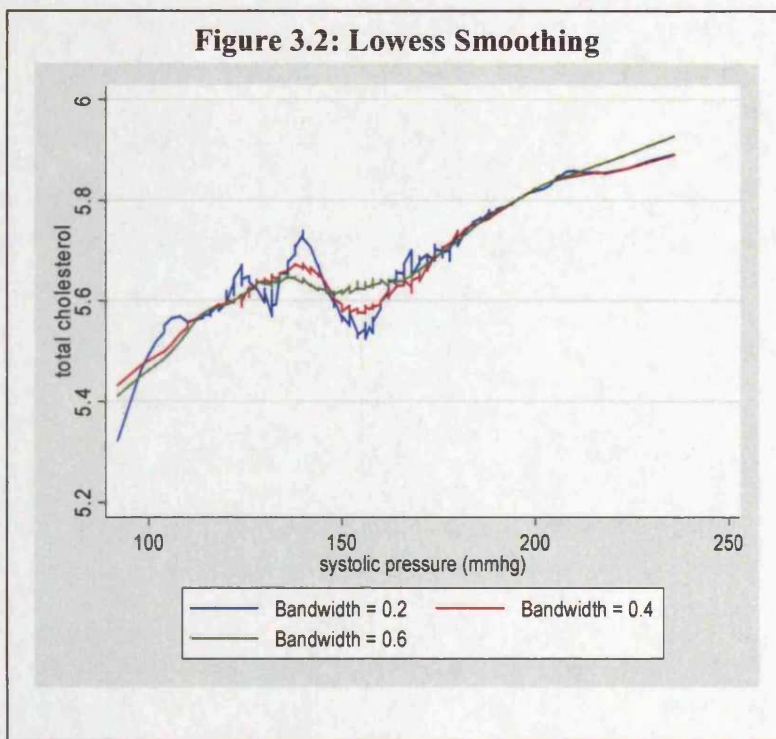
In the above plot, it can be seen that the smoothness of the relationship between SBP and total cholesterol increases as the bandwidth increases.

## 3.6 Survival analysis

In any realistic clinical trial it takes time to accrue the patients for the trial. Often it takes several years. This is also true of retrospective studies which review results of past patients, again because the patients will have been diagnosed or started treatment at different times, often over several years. Either way, this means that patients are followed for survival starting at different times. But the results are analyzed at one time and so at that time, the patients have varying lengths of follow-up.

Mathematically removing a subject from the analysis before the event of interest has occurred is called "censoring" the time until the event. The difficulty is that such subjects have incomplete information. Suppose one subject has no event by the end of the study, when the subject had been in the study for 5 years. So, we know that survival time for that subject is $\geq 5$ years. These subjects cannot be ignored with these censored values and they must be taken into account to use all their data as efficiently as possible and this requires the use of more complex methods.

### 3.6.1 Kaplan-Meier Survival Estimates

The survival rate is expressed as the survivor function (S):

$$S(t) = \frac{number\ of\ individuals\ surviving\ longer\ than\ t}{total\ number\ of\ individuals\ studied}$$

for any positive $t$.

The hazard function, say $h(t)$, represents the "instantaneous" or immediate risk of death or failure for an individual who has survived to time $t$. More specifically, it may be defined as the limit of the probability that an individual who has survived to time $t$ will die in the short interval from $t$ to $t + \Delta t$, as $\Delta t$ goes to zero. This can be expressed in the following equation as

$$h(t) = \frac{f(t)}{S(t)} \qquad\qquad (3.12)$$

where $f(t)$ is the survival function of time $t$.

The cumulative hazard function (H) is the risk of event (e.g. death) at time $t$ and can be estimated by the method of Peterson [50] as:

$$\hat{H}(t) = -\ln(\hat{S}(t))$$

The product limit method of Kaplan and Meier [51] is used to estimate S:

$$\hat{S}(t) = \prod_{t_i \leq t}\left(1 - \frac{d_i}{n_i}\right) \tag{3.13}$$

where $t_i$ is an event time. The $t_i$ values are known survival times ordered so that $t_1 \geq t_2 \geq \ldots\ldots$, $d_i$ is the number of deaths up to point $t_i$ and $n_i$ is the number of individuals at risk just prior to $t_i$. The $n_i$ values take account of censoring in that the change from $n_i$ to $n_{i+1}$ adjusts for values censored between $t_i$ and $t_{i+1}$. $S$ is based upon the probability that an individual survives at the end of a time interval, on the condition that the individual was present at the start of the time interval. $S$ is the product of these conditional probabilities.

If a subject is last followed up at time $t_i$ and then leaves the study for any reason (e.g. lost to follow up) $t_i$ is counted as their censorship time.

### 3.6.2 Cox's Proportional Hazard Model

The proportional hazard model is the most general of the regression models because it is not based on any assumptions concerning the nature or shape of the underlying survival distribution. The model assumes that the underlying hazard rate is a function of the independent variables (covariates); it is assumed that the hazard function follows the same basic shape but that there is no need to formally model the shape parametrically. Thus, in a sense, Cox's regression is usually called a semi-parametric method as it is parametric with respect to the covariates. The model may be written as

$$h(t, x_1, x_2, \ldots, x_m) = h_0(t) \times \exp(\sum_{i=1}^{m} \beta_i x_i) \tag{3.14}$$

where $h(t,...)$ denotes the resultant hazard, given the values of the $m$ covariates $(x_1, x_2, ..., x_m)$ and the survival time $(t)$. The term $h_0(t)$ is called the baseline hazard; it is the hazard for an individual for whom all independent variable values are equal to zero. This model can be linearized by dividing both sides of the equation by $h_0(t)$.

$$\log\left(\frac{h\{(t),(x_1,x_2,...,x_m)\}}{h_0(t)}\right) = \sum_{i=1}^{m} \beta_i x_i \qquad (3.15)$$

The left hand side ratio in the equation (3.15) is called a *relative hazard* function.

According to the equation (3.15), $\beta_i$ can be interpreted as the log hazard ratio corresponding to a change of one unit in $x_i$; this is analogous to the interpretation in the logistic model. In some cases, reporting a one unit change is not of great importance and reporting, say, a 10 unit increase makes it easier to interpret the effect of the explanatory variable. To obtain the effect for a 10 unit increase, the $\beta_i$ is multiplied by 10 and then the exponential of that value is calculated.

### 3.6.3 Assumptions

While no assumptions are made about the shape of the underlying hazard function, the model equations shown above have two assumptions. First, they specify a multiplicative relationship between the underlying hazard function and the log-linear function of the covariates. This assumption is also called the *proportionality assumption*. In practical terms, it is assumed that, given two observations with different values for the independent variables, the ratio of the hazard functions for those two observations does not depend on time. The second assumption of course, is that there is a log-linear relationship between the independent variables and the underlying hazard function.

### 3.6.4 Parametric survival models

Due to the semi-parametric nature of the Cox's proportional hazard model, it is not straightforward to estimate the probability of having an event as it does not assume any shape of the hazard function. With a parametric survival model, however, the survival function has a distributional form and so probabilities can be calculated. Therefore, using survival parametric models has advantages. Furthermore, survival

parametric models can be used in simulations with ease as they will be seen in Chapter 9. This section gives a brief description of the parametric models used in this thesis.

### 3.6.4.1 Exponential survival model

The probability density function and cumulative distribution functions are,

$$f(t) = \lambda e^{-\lambda t} \qquad \text{for } t \geq 0, \lambda > 0 \qquad (3.16)$$

and

$$F(t) = 1 - e^{-\lambda t} \qquad (3.17)$$

respectively.

The survival function is therefore

$$S(t) = e^{-\lambda t} \qquad (3.18)$$

The hazard function

$$h(t) = \lambda \qquad (3.19)$$

which is a constant.

### 3.6.4.2 Weibull survival model

The Weibull distribution is a generalisation of the exponential distribution. Unlike the exponential distribution, it does not assume a constant hazard rate and therefore has broader applications. The Weibull distribution is defined by two parameters $\gamma$ and $\lambda$.

The probability density function and cumulative distribution functions are,

$$f(t) = \lambda \gamma (\lambda t)^{\gamma - 1} e^{-(\lambda t)^{\gamma}} \qquad \text{for } t \geq 0, \gamma, \lambda > 0 \qquad (3.20)$$

and

$$F(t) = 1 - e^{-(\lambda t)^\gamma} \tag{3.21}$$

respectively.

The survival function is therefore

$$S(t) = e^{-(\lambda t)^\gamma} \tag{3.22}$$

The hazard function is

$$h(t) = \lambda \gamma (\lambda t)^{\gamma - 1} \tag{3.23}$$

The value of $\gamma$ determines the shape of the distribution and $\lambda$ determines its scaling. Therefore, $\gamma$ and $\lambda$ are called the shape and scale parameters respectively. When $\gamma = 1$, the hazard rate remains constant as time increases; this is the exponential case as the Weibull distribution reduces to an exponential distribution in that case. The hazard function increases when $\gamma > 1$, and decreases when $\gamma < 1$, as $t$ increases. Thus, the Weibull distribution may be used to model the survival distribution of a population with increasing, decreasing or constant risk.

Weibull regression models are used as a parametric alternative to the Cox model. Since Weibull regression models are comparatively easier to simulate than the semi-parametric Cox models, these will be used for simulations.

### 3.6.4.3 Non-standard parametric survival model

Anderson et al [52] in 1990 discussed a model that was used for estimation and prediction of risk scores for individuals based on several cardiovascular risk factors. Instead of using a traditional parametric survival model or a semi-parametric survival model, they have used a non-standard parametric survival model. They argued that the standard Weibull model assumes the logarithm of time until an event has a constant dispersion and location parameter. They proposed a model which allows for a variable dispersion and/or location parameter for the survival times and claimed that for their data it was a better model then a Weibull one. This is a clever model because it does

31

not require a proportionality assumption as in the case of Cox's model and is more efficient while dealing with survival times.

Let $T$ denote the time until the outcome of interest (MI or stroke in this case). Assume that $x_i$ represent the risk factors measured for an individual where $i = 1$ to $k$. For example, $x_1$ might be systolic blood pressure, $x_2$ total cholesterol and so on. The coefficients $\theta_0$, $\beta_0$, $\beta_1$,.....$\beta_k$ and $\theta_1$ are the parameters that are estimated using this model. Assume that the natural logarithm of the survival time $T$ has a location $\mu$ and a dispersion $\sigma$.

For a known distribution function, $F$:

$$P\left\{\frac{\log(T) - \mu}{\sigma} \leq u\right\} = F(u)$$

Anderson [53] further defined a varying location and dispersion accelerated failure time (VLDAFT) model:

VLDAFT: $\mu = \beta_0 + \beta_1 x_1 + ..... + \beta_k x_k$       (3.25)

and $\log \sigma = \theta_0 + \theta_1 \mu$.

According to Anderson [53], the VLDAFT model allows dispersion to be modelled as a linear function of the covariates, which includes a time-varying component on the scale parameter and the location parameter. He shows

$F(u) = 1 - \exp(- \exp(u))$,       (3.26)

the logarithm of the cumulative hazard function for a failure time is defined by

$$\log[-\log\{\Pr(T \geq t)\}] = \frac{\log(t) - \mu}{\sigma}$$

which implies that $T \sim$ Weibull and $\log(T)$ has an extreme value distribution.

When $\theta_1 = 0$, then the dispersion parameter $\sigma$ is constant and the model is called linear location accelerated failure time model (LLAFT).

In general, a negative $\beta$ coefficient for a variable means that a high value of that variable is associated with high risk.

The coefficients $\theta_0$, $\beta_0$, $\beta_1$,.....$\beta_k$ and $\theta_1$ are estimated using a SAS program explained elsewhere [54]. This program uses the Proc Nlin procedure of SAS. The initial values for this converging procedure are obtained from the Weibull accelerated failure time models estimated in STATA. Two sets of coefficients were estimated using the Framingham data for the risk of MI and stroke.

To find the best fit, assuming that $T$ follows a Weibull distribution, $\beta$ coefficients for LLAFT and VLDAFT model are estimated along with the maximised log likelihoods for each model. Then a difference between the log likelihoods for LLAFT and VLDAFT is tested with the log likelihood ratio test, explained in section 3.4.1, at a 5% significance level to decide which one of them is the better model.

## 3.7 Regression Dilution Bias

The classical regression model assumes that there is no error in the independent variable. Suppose that the relationship between $x$ and $y$ is modelled using a regression model where $x$ is the independent variable. What happens if $x$ is not a true value and is subject to, say, measurement error? If there is measurement error in $x$, that is bound to change the relationship between the $x$ and $y$. The amount of change in the coefficient of $x$ due to the measurement error is called the coefficient of regression dilution bias or the coefficient of measurement error. The following sections illustrate methods to estimate the coefficient of measurement error.

### 3.7.1 Single Variable

The regression slope between a response and predictor variable is underestimated when the predictor variable is measured imprecisely. Repeat measurements of the predictor in individuals in a subset of the study or in a separate study can be used to estimate a multiplicative factor to correct for this 'regression dilution bias'.

Consider a relationship between a response variable $y_i$ (for subject $i = 1, ..., N$) and a predictor variable $x_i$ and the random error $\delta_i$.

$$y_i = \alpha^* + \beta^* x_i + \delta_i \qquad\qquad \delta \sim N(0, \phi^2) \qquad\qquad (3.27)$$

Suppose that $x_i$ is subject to measurement error, so the observed value may not be the true value. Suppose that the relationship between the true measurement and the observed measurement is as follows

$$w_i = x_i + u_i \qquad\qquad x_i \sim N(\mu, \sigma_x^2), \quad u_i \sim N(0, \sigma_w^2) \qquad\qquad (3.28)$$

where $w_i$ is the observed value of $x_i$, $\sigma_x^2$ and $\sigma_w^2$ are the variance of $x_i$ and $\delta_i$.

If $u_i$, $\delta_i$ and $x_i$ are independently distributed it is shown by Snedecor and Cochran [55] that $y_i$ and $w_i$ follow a bivariate normal distribution and the regression of $y_i$ on $w_i$ is linear as follows

$$y_i = \alpha + \beta w_i + \gamma_i \qquad\qquad \gamma_i \sim N(0, \psi^2) \qquad (3.29)$$

where

$$\beta^* = \frac{\beta(\sigma_x^2 + \sigma_w^2)}{\sigma_x^2} \qquad\qquad (3.30)$$

From the above equation, $\beta$, the coefficient from the regression of $y$ on the observed value, underestimates the true value $\beta^*$ by an amount which depends on the relative magnitudes of the measurement error variance and the variance of the true $x$ values.

If there are repeated measurements available, it is possible to obtain an estimate of $\sigma_w$ as follows

Let $\delta$s independent of $x$, then

$$w_1 = x + \delta_1$$

and $\qquad w_2 = x + \delta_2$,

where $w_1$ and $w_2$ are the observed values of $x$ at two different points of time.

Here the assumption is that the actual measurement variable $x$ does not change, but the error terms differ.

$$\therefore w_1 - w_2 = \delta_1 - \delta_2$$

$$\Rightarrow \mathrm{var}|w_1 - w_2| = 2\sigma_\delta^2$$

$$\Rightarrow \sigma_\delta = \frac{SD|w_1 - w_2|}{\sqrt{2}} \qquad\qquad (3.31)$$

Using the above equation the variance of the error can be estimated and hence the revised estimate of $\beta$ can be calculated. This method can also be extended to the multivariate scenario.

## 3.7.2 Multiple variables

If the above univariate scenario is converted into a multivariate scenario, the equation (3.31) rewritten as

$$\beta^* = \beta \times \Gamma \qquad (3.33)$$

where
$$\Gamma = \left(\hat{\Sigma}_x + \hat{\Sigma}_h\right)\hat{\Sigma}_x^{-1} \qquad (3.34)$$

where $\hat{\Sigma}_x$ is the observed variance matrix and $\hat{\Sigma}_h$ is the error variance matrix.

where $\Gamma$ is a matrix of coefficient of regression dilution biases that are to be estimated.

To simplify this, an example of a bivariate problem is given.

Suppose the true values are unchanged and repeated observations are obtained for both variables, say

$$\text{Therefore } Z_1 = X + h_1 \qquad Z_2 = X + h_2$$
$$\text{and } Y_1 = M + \delta_1 \qquad Y_2 = M + \delta_2.$$

Suppose that there are two measurements on two variables. Let X, M denote the true values and Z, Y denote the observed values.

For two variables, equation 3.34 becomes

$$\Gamma = \left[\begin{pmatrix} \text{var}(X) & \text{cov}(X,M) \\ \text{cov}(X,M) & \text{var}(M) \end{pmatrix} + \begin{pmatrix} \text{var}(h) & \text{cov}(h,\delta) \\ \text{cov}(h,\delta) & \text{var}(\delta) \end{pmatrix}\right] \times \begin{pmatrix} \text{var}(X) & \text{cov}(X,M) \\ \text{cov}(X,M) & \text{var}(M) \end{pmatrix}^{-1}$$

cov(Z, Y) can be easily estimated from the data as follows.

$$\hat{\sigma}_z^2 = S_z^2, \text{ the usual sample variance, and } \hat{\sigma}_{zy} = S_{zy} = \frac{1}{n}\sum(z - \bar{z})(y - \bar{y})$$

Assuming that X and $h$ are independent, and similarly for M and $\delta$, then

Var(Z) = Var(X) + Var($h$), Var(Y)=var(M)+ Var($\delta$) and

Cov(Z,Y) = cov(X,M) + cov($h$, $\delta$)

Therefore cov(X, M) = cov(Z, Y) – cov($h$, $\delta$).  $\qquad (3.35)$

To get cov($h$, $\delta$)

$Z_1 - Z_2 = h_1 - h_2$, and $Y_1 - Y_2 = \delta_1 - \delta_2$,

So that cov($h_1 - h_2$, $\delta_1 - \delta_2$) = cov($Z_1 - Z_2$, $Y_1 - Y_2$)

The last term is easy to estimate. The left hand side is

$$\text{Cov}(h_1, \delta_1) + \text{Cov}(h_2, \delta_2) - \text{Cov}(h_2, \delta_1) - \text{Cov}(h_1, \delta_2)$$

and if the errors at different time points are assumed independent, then the last two terms will be 0.

Therefore $2\text{Cov}(h, \delta) = \text{cov}(Z_1 - Z_2, Y_1 - Y_2)$

$$\therefore \text{cov}(h, \delta) = \frac{\text{cov}(Z_1 - Z_2, Y_1 - Y_2)}{2} \tag{3.36}$$

Substituting equation (3.36) in equation (3.35), cov(X, Z) can be calculated. Substituting all the covariance matrices in the equation (3.34), the regression dilution bias of X and M can be calculated.

The above example illustrates how the measurement error is estimated for a bivariate model. In order to extend this to a multivariate scenario, it is the same method with bigger matrices.

### 3.7.3 Misclassification in binary data

**Case 1:**

Suppose that we want to assess the effect of a dichotomised factor with some misclassification. Chu et al [56] suggested a method for taking into account the impact of imperfect sensitivity and specificity. Here the sensitivity refers to the probability of a person who is really exposed being classified as such, while the specificity is analogously defined for those not exposed. The outcome and the exposure are tabulated as follows

**Table 3.1 Observed counts at baseline**

| | Misclassified exposure | | |
|---|---|---|---|
| Disease Status↓ | 1 = present | 0 = absent | Total |
| Case | a | b | $N_1$ |
| Control | c | d | $N_0$ |

Assuming that the sensitivity (*Se*) and specificity (*Sp*) of the exposure and outcome at baseline is same as the sensitivity and specificity of the exposure and outcome at the time of repeated measurement,

the adjusted $\hat{OR}$ is given by

$$\hat{OR} = \frac{(a + N_1 \times (Sp - 1))(N_0 \times Se - c)}{(c + N_0 \times (Sp - 1))(N_1 \times Se - a)} \dots\dots\dots\dots\dots(3.37)$$

and the SE of the adjusted $\log(\hat{OR})$ is given by

$$SE[\ln(\hat{OR})] = \sqrt{\frac{N_1 ab(Se + Sp - 1)^2}{(N_1 Se - a)^2 (N_1 Sp - b)^2} + \frac{N_0 cd(Se + Sp - 1)^2}{(N_0 Se - c)^2 (N_0 Sp - d)^2}} \dots\dots(3.38)$$

Chu et al used a Meta analysis results based on the 10 publications using the superior cotinine validation methods to estimate the sensitivity and specificity [57]. This meta-analysis showed that the sensitivity of the self reported smoking habit ranged from 0.82 to 1.00 and specificity ranged from 0.91 to 1.00.

Using these ranges of sensitivity and specificity from the meta-analysis and simple tabulated values of smoking behaviour and the outcome, the adjusted odds ratios can be estimated.

The sensitivity and specificity can be estimated from the repeated measurements of the data. However, it has to be assumed that the status of the exposure is not changed in the repeated measurement.

However, it is not possible to find an estimate of sensitivity and specificity for every exposure.

Therefore, another method for misclassifications is discussed.

**Case 2:**

Let U be the proportion of members of the stratum two, who are correctly classified in the stratum 2 by the imperfect measures and V be the corresponding proportion of stratum one. Using the information in the table below, the U and V are defined as U = $(b_2 + d_2) / (a_2 + c_2)$ and V = $(b_1 + d_1)/(a_1 + c_1)$.

**Table 3.2 Exposure at 1$^{st}$ and 2$^{nd}$ measurement with outcome at baseline**

|  | Confounder level 1 | |
|---|---|---|
|  | Exposed | Not exposed |
| Cases | $a_1$ | $b_1$ |
| Controls | $c_1$ | $d_1$ |
|  | Confounder level 2 | |
|  | Exposed | Not exposed |
| Cases | $a_2$ | $b_2$ |
| Controls | $c_2$ | $d_2$ |

Then using the method developed by the Walker and Lanes [58] the following formula gives the new adjusted odds ratio.

$$\hat{OR} = OR \times \frac{\left(1 - \dfrac{a_2 \times (1 - U)}{a_1 \times U}\right)\left(1 - \dfrac{d_2 \times (1 - U)}{d_1 \times U}\right)}{\left(1 - \dfrac{b_2 \times (1 - U)}{b_1 \times U}\right)\left(1 - \dfrac{c_2 \times (1 - U)}{c_1 \times U}\right)} \qquad (3.39)$$

where $OR$ and $\hat{OR}$ are the odds ratios before and after the adjustment with the second measurement.

The feasible values for U and V are as follows:

1) $U + V < 1$ ; $0 \leq U \leq 0.365$ ; $0 \leq V \leq 0.4$

2)  $U + V > 1$ ; $0.6 \leq U \leq 1$ ; $0.635 \leq V \leq 1$

Though this method does not estimate the standard error, it is more appropriate where the sensitivity and specificity are not readily available.

# Chapter 4

# 4.0 Introduction to the behavioural Factors

In Chapter 2 possible risk factors for vascular disease were discussed under three headings, namely behavioural, biological and psycho-social factors. In the next three chapters, the risk factors will be examined individually, and their association with the risk of either an MI or stroke will be assessed. In each case there will be a brief literature review followed by a detailed analysis of the relevant CaPS data. This chapter deals with the behavioural risk factors and will be followed by chapters on biological and then psycho-social factors. The modelling strategy is set out in detail for the first risk factor; for subsequent risk factors some of the details will be omitted.

## 4.1 Alcohol Consumption

### 4.1.1 Literature Review

High alcohol consumption is associated with an increased risk of heart disease [59]. This is thought to be due in part to an effect of ethanol on blood pressure. Moderate alcohol intake, however, may be associated with a reduced risk of heart disease relative to abstaining. This could be due to a beneficial effect of ethanol on blood lipids.

There have been many population studies of alcohol and vascular disease. The Whitehall II Cohort Study (UK) measured alcohol consumption in 10,308 civil servants aged between 35 and 55 at baseline. (33% female). A 'U' shaped relationship was found between the volume of alcohol consumption per week and the risk of coronary heart disease [60]. They showed that the teetotallers and those who consumed at least 248 grams of pure alcohol per week when compared to those who consumed 1-80 grams per week had a twofold increase total mortality but the results were not significant for coronary heart disease. A cohort study conducted in Finland with a general population of 5,092 men aged between 24 and 65 showed that acute alcohol consumption is associated with an increased risk of death due to heart attack [61]. Men with six or more drinks at a time have a higher risk of 1.77 (95%CI: 1.01, 3.08) for having ischaemic heart disease compared to those who had no heavy

occasional drinking. The Framingham Study, with 10,333 male and female subjects, conducted in the U.S.A showed a 'J' shaped relationship between alcohol consumption and atrial fibrillation. The relative risks were, 0.97 (95% CI: 0.78, 1.22), 1.06 (95% CI: 0.80, 1.38), 1.12 (95% CI: 0.80, 1.55), and 1.34 (95% CI: 1.01, 1.78) for alcohol categories of 0.1 to 12, 12.1 to 24, 24.1 to 36, and >36 g/day respectively when compared to teetotallers [62]. Several epidemiological studies have shown an inverse relationship between moderate alcohol consumption and the risk of cardiac events [63, 64]. One of these studies estimated the risks of MI with RR=1, 1.02, 0.82, 0.61 in the groups of monthly, weekly, daily consumptions compared to rarely/never [63]. Another review [65] looked at the relationship between alcohol consumption and vascular events in elderly men. It suggested that the reduced risk of cardiovascular disease observed among drinkers in different studies of alcohol consumption can be partly explained by an increased risk among ex-drinkers who stopped drinking because of their health conditions and who are included with the non drinkers, an example of reverse causality [65].

This summary of some of the literature suggests that the effect of alcohol consumption on CHD varies between studies, with some suggesting that higher alcohol consumption leads to greater risks while others suggest alcohol has a protective effect. The 'J' shaped relationships shown by several studies might be attributed to differences in population samples, methods of collecting data and cultural differences between the study populations. It is important to note that a range of outcomes have been used. Though these outcomes are closely related to either MI or stroke, their subtle differences may account for some of the differences between findings.

A Meta analysis study with 116,702 subjects from 61 different studies showed a 'J' shaped relationship between alcohol consumption and MI events for a general population [66]. The inclusion criterion for this meta-analysis was to search for all possible outcomes related to cardiovascular disease such as coronary heart disease, coronary artery disease, coronary event, coronary death, myocardial infarction, ischaemic heart disease and angina pectoris for their risk with alcohol or ethanol consumption. The same authors considered a number of other disease outcomes and found no evidence for J-shaped relationships with alcohol consumption. [67]. They

showed that alcohol consumption appears to give a protective effect for the risk of MI in those who consume 20 - 72gms/day (approximately $2/3^{rd}$ of a pint of beer to 2.5 pints of beer).

Alcohol consumption has been identified as a possible risk factor for stroke for over three centuries [68]. Several studies have shown that the relationship between alcohol consumption and stroke is not straightforward. A systematic review has shown that individual studies reported inconsistent relationships between alcohol consumption and stroke which is not surprising because of the potential biases mentioned previously [69]. However, the majority of the evidence suggests that light and moderate consumption is associated with a protective effect whereas heavy alcohol consumption is associated with an increased risk of stroke (i.e., a 'J' or 'U' shape relationship) [70-73]. A Meta analysis of 35 studies found that, taking teetotallers as a reference group, those who consume between 1 and 12 grams of alcohol per day have a RR of 0.83 (95% CI: 0.75 – 0.91), 12-24 grams have RR of 0.92 (0.78 – 1.06), 24-60 grams have 1.10(0.97 – 1.24) and >60 grams per day have RR=1.64 (95% CI: 1.39-1.93) for having a stroke [74].

Several papers have been published on the Caerphilly Cohort data discussing the effect of alcohol on the risk of vascular events: none of them showed any significant association with risk for any vascular event [75-77]. Previous analyses merely considered the relationship between alcohol consumption and both MI and stroke as a linear relationship whereas this thesis will explore more appropriate methods to identify the actual relationships. It is also important to note that the previous analysis included men with incident heart disease.

### 4.1.2 Alcohol consumption in the Caerphilly study

Alcohol consumption was measured in the Caerphilly cohort in the $2^{nd}$, $3^{rd}$ and $4^{th}$ phases. Information about alcohol consumption was collected in the diet questionnaire. Data regarding the number of drinks and the types of drinks these men had per week was collected. It was then recoded into the amount of alcohol consumed in cc/week. Questions were asked about the consumption levels of different types of alcoholic drinks such as beers, wine, spirits etc to find the volume consumed. This was converted into an intake of alcohol. For example, beer has an average of 5%

alcohol in it. Those who quote that they had consumed 2 pints of beer had consumed $2 \times 5\% \times 568cc = 56.8cc$ of pure alcohol, as there are 568cc in 1 pint. The consumptions of different types were summed to get the final alcohol consumption score, in cc/week. Figure 4.1 shows the distribution of alcohol consumption in each phase.



**Figure 4.1 – Histogram of Alcohol consumption in 3 phases of the Caerphilly study**
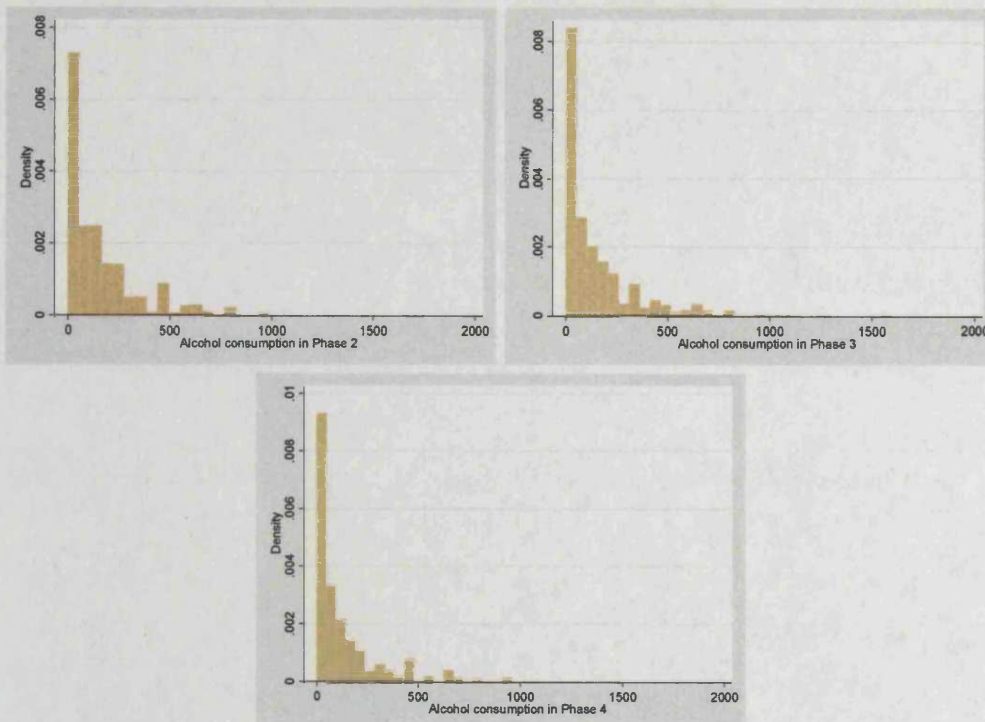
In the above figure, it is obvious that alcohol consumption is not normally distributed. There were 115 teetotallers in the first phase data i.e., 6.2%. This proportion is similar to that found in several other studies (8.6%[2], 5.5% [3], and 5.7% [19]).

After excluding the missing values in the alcohol data, the total number of observations in the 2$^{nd}$ phase considered for the analysis is 1856. The mean and standard deviation of alcohol consumption for each of the three phases are given in the Table 4.1.

**Table 4.1 – Summary statistics of alcohol consumption in 2nd, 3rd and 4th phases**

| Variables | Obs | Mean | Std. Dev | Min | Max |
|---|---|---|---|---|---|
| Alcohol Ph 2 | 1856 | 159.48 | 200.78 | 0 | 1756 |
| Alcohol Ph 3 | 1578 | 146.88 | 184.34 | 0 | 1596 |
| Alcohol Ph 4 | 1514 | 136.80 | 187.23 | 0 | 1487 |

There were 1418 subjects with alcohol consumption measurements in all three phases. The mean consumptions in the 2nd, 3rd and 4th phases for those 1418 subjects were 164, 150 and 140 cc/week respectively, suggesting that the mean alcohol consumption appears to be decreasing over the period of follow-up.

### 4.1.3 The Caerphilly analysis

Before formally modelling the relationship between alcohol consumption and MI or stroke, the nature of the relationship was explored using a smoothed plot of the data. Lowess smoothing, as explained in chapter 3, produces one such plot. The relationship between alcohol consumption and MI or stroke was visualised using Lowess smoothing with a bandwidth of 0.8 and is shown in Figure 4.2. Experimenting with various other bandwidths between 0.7 and 0.99 was carried out but the results were similar. Future plots always use the value 0.8.

**Figure 4.2: Lowess smoothed plot for relationship between alcohol consumption and MI & Stroke**
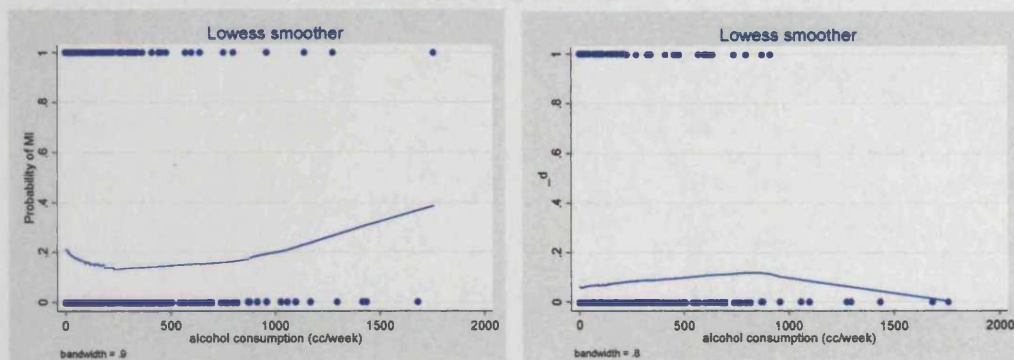


The first plot in the above figure shows the relationship between alcohol consumption and MI in the Caerphilly data. It suggests a non-linear relationship ('J' shape) as mentioned in some of the literature referred to earlier [66]. The relationship between

45

alcohol consumption and stroke appears to be non-linear, possibly due to the few extreme values of alcohol consumption. If these are excluded, the relationship between alcohol consumption and stroke appears linear, similar to that found in the meta-analysis [67].

Many studies in the literature review analysed their data by either using a linear logistic regression or by quoting the odds ratios comparing teetotallers with different categorical levels of alcohol consumption, such as 'low', 'medium' or 'high'. These are crude methods of assessing the status of alcohol consumption status as a risk factor. On the one hand using categorical data loses valuable information but makes fewer assumptions about the nature of the model while the linear logistic model makes stronger assumptions than a fractional polynomial model.

We decided to use the reported consumption as a continuous variable to preserve the detail but to examine carefully the model to be fitted. The range of consumption was very large; in particular there were some extreme values in excess of 1000cc per week. It is somewhat implausible that a linear relationship would be appropriate for this full range of values and so it was decided to transform the alcohol consumption to a less skew distribution. An important consideration is the large number who reported no consumption. Any transformation should be able to handle those and to transform those to a value not too far removed from the rest of the distribution. A logarithmic transformation is a natural one to take but a constant has to be added to the consumption to cope with zero values. As a result the chosen transformation was

$$logalc = \log(alc + 4)$$

The distribution of logalc is shown in Figure 4.3. 'logalc' is used instead of actual alcohol consumption data for all the future analyses of MI.

**Figure 4.3: Histogram of logalc**
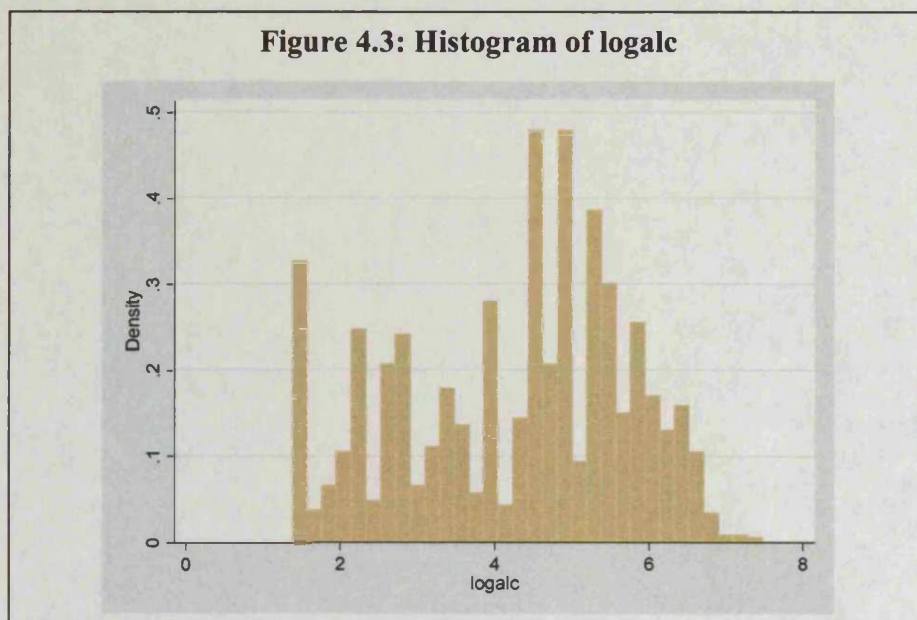


## 4.1.3.1 MI and alcohol consumption

The relationship between MI and alcohol consumption (i.e, logalc) is shown in Figure 4.4. The bandwidth for smoothing the data was 0.8. as described earlier

**Figure 4.4: Relationship between logalc and MI**
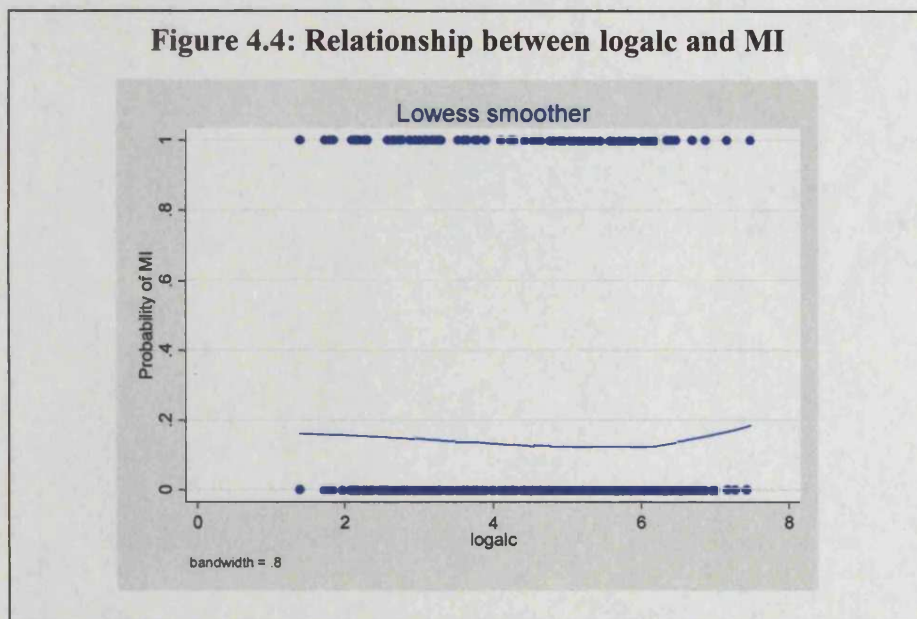


The relationship appears to be non-linear and hence a fractional polynomial regression was used to estimate the relationship between 'logalc' and MI.

### 4.1.3.1.1 Logistic regression

There are 1797 men with 243 first MI events between them. The logistic regression model for this relationship is given below.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = -1.59 - 0.63\text{logalc} \dots\dots\dots (4.1)$$

where $p$ is the probability of an MI.

The standard errors and $p$ values for the constant and coefficient of logalc and the constant in the model are SE = 0.20, 0.05 and $p$ = 0.0001, 0.17

The slope of the above equation is negative suggesting that the increase in alcohol consumption decreases the risk of MI. Figure 4.4 suggests that this may be true for values of logalc below 6 (consumption less than 400 cc per week) but not for larger consumptions and so it is plausible that a non-linear relationship holds. Therefore, a fractional polynomial logistic model was fitted and the improvement in fit assessed. Table 4.2 gives the coefficients of the fractional polynomial model.

### Table 4.2 – Fractional polynomial regression of logalc

| MI | Coef | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| Logalc1 | -0.019 | 0.010 | 0.083 | -0.039 | 0.002 |
| Logalc2 | 0.009 | 0.005 | 0.095 | -0.002 | 0.021 |
| Constant | -1.932 | 0.095 | 0.000 | -2.118 | -1.746 |

where $\text{logalc1} = \text{logalc}^3 - 77.84$, $\text{logalc2} = \text{logalc}^3 \times \log(\text{logalc}) - 112.98$

This model has been chosen as the best fit to the data using the method explained in section 3.4. Substituting these values in the equation, the fractional polynomial regression equation is as follows.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = -(0.02)\text{logalc}^3 + (0.01)\text{logalc}^3 \times \log(\text{logalc}) - 1.47 \dots\dots\dots (4.2)$$

where $p$ is the probability of an MI.

The relationship estimated by the fractional polynomial regression from the above equation can be visualised by the following plot in Figure 4.5 along with the Lowess fit.



**Figure 4.5: Predicting MI using logalc by Fractional polynomial regression**
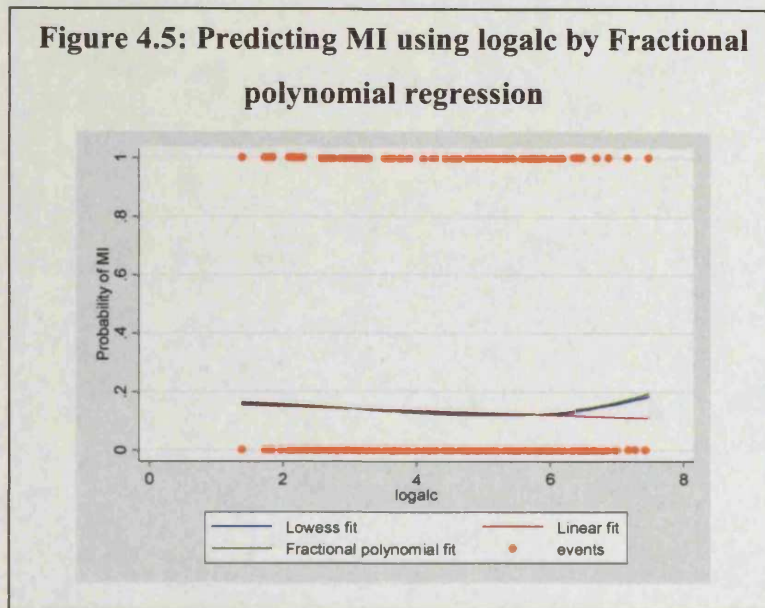
The logistic fractional polynomial regression fits logalc very closely in the first part of the above figure. The deviance gain of this model over the linear model is 1.71, showing that there is no statistically significant difference between linear logistic and fractional polynomial logistic in modelling the MI risk as a function of alcohol consumption, as the minimum needed for statistical significance, for 2 extra parameters, is 5.99

Although the fractional polynomial model did not give a significant improvement the fact that it gave a fitted shape similar to that of the Lowess plot, and that the meta-analysis described earlier suggested a J-shaped model was appropriate, led that to be the one chosen for further analysis

### 4.1.3.1.2 Cox's Proportional Hazard Model

A Cox's proportional hazards model (section 3.6.2) gave an estimated hazard ratio of 0.94 (95% CI: 0.87 – 1.03). While not significant, the fact that the estimate is less than 1 suggests that the risk of an MI decreases as alcohol consumption increases.

When Fractional polynomials were used for Cox's proportional hazard model, the resulting model coefficients are shown in . Table 4.3.

**Table 4.3 – Cox's Proportional hazard model of logalc and MI**

| MI | Coef | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| Logalc1 | 0.984 | 0.009 | 0.094 | 0.965 | 1.003 |
| Logalc2 | 1.008 | 0.005 | 0.107 | 0.998 | 1.019 |

where $\text{logalc1} = \text{logalc}^3 - 77.84$, $\text{logalc2} = \text{logalc}^3 \log(\text{logalc}) - 112.98$

The information in the above table can be interpreted as follows

$$\log\left(\frac{h\{(t),(\text{logalc})\}}{h_0(t)}\right) = -0.01 \times \text{logalc1} + 0.01 \times \text{logalc2} \dots\dots\dots\dots (4.3)$$

The hazard ratios fitted by the above model are plotted below.



**Figure 4.6: Hazard ratios for having an MI**
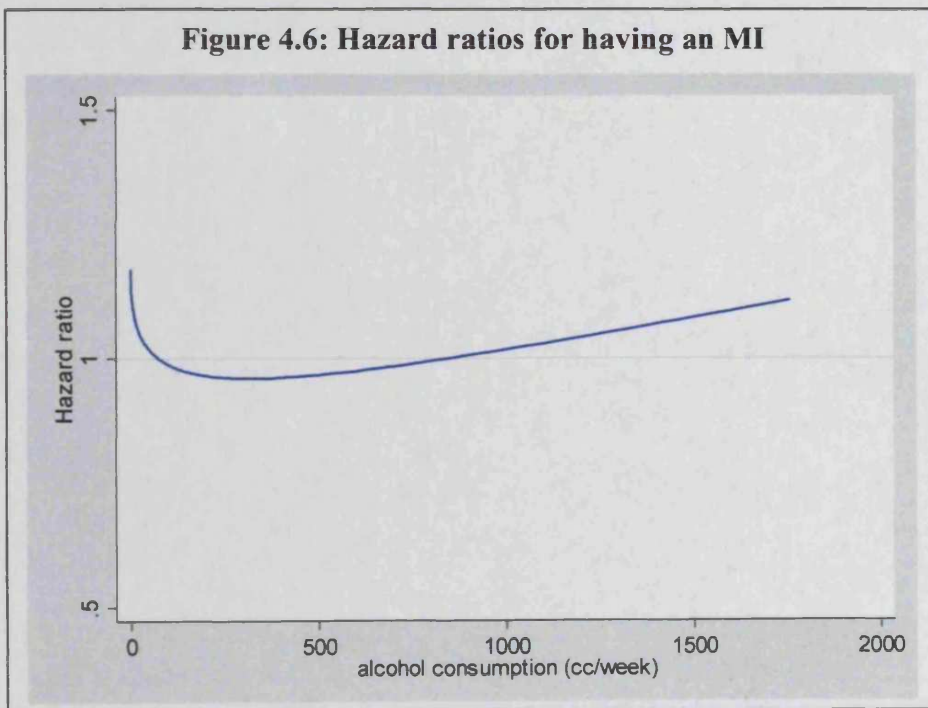
The improvement of the deviance for the fractional polynomial model is 1.5. This shows that the fractional polynomial model is not statistically significantly better then the linear model but for the reasons stated above it was decided to use this model.

50

For further analyses of the other risk factors, if the fractional polynomial regression were not found to be significantly different from the linear logistic model and the literature review suggests that there is no non-linearity in the relationship; the equations of the logistic regression model will be shown.

A simple way to understand the relationship between the risk of MI and alcohol consumption over time is to plot Kaplan-Meier survival curves. For these purposes only, alcohol consumption was divided into four different categories, defined by teetotallers, occasional consumers (consume less than 135 cc/week), moderate consumers (consume alcohol between 135 and 435 cc/week) and high consumers (consume more than 435 cc/week). These groups were based on groups used by studies in the literature review.



**Figure 4.7: Predicting MI for three different groups of alcohol consumption**
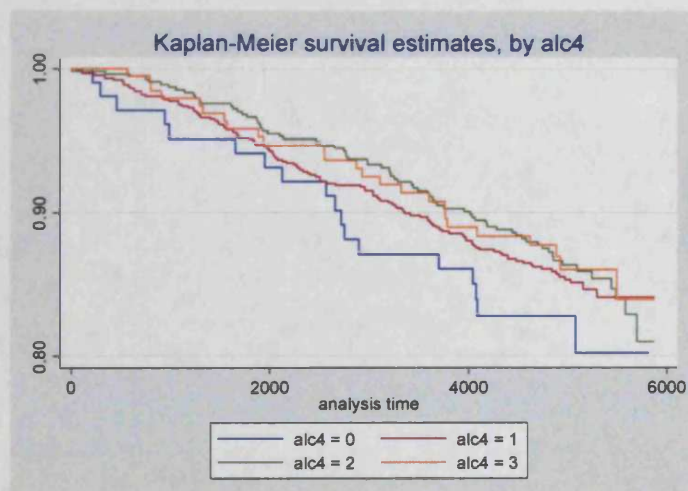
where alc4=0 Teetotallers
         alc4=1 Men who consume less than 135cc/week
         alc4=2 Men who consume between 135 & 435 cc/week
         alc4=3 Men who consume more than 435cc/week

In the above figure, the teetotallers appear to have a higher risk of having an MI, though it is not significant (Peto-Peto test for equality of survivor functions showed a $p$ value of 0.71).

#### 4.1.3.2 Stroke and alcohol consumption

The relationship between alcohol consumption and stroke is shown in the second plot of Figure 4.2. It appears to be linear for all but the highest levels of consumption. If these extreme values are excluded from the analysis, linear regression can be used to model the relationship between alcohol consumption and stroke. Since the literature review suggests a linear relationship [67], it was thought to be appropriate to go with the exclusion of these extreme values.

#### 4.1.3.2.1 Logistic Linear regression

Coefficients of the linear model for the relationship between alcohol consumption and stroke are given in the table 4.4.

**Table 4.4 – Logistic linear regression of alc and stroke**

| MI | Coef | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| ALCOHOL | 0.001 | 0.000 | 0.012 | 0.000 | 0.002 |
| CONST | -2.722 | 0.120 | 0.000 | -2.957 | -2.486 |

Substituting these values in the equation, the linear logistic regression model is as follows.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = (0.001)\text{alc} - 2.722 \ldots\ldots\ldots (4.4)$$

where $p$ is the probability of stroke for all men.

The relationship estimated by the linear logistic regression for the equation (4.4) can be visualised by the following plot in Figure 4.8 along with the Lowess fit.

**Figure 4.8: Predicting stroke for all logistic linear regression**
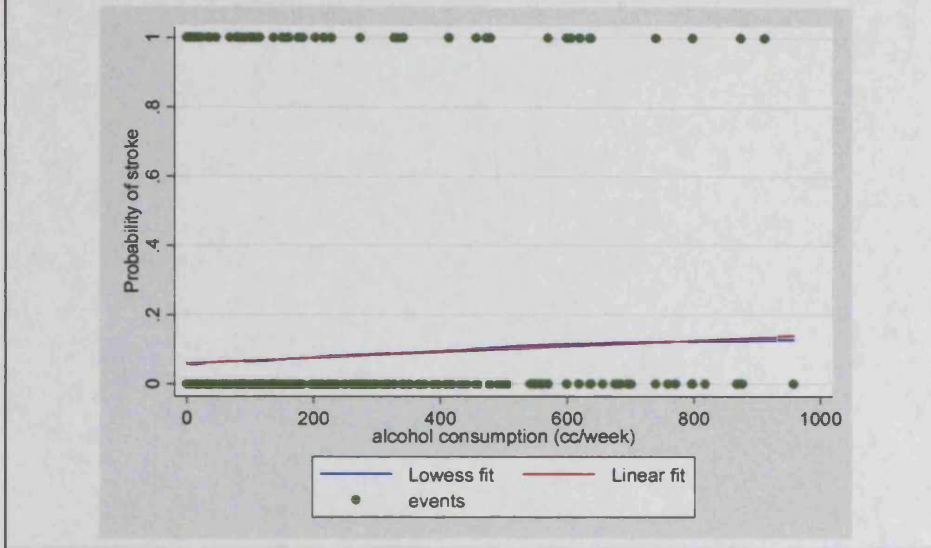


From the above figure, it appears that the linear logistic model fits the relationship between alcohol consumption and stroke reasonably well. The above equation expresses the log odds of stroke for a 1cc/wk increase in alcohol consumption, a tiny one. For every 100cc/wk increase in alcohol consumption, the odds of stroke increase by 11%.

### 4.1.3.2.2 Cox's Proportional Hazard Model

Cox's proportional hazard model was used to estimate the hazard ratios for the risk of stroke with alcohol consumption. The hazard ratio for an increase of one unit of alcohol consumption is 1.001 [95% CI: 1.0003 – 1.0020]. The $p$ value and standard error are 0.01 & 0.0004, showing that the relationship between alcohol consumption and the risk of stroke is significant. For an increase of 100cc/wk which is approximately equal to 3.5 pints of beer per week, the hazard ratio for having a stroke increases by 11%.

Similar to the relationship between alcohol consumption and MI, the Kaplan-Meier survival plots of the four groups of alcohol consumption for having a stroke are shown in Figure 4.9.

**Figure 4.9 Predicting stroke for three different groups of alcohol consumption**
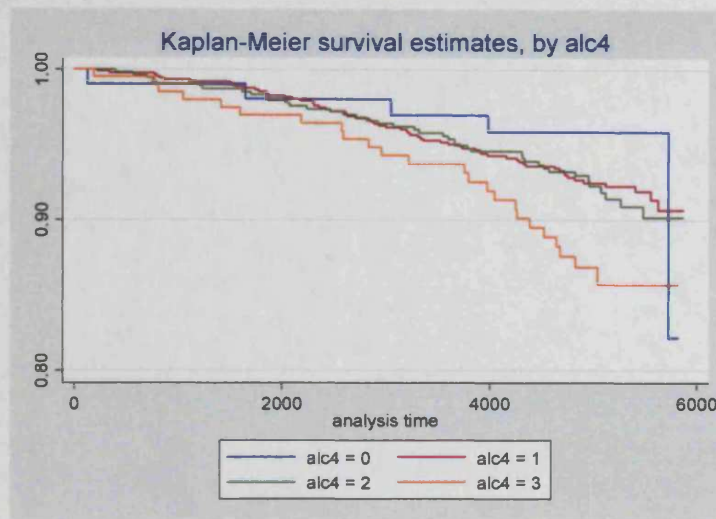


Kaplan-Meier survival estimates, by alc4

where alc4=0 Teetotallers
alc4=1 Men who consume less than 135cc/week
alc4=2 Men who consume between 135 & 435 cc/week
alc4=3 Men who consume more than 435cc/week

When the three different groups of alcohol consumption were compared for the risk of stroke, it appears that the heavy consumption group has more risk (Peto-Peto test for equality of survivor functions showed a $p$ value of 0.04). Both these plots, along with the models fitted, suggest that the risk of stroke increases significantly as the consumption levels increase.

### 4.1.4 Conclusions

A Meta analysis study discussed in the literature review has shown a similar 'J' shape relationship between MI and alcohol consumption similar to the one shown in the first plot of Figure 4.2 [67]. It showed that alcohol consumption over 83cc, or 2.76 on the log scale per day increases the risk of MI. The model fitted to the Caerphilly data has a minimum risk of MI at 6 units on the log scale (Figure 4.5) which is very different to what the meta-analysis has shown. Because of the very few subjects with very high alcohol consumption levels in the Caerphilly data, the relationship between logalc and MI appears to be 'U' shaped.

Several population studies that were considered for the meta-analysis paper, with different sample sizes and different follow-up times, showed a 'J' shape relationship

54

between alcohol consumption and MI. Some of the studies included in the Meta analysis had different relationships with the outcomes. Five of them are plotted in Figure 4.10 using the information available from their published results (studies 1-5 [62, 78-81]). These studies were selected because the data required for this comparison was available in the published papers. These plots below show the relative risks of different alcohol consumption groups compared with teetotallers.

It was thought that a better idea is to use logarithms of the alcohol consumption scores instead of the true values to bring the extreme values closer which resulted in showing 'U' shape relationship as some of the studies in the literature review. Studies from different cohorts showed predominantly a 'U' shape relationship between alcohol consumption and MI. This appears to suggest that the logarithmic values of alcohol consumption show a 'U' shaped relationship.

The first study shown in the figure below had its alcohol consumption data measured in grams of alcohol consumed/day [82]. There were divided into 5 groups as teetotallers, <12gms/day, 12-24gms/day, 24-36 gms/day and >36gms/day. Other studies' alcohol consumption levels can be observed from the plots given below.

**Figure 4.10 – Fitted models from different studies**
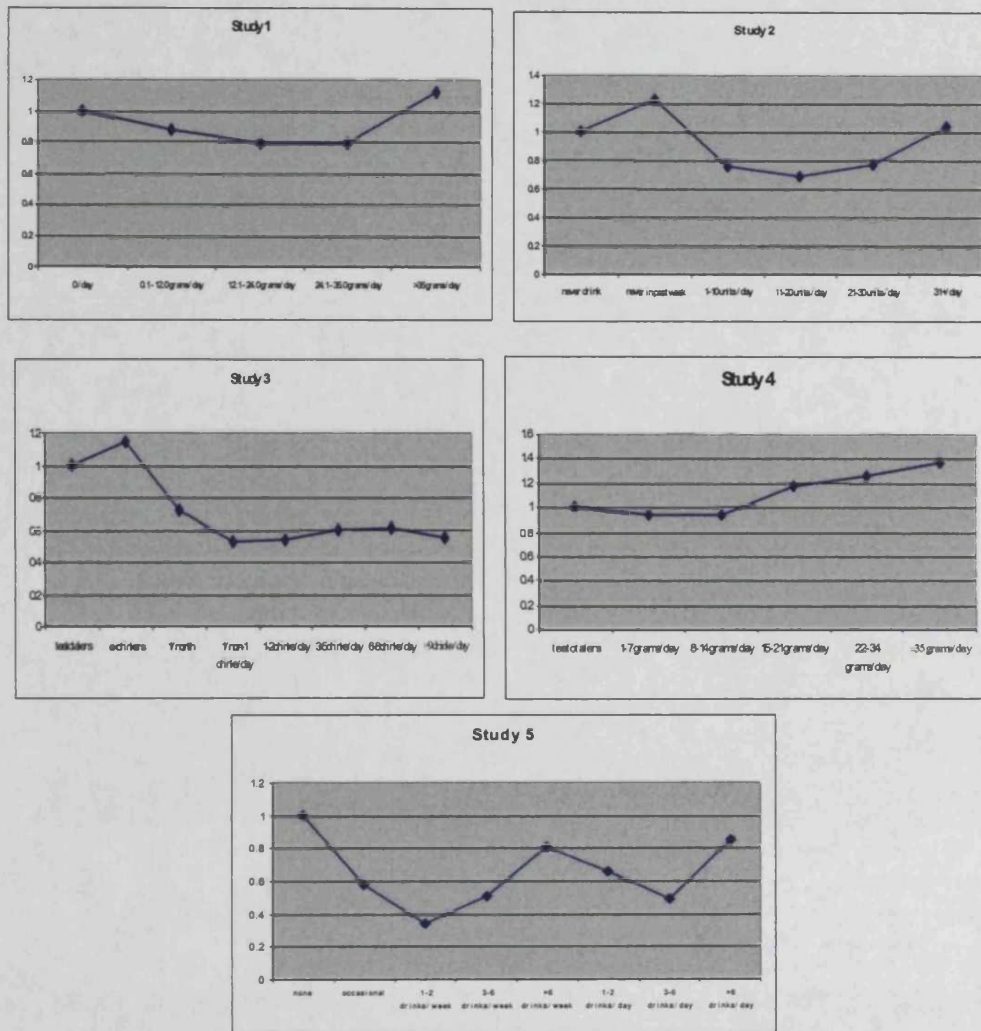
For example, study 3 in Figure 4.10 has a sample of men and women included, whereas age groups of subjects in each study differ. One of the major differences in all these studies was the follow-up time that varied from 4 years to 21 years. The differences in the follow-up time of each study are not addressed in the Meta analysis. Therefore, some other sophisticated models are needed for estimating these inter-study biases.

The overall conclusion from these analyses is that the risk of MI is lower in the moderate alcohol consumption group whereas the risk of stroke increases as the alcohol consumption increases.

56

Increase in alcohol consumption appears to show an increased risk of stroke in the Caerphilly analysis with a linear relationship. A meta analysis study in the literature review showed that that the relationship between alcohol consumption and stroke is not linear. A Meta analysis of 35 studies found that, taking teetotallers as a reference group, those who consume between 1 and 12 grams of alcohol per day have a RR of 0.83 (95% CI: 0.75 – 0.91), 12-24 grams have RR of 0.92 (0.78 – 1.06), 24-60 grams have 1.10(0.97 – 1.24) and >60 grams per day have RR=1.64 (95% CI: 1.39-1.93) for having a stroke [74]. This shows that there might be a J shape or a U shape relationship for stroke as well. However this is not evident from the analysis of the Caerphilly study data but it is not clear why this different relationship has been found.

## 4.2 Smoking

### 4.2.1 Literature Review

There are approximately 4000 chemicals in cigarette smoke, many of them toxic. The ingredients in cigarettes affect everything from the internal functioning of organs to the efficiency of the body's immune system. The effects of cigarette smoking are destructive and widespread. Toxic ingredients in cigarette smoke travel throughout the body, causing damage in several different ways. Nicotine reaches the brain within 10 seconds after smoke inhalation and can be found in every part of the body of a smoker [83]. Carbon monoxide binds to haemoglobin in red blood cells, preventing affected cells from carrying a full load of oxygen. Smoking affects the function of the immune system and may increase the risk for respiratory and other infections. There are several likely ways that cigarette smoke does its damage. One is oxidative stress that mutates DNA, promotes atherosclerosis, and leads to chronic lung injury [83].

There is much evidence that smoking increases a person's risk of cardiovascular disease, cerebrovascular disease (ischaemic and haemorrhagic stroke), aortic aneurysm and peripheral arterial disease. In Sir Richard Doll's study [84] the risk of mortality from any cardiovascular disease in all cigarette smokers was 1.6 times that of those who never smoked, with the figure rising to 1.9 times in heavy smokers when compared to non smokers [84]. Over the age of 60 years, the risk of heart attack doubles in smokers compared to non-smokers, but under the age of 50 years smoking is associated with a more than five fold increase in risk [85].

A 12 year follow-up study conducted in the US with a one million sample size in 1976 showed that heavy smokers have 1.82 times more risk of dying with CVD compared to non-smokers [86]. The Whitehall study in 1983, with a 3 year follow-up, showed that those who smoke more than 2 cigarettes per day (more than 20mg of carbon monoxide) have 1.47 times higher risk of CHD mortality compared to smokers who smoke less than 2 cigarettes per day [87]. A cohort study conducted on 17,475 male British civil servants aged between 40-64 and 8,089 male British residents aged between 35-69 showed that the risk of CHD in men with smoking 1-9 cigs/day, 10-19

cigs/day and ≥20 cigs/day is 1.23, 1.72, and 1.89 respectively compared to non-smokers for a follow-up of 12 years [88].

There are several studies which have shown that smoking is a major risk factor for stroke [89-91]. The RR for stroke for those who smoked more than 20 cigarettes per day when compared with non-smokers was 1.47 [95% CI: 1.08 – 2.00] in the Renfrew/Paisley Study in Scotland [89]. Kurth et al showed that the relative risk of stroke in subjects who smoke less than 20 and more than 20 cigarettes per day was 1.64 [95% CI: 0.60 – 4.45] and 2.34 [95% CI: 1.38 – 3.96] respectively compared to non-smokers[90]. In the Caerphilly study the results showed that those who smoke more than 15 cigarettes a day have an increased risk of 82% for having ischaemic stroke [91] compared to non-smokers.

An important issue concerns the way in which the 'levels of smoking' are measured. Most of the studies' measurements were based on the questionnaire answers provided by the subjects. However, it should be noted that smokers tend to claim that they don't smoke whereas non-smokers are unlikely to pretend otherwise. There is little one could do to avoid misclassification of their smoking habits by the subjects. The effect will generally to be to under-estimate the effect of smoking and so results are likely to be conservative, in that the misclassification tends to blur the distinction between smokers and non-smokers. Even with this uncertainty in the reliability of measurement of smoking behaviour, smoking is graded as a very important predictor and cause of CVD.

## 4.2.2 The Caerphilly analysis

Smoking is not a continuous measure. It was measured using a questionnaire that divided the whole sample into nine different categories. These nine categories appear to be ordinal due to the prior beliefs for predicting the outcome. At the time of the data collection, men answered their smoking habit from the choices of, never smoked, given up smoking before 10 years, between 5-9 years, between 1-4, given up in the last year, cigar/pipe smokers, 1-14 cigarettes/day, 15-24 cigarettes/day and > 24 cigarettes/day. The bar chart in Figure 4.11 shows the frequencies of the nine different categories of smoking in the phase two of the Caerphilly cohort.

**Figure 4.11**
**Bar Chart of Smoking variable in Phase 2**
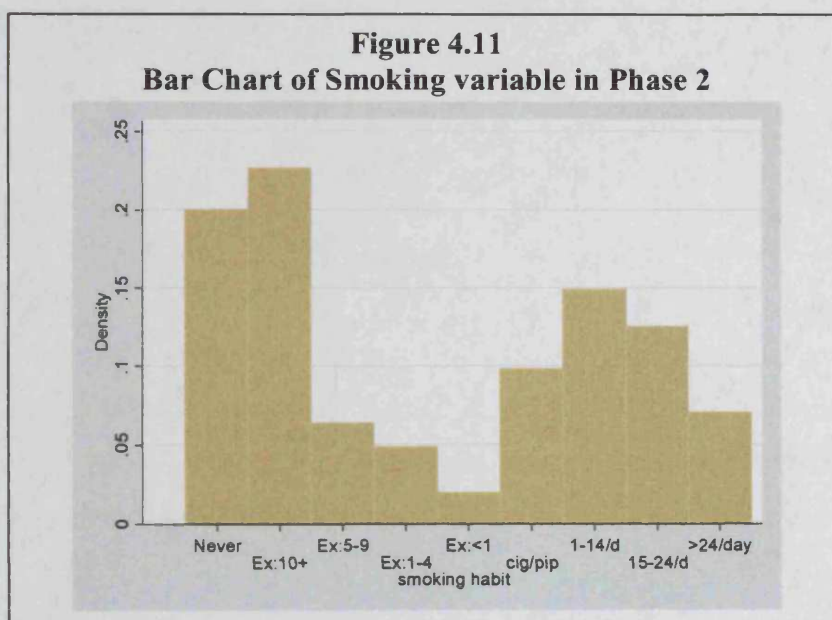


Figure 4.12, with 2 plots, shows the probability of having an MI or stroke for the different groups.

## Figure 4.12
### Probability of MI and stroke respectively for all smoking categories
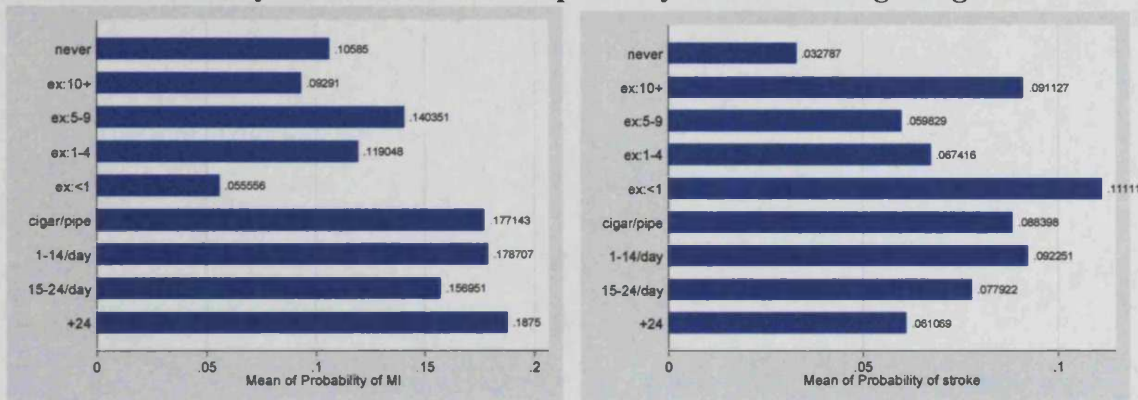


The risk of MI for ex-smokers less then one year is less than the risk of MI for men who were ex-smokers and have not smoked for between 5-9 years. One would expect these to be reveresd. At phase 2 of the study, very few subjects (1.94%) were ex-smokers for less than a year. Due to the small numbers in some categories the data was re-grouped into a smaller number of coherent categories.

## Figure 4.13
### Probability of MI and stroke respectively for three smoking categories
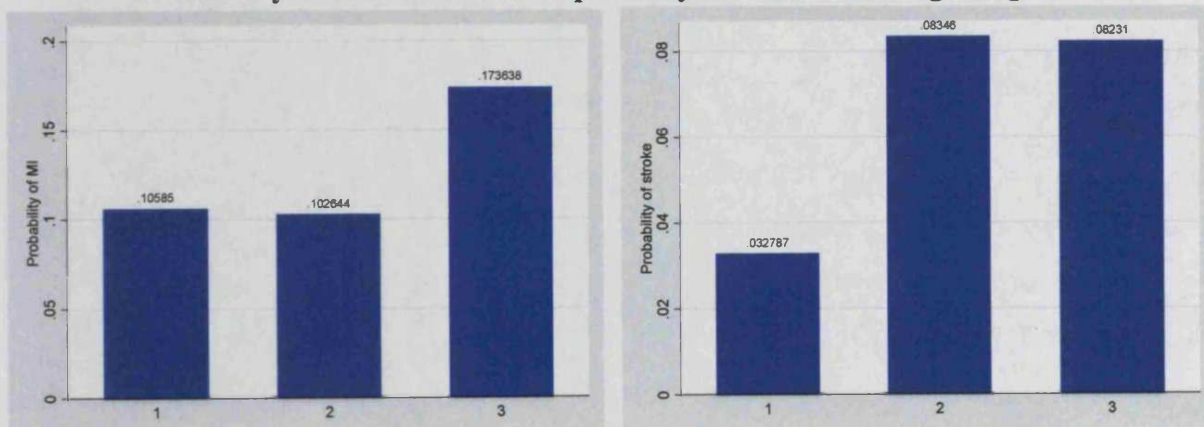


1 = non-smokers
2 = ex-smokers
3 = current smokers

The risk of MI for current smokers is higher than for either of the other categories. The risk of stroke appears to be very similar for ex and current smokers compared to the risk of stroke for non-smokers.

61

### 4.2.2.1 Smoking and MI

The odds of current and ex-smokers compared to non smokers estimated from the logistic regression model considering events occurring at any time in the follow-up period are given in the following Table 4.5.

**Table 4.5 – Logistic regression of Smoking and MI**

| MI | Odds Rat | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| Ex-smokers | 0.966 | 0.208 | 0.873 | 0.634 | 1.473 |
| Cur. Smokers | 1.773 | 0.347 | 0.003 | 1.210 | 2.604 |

The results show that there is no significant difference between ex-smokers and non-smokers for having an MI. The effect of smoking on MI is significant when current smokers and non smokers are compared. The hazard ratios are given in Table 4.6 for a corresponding survival analysis.

To show how the risk changes in each category of redefined smoking data, Kaplan-Meier survival plots are shown in Figure 4.14.



**Figure 4.14**
**Kaplan Meier survival estimates for MI with smoking behaviour**
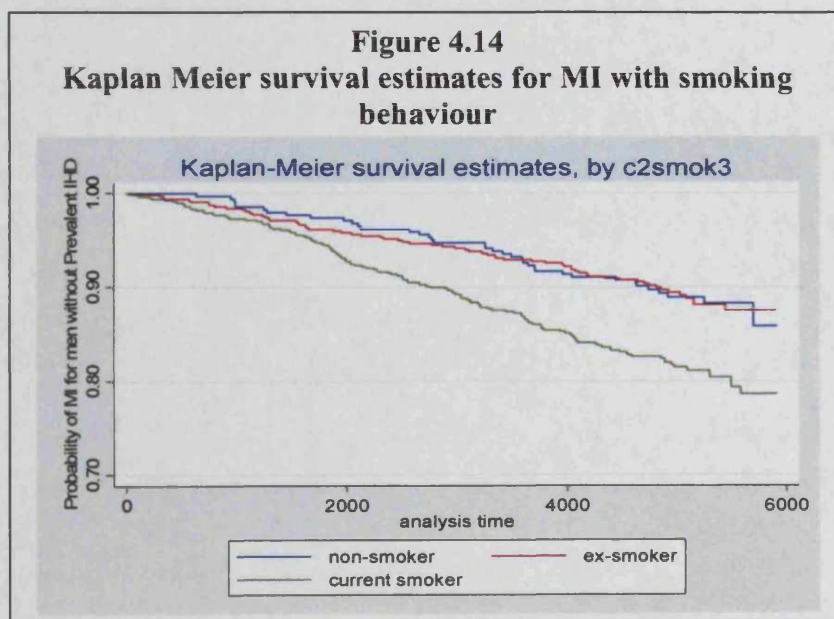
**Table 4.6 – CPHM of Smoking and MI**

| MI | Haz Rat | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| Ex-smokers | 1.013 | 0.206 | 0.948 | 0.680 | 1.511 |
| Cur. Smokers | 1.844 | 0.338 | 0.001 | 1.287 | 2.642 |

Figure 4.14 suggests that there is no difference between non-smokers and ex-smokers for having an MI. However, it appears that the smokers have a significantly greater risk of an MI then both the other categories, with the survival curves steadily diverging.

### 4.2.2.2 Smoking and stroke

The odds ratios of the ex-smokers and current smokers compared to the non-smokers are given in the following Table 4.7.

**Table 4.7 – Logistic regression of Smoking and stroke**

| MI | Odds Rat | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| Ex-smokers | 2.687 | 0.871 | 0.002 | 1.418 | 5.077 |
| Cur. Smokers | 2.652 | 0.849 | 0.002 | 1.409 | 4.947 |

These results show that the risk of stroke is almost the same for ex-smokers and current smokers compared to non-smokers. The hazard ratio obtained by Cox's proportional hazard model comparing non-smokers and ex-smokers is 2.70 [95% CI: 1.45 – 5.06, p = 0.002 ] whereas, comparing the non-smokers with smokers, the hazardss ratio was 2.85 [95% CI: 1.54 – 5.84, p = 0.002]. Kaplan-Meier survival estimates are given in the Figure 4.15.



**Figure 4.15**
**Kaplan Meier survival estimates for smoking behaviour for stroke**
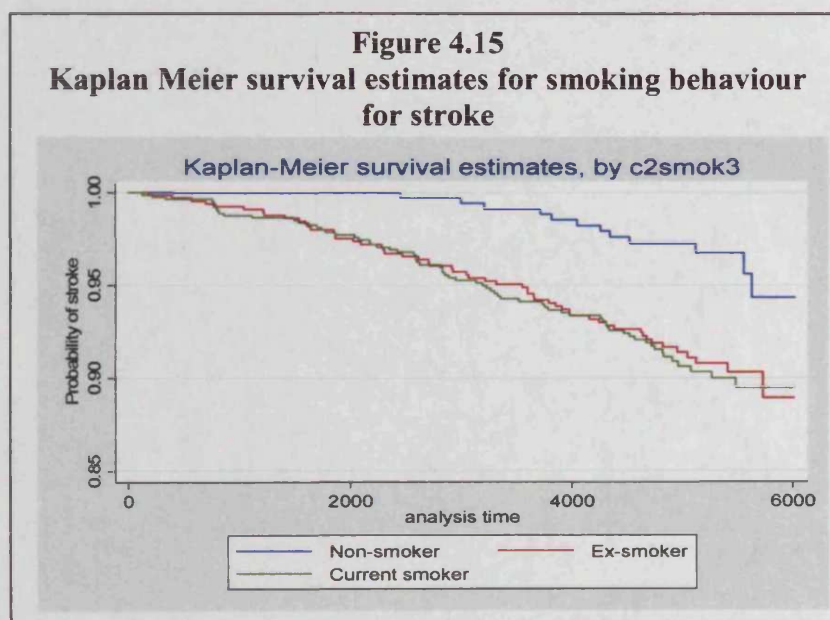
The Kaplan-Meier survival plots show that the risk of stroke is about the same in ex and current smokers. Nevertheless, there appears to be a significant difference between non-smokers and the remaining two smoking categories.

## 4.2.3 Conclusions

When the odds of having an MI with smoking in the Caerphilly men is compared with the odds ratios quoted in the Sir Richard Doll's study [92], they appear to be similar. The OR of MI for the smoking Caerphilly men when compared with non-smokers is 1.844. The OR quoted in the Doll's study was 1.9, very close to the Caerphilly results. It is also important to note that Doll's study compares different groups of smoking categories as it has a bigger dataset. In the Caerphilly analysis, due to the small sample size, there were only three categories considered. A study in the US gave an odds ratio of 1.82 when smokers and non smokers were compared for having an MI [86], which is similar to the Caerphilly results. The slight variations in the risks may be attributed to the difference in the location, sample size, outcome and follow-up period of the study and several other factors.

When the literature from other studies was compared with the Caerphilly analysis, the risk of stroke again appeared to show similar results to those of the literature. It appears from the results that the risk of MI for non-smokers and ex-smokers is similar whereas the risk of stroke for ex-smokers and smokers is similar. This suggests that the risk of cardiovascular disease for those who quit smoking falls fairly rapidly to that of non-smokers whereas the risk of a stroke is maintained at the level of current smokers. Information regarding the partners' smoking habits was not collected in the Caerphilly study and therefore the effect of passive smoking on non-smoking Caerphilly men cannot be estimated.

## 4.3 Leisure activity

### 4.3.1 Literature review

A low level of leisure activity is a major risk factor for ill health and mortality from all causes. People who do not do sufficient leisure activity have a greater risk of cardiovascular disease [93]. Being physically active is also thought to improve mental health and reduces other risk factors such as obesity, high blood pressure and high blood cholesterol.

For middle-aged men, regular leisure-time physical activity is associated with lower all-cause mortality and with lower morbidity and mortality from cardiovascular disease [94-100]. There is also evidence that older people benefit from physical activity [101, 102]. Some researchers have raised concerns about possible hazards associated with vigorous activity. The BRHS reported increased CHD events in those who reported vigorous compared with moderate levels of physical activity [98]. This increased risk was seen for CHD but not for stroke [100]. The excess risk was confined to men with hypertension who were vigorously active [103]. The British regional heart study in 1998 reported that the risk of age adjusted CVD mortality for light activity=0.54(0.32-0.90), moderate activity=0.26(0.12-0.58) and heavy=0.43(0.25-0.73) compared with very little activity [104]. Physical inactivity showed a RR = 1.9 compared with moderate physical activity for having an MI in a study by James C et.al in 60-70 years old Harvard alumni [105]. A study by Fang et.al [106] showed hazard ratios of low physical activity of 1.24 (CI=1.06-1.44) compared with moderately active in 9790 subjects aged 25 to 74 years and followed-up for 17 years.

A Finnish study on 25-74 year old 18,892 men and women showed that moderate and high physical activity have RRs of 0.57, 0.52 respectively for having a first stroke compared to low activity [107]. A meta analysis of 31 studies for the risk of stroke by physical inactivity showed that moderately intense physical activity had a RR of 0.64(95% CI: 0.48–0.87) and moderate physical activity RR of 0.85(95% CI: 0.78–0.93) compared with inactivity for the risk of first stroke [108].

## 4.3.2 The Caerphilly analysis

Data were collected using questions derived from the Minnesota Leisure Time Physical Activity (MLTPA) questionnaire. It is used to estimate the energy expenditure expressed as an activity index (AI) in kcal/day from a record of leisure activity in the Caerphilly cohort [109]. Participants were asked whether they had undertaken a variety of activities (including walking, sports, house maintenance and gardening) during the last 12 months and, if so, the frequency and average duration of this activity. Each activity was assigned an intensity code and energy expenditure was calculated (as an activity index), in total, and for each level of intensity (light, moderate, and heavy) [110]. For example, light AI was defined by summing those activities having intensity codes 2.0, 2.5, 3.0, 3.5, and 4.0 (walking, bowling, sailing). Moderate AI was obtained by summing activities with intensity codes of 4.5, 5.0, and 5.5 (golfing, digging, dancing) and heavy AI was defined by summing all activities having intensity codes > 6.0 (climbing stairs, swimming, jogging). Using all these codes and considering their intensities, energy expenditure of individuals was calculated. To make Caerphilly results comparable with other studies round the world, energy expenditure data was divided into three equal thirds i.e., light, moderate and heavy activity, as was done for the Caerphilly results published elsewhere [110].

## 4.3.2.1 Leisure activity and MI

The odds ratios of moderate and heavy activity compared to light activity for having an MI are given in Table 4.8.

**Table 4.8 – Logistic regression of leisure activity and MI**

| MI | Odds Rat | Std Err | P>|z| | 95 % CI | |
|----|----------|---------|-------|---------|---|
| Leisure 1 | 0.662 | 0.111 | 0.014 | 0.477 | 0.919 |
| Leisure 2 | 0.645 | 0.108 | 0.009 | 0.465 | 0.895 |

**where leisure 1 and leisure 2 are moderate and heavy leisure activity compared with light leisure activity**

As in the various studies in the literature review, the Caerphilly data shows that the odds ratios of moderate and heavy activities show a significant reduction in risk of an MI when compared with men with light leisure activity. Cox's proportional hazard model showed very similar results with the hazard ratios of 0.66 [95% CI: 0.49 – 0.89] and 0.66 [95% CI: 0.49 – 0.90] ($p$ values and standard error for both groups

were 0.007, 0.008 and 0.102, 0.103) for comparing men with no leisure activity and moderate and high leisure activities respectively. Kaplan-Maier survival curves of these three groups are given in Figure 4.16 illustrate this relationship between the time and the risk of leisure activity in the three different groups.

**Figure 4.16**
**Kaplan Meier survival estimates for MI with leisure activity**
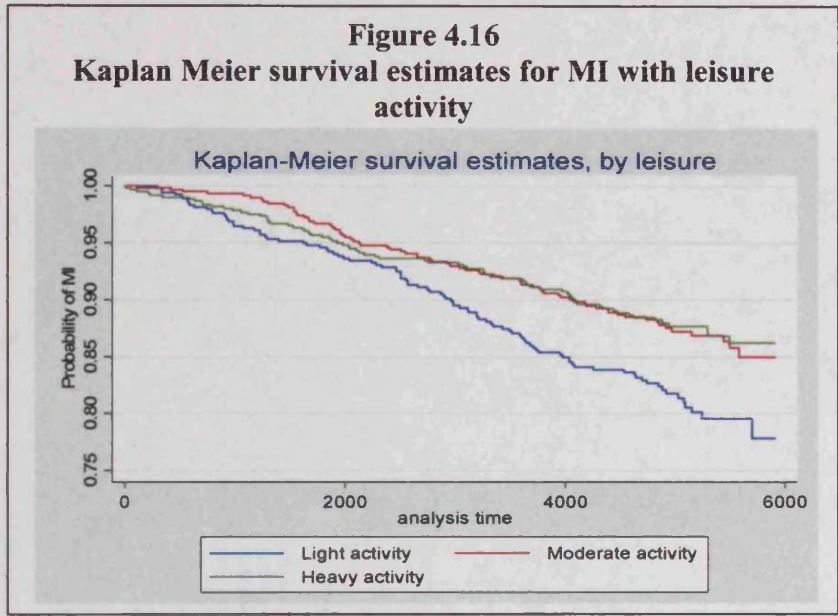


### 4.3.2.2 Leisure activity and Stroke

The odds ratios of moderate and heavy activity compared to light activity for having a stroke is given in Table 4.9.

**Table 4.9 – Logistic regression of Leisure activity and stroke**

| MI | Odds Rat | Std Err | P>\|z\| | 95 % CI | |
|---|---|---|---|---|---|
| Leisure 1 | 0.707 | 0.149 | 0.126 | 0.453 | 1.103 |
| Leisure 2 | 0.924 | 0.196 | 0.707 | 0.610 | 1.399 |

**where leisure 1 and leisure 2 are moderate and heavy leisure activity compared with light leisure activity**

From the above table, it is clear that different levels of leisure activity do not show any significant difference from no leisure activity for having a stroke. When Cox's proportional hazard model was used to estimate the hazard ratio of men with moderate and heavy activity compared to light activity men for having a stroke, the

hazard ratios were 0.68 [95% CI: 0.44 – 1.05] and 0.90 [95% CI: 0.61 – 1.34] ($p$ values and standard error for both groups were 0.081, 0.610 and 0.149, 0.183). There appears to be statistically no significant difference between any activity levels of men for the risk of stroke.

## 4.3.3 Discussion

Heavy and moderate leisure activity have been shown to be associated with significantly reduced risks, when compared to low leisure activity, of having an MI These Caerphilly results suggest that there is at most a small difference between the effects of heavy and moderate leisure activity. These results are consistent with those reported in the literature review Though some studies reported that the risk of MI is increased in men who undertake vigorous leisure activity compared to moderate activity [111], the Caerphilly study showed no significant difference between those two groups. The British regional heart study reported the risk of age adjusted CVD mortality for light activity=0.54(0.32-0.90), moderate activity=0.26(0.12-0.58) and heavy=0.43(0.25-0.73) compared with very little activity [104]. Though the Caerphilly results are not as highly significant as these results, there appear to be some similarities. This might be due to several reasons such as difference in populations, relatively low power for identifying such a difference due to the sample size, measuring procedures and all other potential factors. An important point that should be considered is that the overall message is one of health benefits of increased leisure activity, whether moderate or vigorous.

The risk of stroke in men who undertake moderate or high leisure activity compared with low leisure activity appears to show no statistical significance, in contrast to the results of a meta analysis [108]. The fact there were few events of stroke in the Caerphilly men could be a possible explanation, or it could be that there really is no difference between the different activity groups in respect of having a stroke. Though the literature review appears to show significantly decreased risks of stroke with leisure activity, the data analysis carried out in the previous section did not show a statistically significant difference. This might be due to low numbers of subjects.

# 4.4 Regression dilution bias (RDB)

There is a need for two measurements of variables that are taken at two different points of time to quantify the measurement error of the variable. Alcohol consumption was measured in exactly the same way in subsequent phases ($3^{rd}$ and $4^{th}$) of the Caerphilly cohort study. The same questionnaire was used on each occasion, which made it easy to calculate the final alcohol consumption score for each individual. However, leisure activity was measured using different questionnaires in the subsequent phases, which makes it difficult to compare these with the $2^{nd}$ phase measurements. The one used in the $2^{nd}$ phase was the thorough one (MLTPA) whereas the rest of the phases had just one question asking the subjects whether they were doing any leisure activity or not.

## 4.4.1 Regression Dilution Bias of alcohol consumption

Measurement error in alcohol consumption can be estimated with the methods explained in chapter 3. The information required for calculating error factor [112] is given in Table 4.10.

**Table 4.10: Summary statistics for regression dilution bias for MI and alcohol consumption**

|       | Obs  | Mean   | Std.Dev | Min     | Max    |
|-------|------|--------|---------|---------|--------|
| a1    | 1418 | 106.30 | 82.13   | 2.66    | 417.35 |
| a2    | 1418 | 99.30  | 81.04   | 2.66    | 401.58 |
| a3    | 1418 | 94.60  | 79.34   | 2.66    | 390.17 |
| b1    | 1418 | 176.68 | 156.19  | 0.87    | 839.41 |
| b2    | 1418 | 163.75 | 153.39  | 0.87    | 802.55 |
| b3    | 1418 | 154.99 | 149.45  | 0.87    | 775.99 |
| a2-a1 | 1418 | -7.00  | 58.57   | -308.87 | 288.03 |
| a3-a1 | 1418 | -11.70 | 66.73   | -321.44 | 262.86 |
| b2-b1 | 1418 | -12.93 | 114.10  | -644.76 | 608.74 |
| b3-b1 | 1418 | -21.69 | 129.37  | -669.76 | 520.08 |

where $a_i = \text{logalc}^3(i)$, $b_i = \text{logalc}^3(i)*\log(\text{logalc}(i))$

$\text{logalc}(i)=\log(4+\text{alc in phase } (i+1))$ where $i = 1,2 \,\&\, 3$

There were 1418 men with all the information available for estimating the error coefficients. Estimation of coefficient of measurement error was carried out using the model fitted for all men who were present in all phases of the cohort. The relationship between alcohol consumption (logalc) and MI was considered to illustrate this example. The fractional polynomial regression equation estimating the relationship

between MI and logalc can be represented in the form of an equation as follows (similar to equation 4.2):

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \times \text{logalc}^3 + \beta_2 \times \text{logalc}^3 \log(\text{logalc}) \dots\dots(4.5)$$

where $\beta$s are estimated when the log likelihood function is maximised. The adjusted coefficients takes the variations with in subjects into account are denoted by $\hat{\beta}$; their values are given in Table 4.11.

**Table 4.11: $\beta$s adjusted with measurement**

|  | $\beta_0$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|
| Base Line $\beta$s (phase 2) | -2.2224 | -0.0122 | 0.0067 |
| $\beta$s adjusted for Phase 3 | -1.9034 | -0.01823 | 0.0101 |
| $\beta$s adjusted for Phase 4 | -1.8748 | -0.01961 | 0.0109 |

The Figure 4.17 shows the relationship between the risk of MI and alcohol consumption when using the unadjusted and adjusted coefficients.



**Figure 4.17: Predicting MI for all men using baseline and RDB adjusted models**
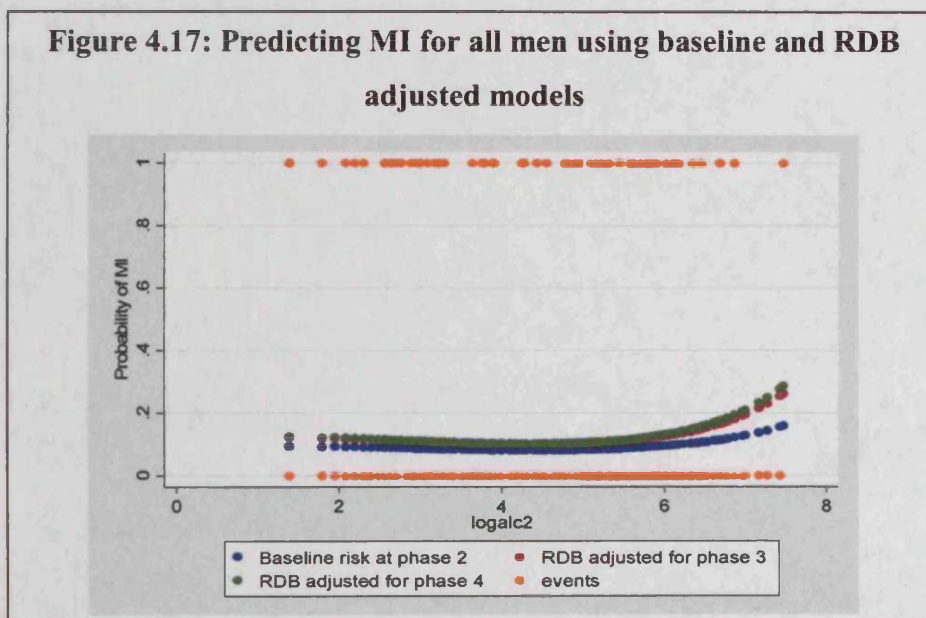
This shows that the model adjusted for variations shows a stronger association between the risk of MI and alcohol consumption but the interpretation must be cautious because of the long time intervals between the measurements.

## 4.4.2 Regression Dilution Bias for Smoking

Repeated measurements were made in the subsequent phases of the cohort on the smoking behaviour of the Caerphilly men. Assuming that the smoking status has not changed from phase 2 to phase 3 of the Caerphilly study, the sensitivity and specificity for the smoking status in phase 2 and 3 was 0.76 and 0.97 respectively. Using these sensitivity and specificity values along with the tabulated values of smoking status and MI in the phase 2 of the Caerphilly study in Table 4.12,

**Table 4.12: Smoking behaviour in Phase 2**

|  | MI | |
|---|---|---|
| Smoking Status↓ | 1 = event | 0 = no event |
| Smoker | 67 | 488 |
| Non-smokers | 70 | 727 |
| Total | 137 | 1215 |

the adjusted odds ratio is given by the formula

$$\hat{OR} = \frac{(a + N_1 \times (Sp-1))(N_0 \times Se - c)}{(c + N_0 \times (Sp-1))(N_1 \times Se - a)} = 3.76$$

Using the formula for standard error of the adjusted OR in equation 3.38, the 95% CI is (1.50, 9.40). This adjusted odds ratio is calculated on an assumption that that there is no change in the smoking habit from phase 2 to phase 3. However, this is far from true. There is will some changes in smoking habit as some of them would quit smoking. Therefore, assuming that smoking habit does not change from phase 2 to phase 3 is not going to be right.

According to a meta-analysis based on 10 studies, the sensitivity and specificity of smoking behaviour measured by superior cotinine validation methods showed the sensitivity ranged from 0.82 to 1 and specificity ranged from 0.91 to 1 [57].

If it is assumed that the sensitivity and specificity is 0.90 in the Caerphilly analysis instead the one obtained from the Caerphilly data, the adjusted odds ratio was 1.56

(95% CI: 1.003, 2.438). If the sensitivity and specificity were assumed to be 0.80, the adjusted odds ratio was 1.84 (95% CI 1.01, 3.33).

Since the method is depended on the sensitivity and specificity of smoking habit in $2^{nd}$ and $3^{rd}$ phase and assumed that there is no change in the smoking habit between the phases, these results should not be taken seriously. By using the sensitivity and specificity from the meta-analysis study showed that the adjusted odds ratio is not that high compared to the adjusted odds ratio calculated using the sensitivity and specificity of the Caerphilly data. Therefore, more caution is needed in using these results.

The other method mentioned in the statistical methods chapter (section 3.7.3 Case 2) is was not used to this data because the U and V values were not in the feasible regions of the model.

## 4.5 Conclusions

This chapter performs an analysis of biological risk factor and concludes that there is no significant association between alcohol consumption and the risk of MI. The risk of stroke with an increase in alcohol consumption increases as alcohol consumption increases.

There are several methods to measure the regression dilution bias for the data assumed to be normally distributed. To make things easier, alcohol consumption was taken on the log scale after adding 4 to the actual consumption scores to avoid taking the logarithm of teetotallers (i.e., alcohol consumption = 0). This transformation converts alcohol consumption score into a distribution that was approximately normal. Regression dilution bias was estimated with 'logalc'. The other major assumption is about the direction of the measurement error that was assumed to be on both directions of the scale. However, this may not be true in the case of alcohol consumption. For those who quote themselves as teetotallers, the measurement error cannot be less then zero.

Smoking analysis showed that the risk of MI is similar for ex-smokers and non-smokers. Smokers showed a very high risk of MI compared to the non-smokers. Current and ex-smoking showed an increased risk when compared with non-smokers for having stroke. These stroke results did not show similar results as in the studies from the literature review may be because of low sample sizes.

Leisure activity showed an increased risk of MI when moderate and high activity groups were compared with the low activity group. However, it was not evident that high leisure activity had any negative effects i.e., the risk of MI was significantly lower for high activity men compared to the low activity men. The literature review suggested that the high leisure activity has an increased risk of MI but the Caerphilly analysis did not confirm this.

# Chapter 5

# 5 Biological factors

Previous chapter has dealt with the behavioural factors that may affect the risk of MI or stroke. This chapter analyses with the biological risk factors such as blood pressure, cholesterol levels and body mass index.

## 5.1 Literature review to blood pressure

In older men with systolic hypertension, the pressure on the arterial wall is significantly greater than in younger men. This is thought to be due to an increased rigidity of the arterial wall. Mechanical changes in the arterial wall, such as atherosclerosis, increase overall arterial rigidity. This decrease in compliance results in higher blood pressures as the large vessels become less able to reduce the pressure generated by the left ventricle.

The importance of high blood pressure as a risk factor of VAD has been demonstrated by several studies. It is also been shown that, compared to individuals with normal blood pressure, those with high blood pressure more commonly possess other VAD risk factors. The aim here is to show the importance of blood pressure as a risk factor for VAD.

The Framingham study [113] showed a 41% and 34% greater risk of cardiovascular disease for an increase of 1 standard deviation of SBP and DBP respectively, for men aged between 35-64. Wilhelmsen et al [114] showed those who have SBP≥176 units have a RR of 1.92 for an MI event compared to those who had SBP<145, when adjusted for various standard risk factors. These two studies together with many others, have demonstrated the importance of SBP as a risk factor to CVD.

A study followed 11,150 male physicians aged 45-60, with no history of CVD or antihypertensive treatment, for 2 years. It showed that for every 10mmHG increase in SBP and DBP, the risk of CHD is increased by 31% and 46% respectively [115]. At first sight, the results of this study may suggest that the effect of DBP on the risk of VAD is greater then the effect of SBP. Since SBP has a greater range than DBP, these

results are a little misleading. A better comparison is to consider the increased risk associated with an increase of 1 SD. Following a similar argument, several other studies showed that SBP is a better predictor of CVD than DBP [116, 117]. A meta analysis by the Prospective Studies collaboration with one million adults showed that a 20mmHg increase in SBP is approximately equivalent to a 10mmHg increase in DBP in showing a two fold increase in risk for CVD [118]. The standard deviation of SBP is almost twice the standard deviation of DBP [118]. Though DBP appears to show a larger odds ratio, the actual risks of MI due to SBP and DBP are very similar.

Using data from the Framingham and Whitehall Studies, Clarke et al measured the regression dilution bias of blood pressure over 16 years and 26 years respectively [119]. They show that uncorrected associations of disease risk with baseline measurements underestimate the strength of the real associations with the usual levels of these risk factors during the first decade of exposure by about one-third, the second decade by about one-half, and the third decade by about two-thirds. Since there are three measurements of blood pressure in the CaPS, it would be interesting to investigate the effect of regression dilution on the CaPS data and compare the results with the Framingham study.

Interestingly, some researchers have suggested that the same path of physiological mechanisms for SBP and DBP results in an increase of pulse pressure, which is defined as the difference between systolic and diastolic blood pressures. Pulse pressure is thought to reflect arterial compliance, which in turn reflects the elasticity of the arterial wall. It is argued that a smaller difference between diastolic and systolic pressure indicates greater arterial elasticity as the arterial walls have absorbed more of the peak pressure due to systole. There is evidence that elevation in pulse pressure itself may be an important independent risk factor for all-cause and cardiovascular mortality, and analysis of an American study with 4736 men showed that every 10mmHg increase in pulse pressure increases the risk of heart failure by 32% [120]. Data for the Framingham Heart Study has suggested that neither SBP nor DBP is superior to PP in predicting coronary heart disease risk [121]. Other studies also showed that pulse pressure is superior in predicting heart disease than SBP or DBP [122-124]. This suggests that the pulse pressure is a better predictor of VAD.

Most of the studies had a general population as their study subjects. Some studies, however, used population samples with no previous heart disease [122]. For men with prevalent IHD, the risk may not be the same as men without prevalent IHD because antihypertensive drugs may already been prescribed for them, leading to a reduction in blood pressure. This will have the effect of reducing the estimated effect of blood pressure on the risk of heart disease. Several studies discussed the effects of antihypertensive drugs but those studies will be discussed in a literature review of intervention studies in chapter 9.

Statistical procedures used to identify and model plausible relationships between SBP, DBP, pulse pressure, total cholesterol, HDL cholesterol and BMI with MI and stroke are very similar. Since these variables are continuous, the procedure used to identify a suitable model is similar to that for alcohol consumption and MI, because alcohol consumption was considered as a continuous variable in the previous chapter.

## 5.2 Analysis of blood pressure

For the sake of clarity, the analyses for SBP, DBP and PP are presented separately.

### 5.2.1 Systolic Blood pressure (SBP)

Systolic blood pressure was measured in all four phases of the Caerphilly cohort. As this thesis is based on data from the cohort after it had been augmented in phase 2, the analysis includes measurements from the $2^{nd}$, $3^{rd}$ and $4^{th}$ phases only. The distributions of SBP in the three phases are given in the Figure 5.1. The purpose of these histograms is primarily to visualise the distribution of the data in order to use this information for further analyses and secondly to understand and possibly explain if there are any changes in SBP distributions over the period of three phases of follow-up.



Figure 5.1 – Histograms of SBP in $2^{nd}$, $3^{rd}$, and $4^{th}$ phases
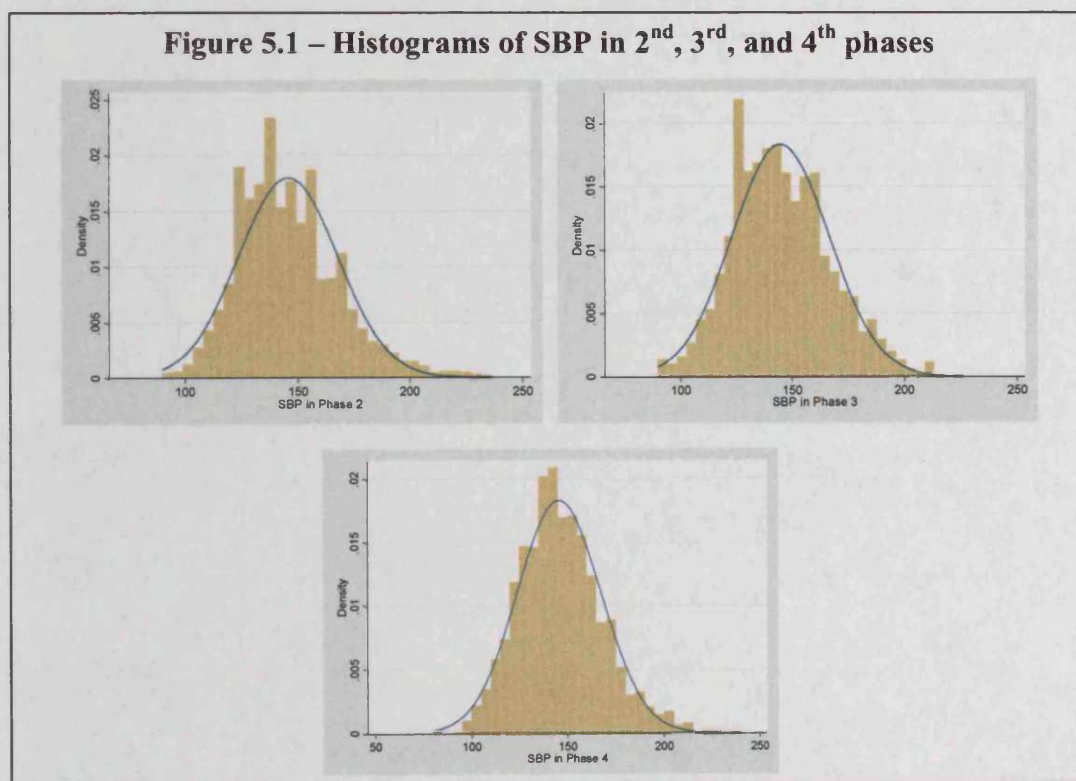
In the above Figure 5.1, all the three plots show fairly normal distributions. The summary statistics of these three measurements are given in the Table 5.1.

**Table 5.1 – Summary statistics of SBP in 2$^{nd}$, 3$^{rd}$ and 4$^{th}$ phases**

| Variables | Obs | Mean | Std. Dev | Min | Max |
|-----------|-----|------|----------|-----|-----|
| SBP in Ph 2 | 1839 | 146.58 | 22.71 | 90 | 236 |
| SBP in Ph 3 | 1610 | 144.85 | 22.44 | 74 | 226 |
| SBP in Ph 4 | 1416 | 145.84 | 22.15 | 80 | 240 |

After excluding the missing values in the SBP, the total number of observations available in the end phase is given for each phase in Table 5.1. From the above table, it can be observed that the standard deviations and means of SBP measurements in three phases hardly differ. If subjects with all three measurements are considered, the mean and standard deviations of SBP for 1352 men in the three phases are 146.1 (SD=22.4), 144.9 (SD=21.9) and 146.2 (SD=21.9) respectively. This supports the view that there was little difference in mean systolic blood pressure between phases.

### 5.2.1.1 SBP and MI

### 5.2.1.1.1. Linear logistic regression

Lowess smoothing is used to give a graphical approximation of the relationship between SBP and MI. Figure 5.2 shows the relationship between MI and SBP.



**Figure 5.2: Predicting MI using baseline SBP by logistic regression**
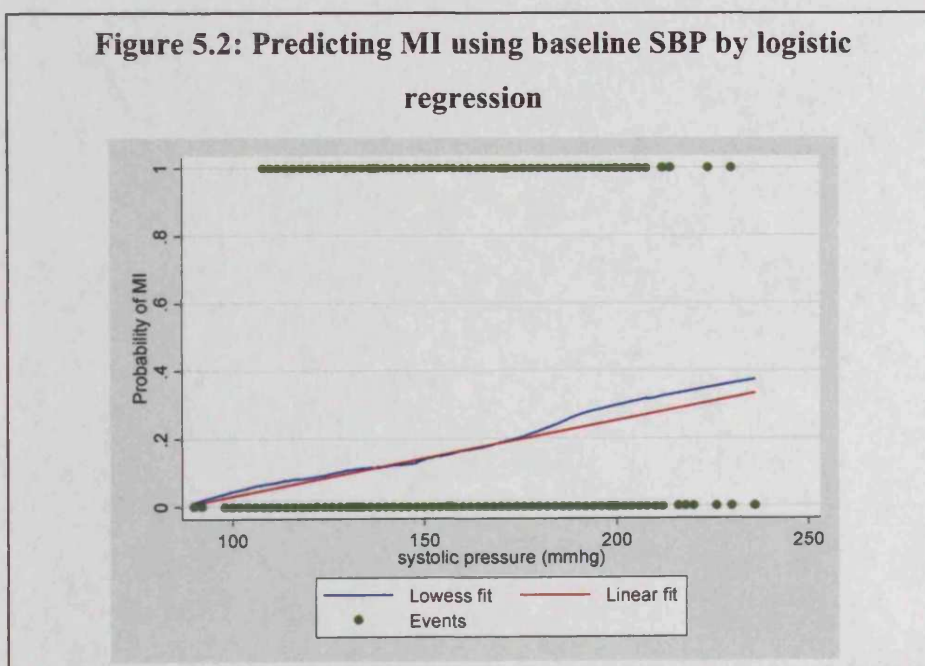
The regression coefficient table of the linear logistic regression is given in Table 5.2.

**Table 5.2 – Logistic regression of baseline SBP and MI**

| MI | Coef | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| SBP | 0.018 | 0.003 | 0.000 | 0.012 | 0.024 |
| Const | -4.523 | 0.453 | 0.000 | -5.419 | -3.638 |

Substituting these values in the equation, the regression equation is as follows.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.018SBP - 4.523 \dots\dots\dots(5.1)$$

where $p$ is the probability of MI.

The above equation means that the log odds of MI increases by 0.018(1.8%) for every 1 mmHg increase in SBP. However, this is not helpful to understand how much the risk of MI increases by SBP because the one unit increase in SBP is very small. For a 10 mmHg increase in SBP, there is an increased odds of 19.4% for having an MI. Fractional polynomial regression model was also used, but showed no significant improvement over the linear logistic model (gain in deviance 0.19).

### 5.2.1.1.2 Cox's proportional hazard model (CPHM)

The hazard ratios estimated by the CPHM are given in Table 5.3 and the hazard function is shown in the equation 5.2.

**Table 5.3 – CPHM of baseline SBP and MI**

| MI | Haz.r | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| SBP | 1.017 | 0.003 | 0.000 | 1.012 | 1.022 |

The information in the above table can be interpreted to the equation as follows

$$\log\left(\frac{h\{(t),(SBP)\}}{h_0(t)}\right) = 0.017 \times SBP \quad \dots\dots\dots\dots\dots\dots\dots (5.2)$$

CPHM showed a significantly increased risk of MI for every unit increase in SBP. The gain with a fractional polynomial model was not significant (deviance gain 0.17) suggesting that a linear model is adequate.

### 5.2.1.2 SBP and Stroke

### 5.2.1.2.1 Logistic regression

To have a basic idea of how the relationship between SBP and stroke looks, a Lowess smoothed plot is shown in Figure 5.4 below along with the appropriate model fitted to it.



**Figure 5.4: Relationship between SBP in phase two and stroke for all men**
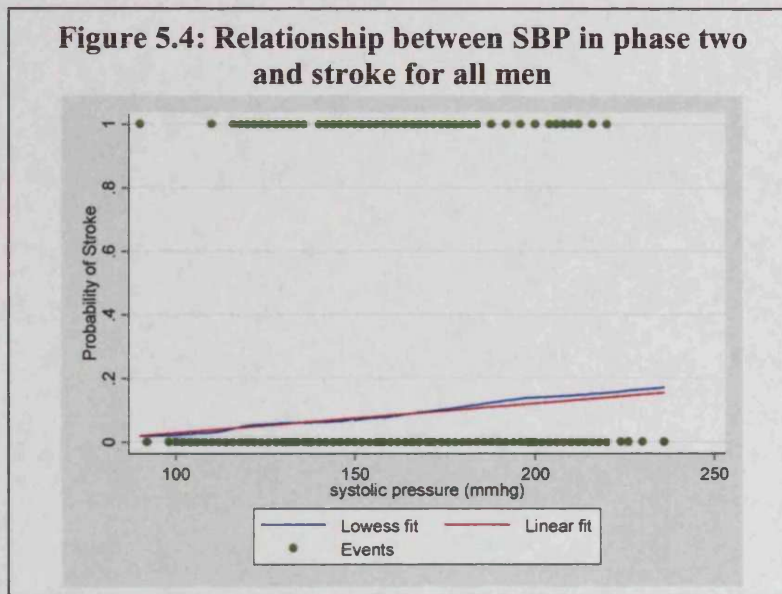
The regression coefficient table for baseline SBP and stroke is given in Table 5.4. The odds ratio is 1.013 with 95% confidence interval as [1.006, 1.021]. For an increase of 10 mmHg in SBP, the odds of a stroke increases by 14%.

**Table 5.4 – Logistic regression of baseline SBP and stroke**

| Stroke | Coef | Std Err | P>|z| | 95 % CI | |
|--------|------|---------|-------|---------|---|
| SBP | 0.013 | 0.004 | 0.000 | 0.006 | 0.021 |
| Const | -4.535 | 0.579 | 0.000 | -5.671 | -3.400 |

Table 5.5 can be represented in an equation as follows

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = -4.535 + 0.013 \times SBP \dots\dots\dots\dots\dots (5.3)$$

where $p$ is the probability of having a stroke.

From the above table and equation, SBP appears to be a highly significant risk factor for stroke. The deviance gain using fractional polynomial regression was not significant (Gain=0.14).

**5.2.1.2.2 Cox's proportional hazard model**

Cox's proportional hazard model also gave hazard ratios very similar to the odds ratios (1.013, 95% CI [1.01, 1.02]) given by the linear model. The hazard ratio can be rewritten in the form of an equation follows

$$\log\left(\frac{h\{(t),(SBP)\}}{h_0(t)}\right) = 0.013 \times SBP \quad \text{................................ (5.4)}$$

A fractional polynomial model showed no significant improvement in the deviance (gain=0.12). This shows that the linear model is adequate for these data.

## 5.2.2 Diastolic Blood pressure

Diastolic blood pressure was measured along with SBP in all four phases of the Caerphilly cohort and data from phases 2, 3 and 4 are used here. Figure below shows the distributions of the three measurements taken.



**Figure 5.5 – Histograms of DBP in 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> phases**
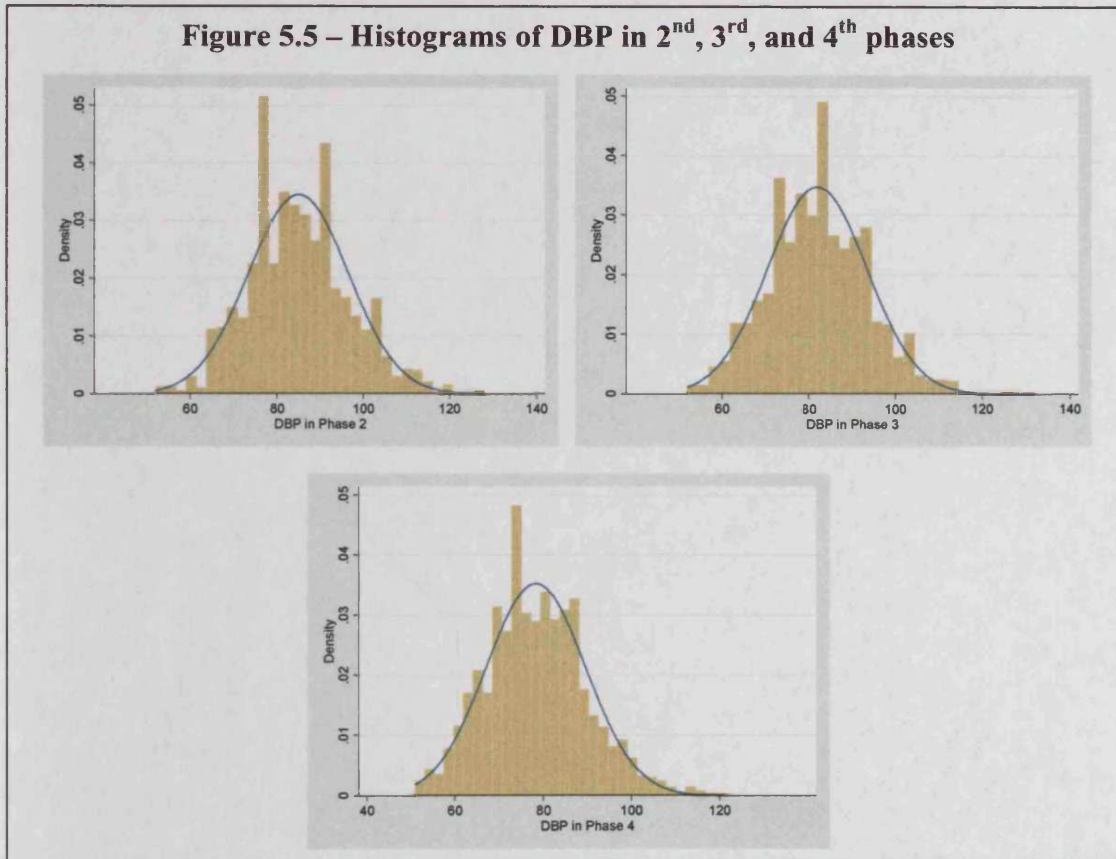
The means and standard deviations of DBP measured in three phases are given in Table 5.5.

**Table 5.5 – Summary statistics of DBP in 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> phases**

| Variables | Obs | Mean | Std. Dev | Min | Max |
|---|---|---|---|---|---|
| DBP in Ph 2 | 1836 | 84.83 | 11.84 | 52 | 144 |
| DBP in Ph 3 | 1598 | 81.88 | 11.65 | 52 | 132 |
| DBP in Ph 4 | 1397 | 78.63 | 11.43 | 52 | 121 |

From the above histograms, it is evident that DBP too follows nearly a normal distribution but the means fall slightly from 2<sup>nd</sup> phase to the 4<sup>th</sup> phase. There were 3

men with DBP lower than 50 and 8 men with DBP more than 120 in phase 2. These three very low values, in particular, had a considerable effect on the modelling. When those eleven observations were excluded from the analysis, the mean of the DBP measured in the 2$^{nd}$ phase of the cohort dropped slightly. However, the reduction over time was also shown when means and standard deviations of men with all three measurements were considered (means 84.9, 82.1 and, 78.6 respectively). The standard deviation for all three phases was 11.4. So the mean DBP fell substantially much grater then the changes in SBP.

### 5.2.2.1 DBP and MI

### 5.2.2.1.1 Linear logistic regression

A Lowess smoothed plot for the relationship between MI and DBP is shown along with a linear logistic fit in Figure 5.6 below.



Figure 5.6: Relationship between DBP in phase two and MI
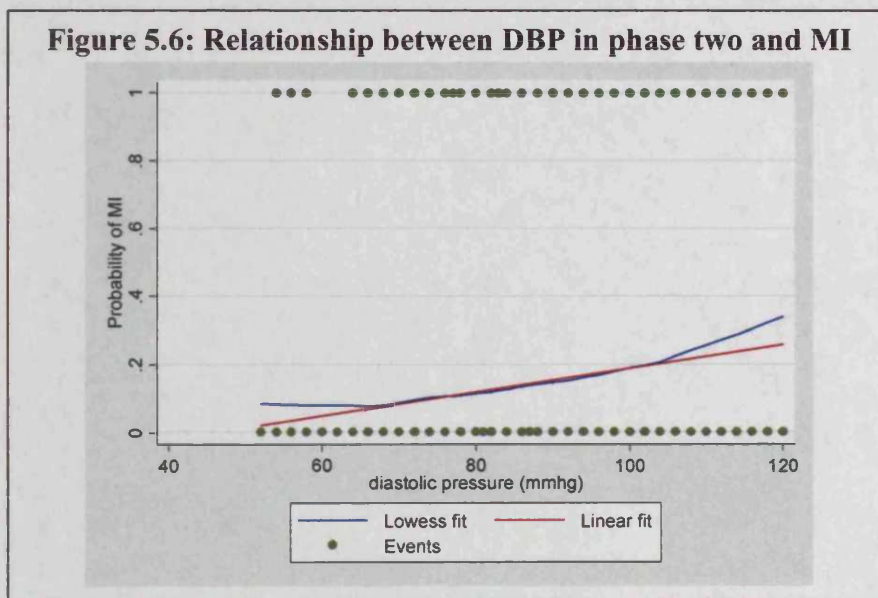
The relationship appears to be non-linear. However, the gain of a fractional polynomial model over the linear fit is only 0.88 with 4 degrees of freedom and that is statistically not significant. Furthermore, none of the literature review showed a non-linear relationship between DBP and MI. Therefore, a linear model is appropriate.

The regression coefficient table for baseline DBP and MI is given in Table 5.6.

**Table 5.6 – Logistic regression of baseline DBP and MI**

| MI | Coef | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| DBP | 0.029 | 0.006 | 0.000 | 0.017 | 0.041 |
| Const | -4.361 | 0.542 | 0.000 | -5.423 | -3.299 |

The results are statistically significant and show that DBP is a very significant predictor of MI.

Table 5.6 can be represented in an equation as follows

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = -4.361 + 0.029 \times DBP \quad\text{........................}\quad (5.5)$$

where $p$ is the probability of having MI.

To an increase of 10 mmHg in DBP results in an increase in odds of MI of 34%.

**5.2.2.1.2 Cox's Proportional hazard model**

The hazard ratios for every 1mmHg increase in DBP are given the Table 5.7.

**Table 5.7 – CPHM of baseline DBP and MI**

| MI | Haz ratio | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| DBP | 1.028 | 0.006 | 0.000 | 1.016 | 1.039 |

This can be rewritten in the form of an equation with time covariate as follows

$$\log\left(\frac{h\{(t),(DBP)\}}{h_0(t)}\right) = 0.028 DBP \quad\text{.............................................}\quad (5.6)$$

The gain in deviance by the fractional polynomial Cox's model over the above model is 1.61 and that is statistically not significant.

### 5.2.2.2 DBP and Stroke

### 5.2.2.2.1 Logistic regression

A Lowess smoothed plot for the relationship between stroke and DBP for all men is given in the figure below. A linear model is also fitted.

**Figure 5.7: Relationship between DBP in phase II and stroke**
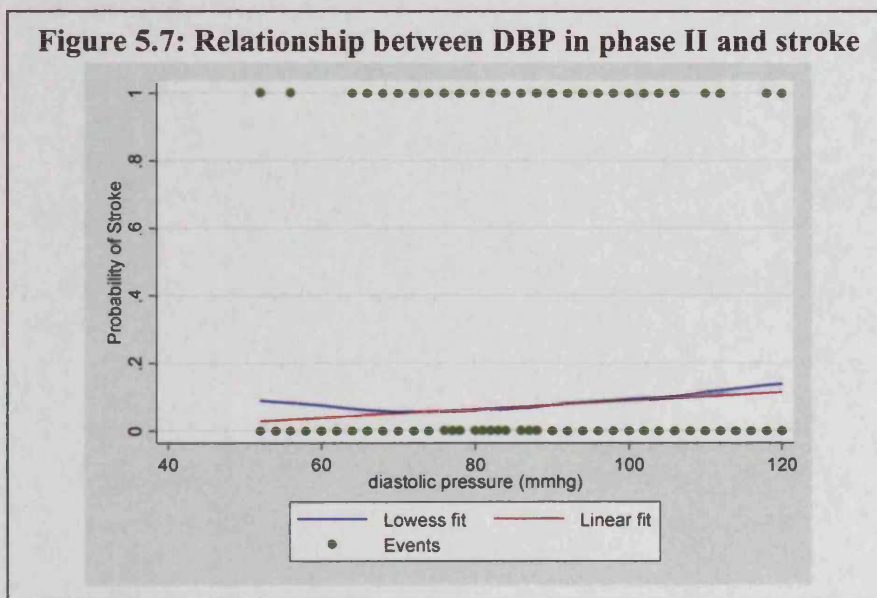


The relationship appears to be linear. The gain of a fractional polynomial model over the linear fit is 0.48 with 4 degrees of freedom, which is statistically not significant. Therefore, a linear model can be concluded as being adequate.

The log odds regression coefficient table for baseline DBP and stroke for all men is given in Table 5.8.

**Table 5.8 – Logistic regression of baseline DBP and stroke for all men**

| Stroke | Coef | Std Err | P>\|z\| | 95 % CI | |
|---|---|---|---|---|---|
| DBP | 0.018 | 0.008 | 0.024 | 0.002 | 0.034 |
| Const | -4.138 | 0.687 | 0.000 | -5.550 | -2.857 |

The results are statistically significant.

Table 5.8 can be represented in an equation as follows

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = -4.138 + 0.018 \times DBP \dots\dots\dots\dots \quad (5.7)$$

where $p$ is the probability of having stroke. For an increase of 10 mmHg of DBP, the odds of stroke increases by 20%.

### 5.2.2.2.2 Cox's Proportional Hazard Model and Weibull regression model

Cox' proportional hazard model also showed similar results. Table 5.9 gives the hazard ratios for an increase of 1mmHg in DBP.

**Table 5.9 – CPHM of baseline DBP and stroke**

| MI | Haz ratio | Std Err | P>|z| | 95 % CI | |
|----|-----------|---------|-------|---------|---|
| DBP | 1.018 | 0.008 | 0.017 | 1.003 | 1.033 |

The Table 5.7 can be rewritten in the form of an equation as follows

$$\log\left(\frac{h\{(t),(DBP)\}}{h_0(t)}\right) = 0.018 DBP \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (5.8)$$

A fractional polynomial model for the Cox's proportional hazard model showed no significant gain (deviance gain 0.44) suggesting that the linear model is adequate.

## 5.2.3 Pulse Pressure

Pulse pressure is the difference between SBP and DBP. As in the case of previous risk factors, only three measurements (from phases 2, 3 and 4) will be considered in the analysis. The following Figure 5.8 gives the distribution of the three measurements of pulse pressure.

**Figure 5.8 – Histograms of PP in 2ⁿᵈ, 3ʳᵈ, and 4ᵗʰ phases**
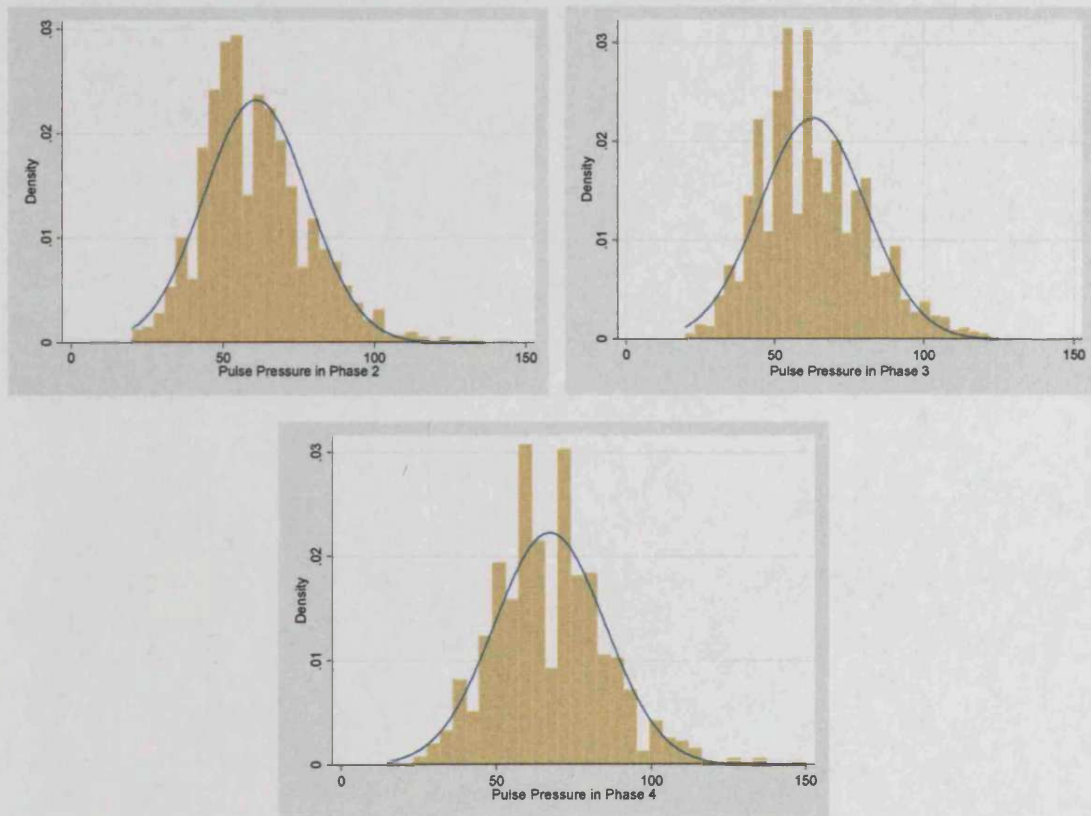


The summary statistics of PP are given in Table 5.10.

**Table 5.10 – Summary statistics of PP in 2ⁿᵈ, 3ʳᵈ and 4ᵗʰ phases**

| Variables | Obs | Mean | Std. Dev | Min | Max |
|-----------|------|-------|----------|-----|-----|
| PP in Ph 2 | 1836 | 61.72 | 17.63 | 20 | 136 |
| PP in Ph 3 | 1598 | 63.27 | 17.87 | 20 | 130 |
| PP in Ph 4 | 1397 | 67.59 | 17.69 | 15 | 150 |

The mean of PP seems to have increased over the period. When men with all three measurements were considered, the mean increased over the period of the study (phase 2= 61.1 (SD=17.4), phase 3= 62.8 (SD=17.1), phase 4=67.6 (SD=17.7)). Mean SBP stayed fairly constant at the end of phase 4 but the mean of DBP fell. So the increase in the average pulse pressure is not surprising.

### 5.2.3.1 PP and MI

A Lowess smoothed plot for the relationship between MI and PP, along with logistic and fractional polynomial models is given in Figure 5.9 below.



Figure 5.9: Relationship between PP in phase two and MI
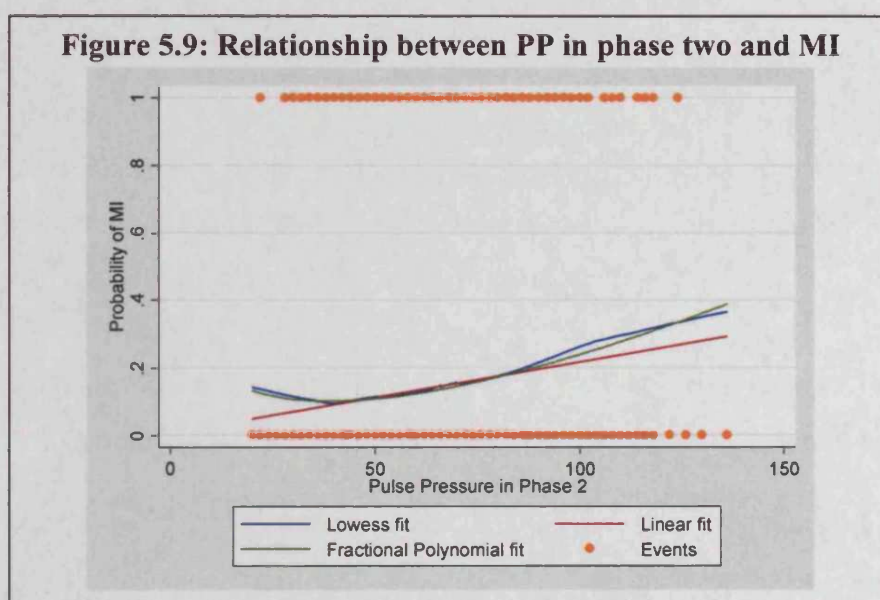
The gain in deviance of the fractional polynomial model over the linear model is 1.19; that is statistically not significant and again suggests the linear model is adequate. Furthermore, none of the studies in the brief literature review showed a non-linear relationship between pulse pressure and MI therefore, a linear model was used.

### 5.2.3.1.1 Logistic regression

The regression coefficient table for baseline PP and MI for men without prevalent IHD is given in Table 5.11.

**Table 5.11 – Linear logistic regression of baseline PP and MI**

| MI | Coef | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| PP | 0.016 | 0.004 | 0.000 | 0.009 | 0.023 |
| Const | -2.885 | 0.257 | 0.000 | -3.431 | -2.433 |

Substituting these values in the equation, the regression equation is as follows.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.016PP - 2.885 \quad\text{............}(5.9)$$

where $p$ is the probability of having MI.

For an increase in PP by 10 mmHg, the odds of MI increase by 18%.

### 5.2.3.1.2 Cox' proportional hazard model

CPHM showed similar results when compared to linear logistic regression in the above section. The gain obtained by the fractional polynomial model over the linear model was 1.36 which is not significant. The hazard ratio for an increase of 1 mmHg in PP is given in the Table 5.12.

**Table 5.12 – CPHM of baseline PP and MI for men without prevalent IHD**

| MI | Haz.r | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| PP | 1.016 | 0.003 | 0.000 | 1.011 | 1.024 |

This can be rewritten in the form of an equation as follows

$$\log\left(\frac{h\{(t),(PP)\}}{h_0(t)}\right) = 0.016PP \quad\text{.......................................}(5.9)$$

For an increase in PP by 10 mmHg, the odds of MI increase by 18%.

### 5.2.3.2 PP and Stroke

A Lowess smoothed plot for the relationship between stroke and PP for all men along with the different modelling strategies fitted is shown in the Figure 5.10.



Figure 5.10: Predicting Stroke using baseline PP
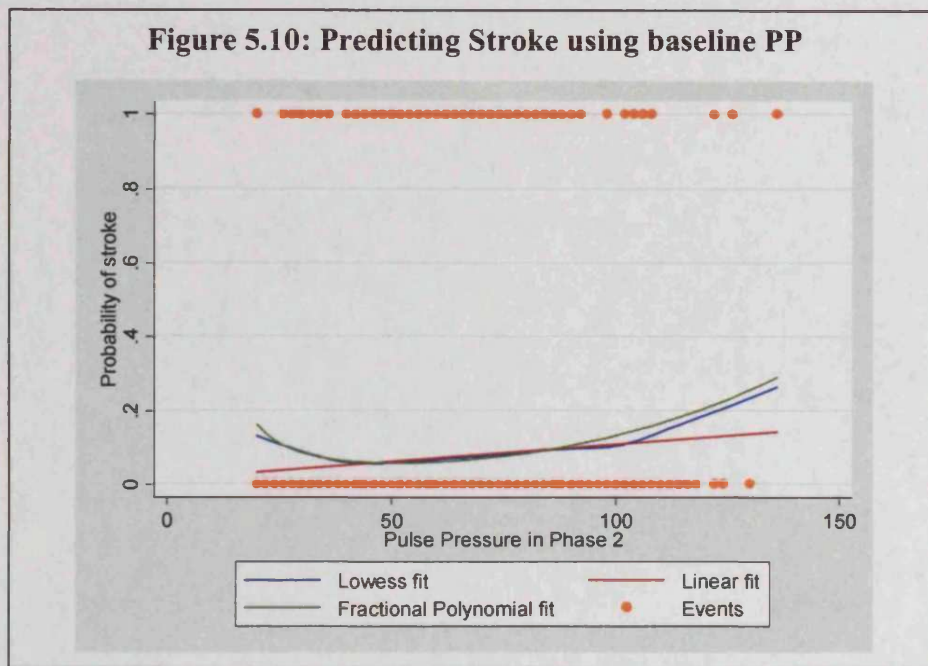
The relationship appears non-linear. The model can be improved if fractional polynomial regression is used. The gain in the deviance of a fractional polynomial model over the linear logistic model is 5.17 with 4 degrees of freedom. It shows that the fractional polynomial relationship between pulse pressure and stroke is statistically not significant over the linear logistic model. There is no such evidence from the literature review that suggests a non-linear relationship between stroke and PP therefore, a linear model was considered appropriate.

### 5.2.3.2.1 Linear logistic regression

The regression coefficient table for baseline PP and stroke for all men is given in Table 5.13.

Table 5.13 – Linear logistic regression of baseline PP and stroke

| Stroke | Coef | Std Err | P>|z| | 95 % CI | |
|--------|------|---------|-------|---------|---|
| PP | 0.013 | 0.005 | 0.006 | 0.004 | 0.023 |
| Const | -3.431 | 0.334 | 0.000 | -4.086 | -2.775 |

where $p$ is the probability of having an MI.

Substituting these values in the equation, the linear logistic regression equation is as follows.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.013PP - 3.43 \dots\dots\dots\dots(5.16)$$

where $p$ is the probability of having a stroke.

For an increase of 10 mmHg in PP, the odds of stroke is increased by 13%

## 5.2.3.2.2 Cox's Proportional hazard model

Cox' proportional hazard model also showed similar results with a hazard ratio of 1.17 [95% CI: 1.09 – 1.28] for every 10 mmHg increase in PP. The hazard ratios for an increase of 1mmHg of PP are shown in Table 5.14. If fractional polynomial regression were to be used in CPHM, the gain in the deviance as 3.03 which is statistically not significant.

**Table 5.14 – CPHM of baseline PP and stroke**

| MI | Hazard ratio | Std Err | P>|z| | 95 % CI | |
|----|-------------|---------|-------|---------|---|
| PP | 1.015 | 0.005 | 0.001 | 1.006 | 1.025 |

The hazard ratio in the above table can be written in the form of an equation as follows

$$\log\left(\frac{h\{(t),(PP)\}}{h_0(t)}\right) = 0.015PP \dots\dots\dots\dots\dots\dots\dots\dots(5.10)$$

## 5.3 Regression Dilution Bias of blood pressure

SBP and DBP were measured in both subsequent phases of the Caerphilly study. There were 1352 men with measurements in phases 2, 3 and 4. The summary statistics of the systolic, diastolic and pulse pressures for all the men is given in the Table 5.15.

**Table 5.15: Summary statistics for regression dilution bias for MI and SBP, DBP and PP**

| Variable | observations | Mean | Standard deviation |
|----------|--------------|------|---------------------|
| SBP in P2 | 1352 | 146.07 | 22.37 |
| SBP in P3 | 1352 | 144.89 | 21.95 |
| SBP in P4 | 1352 | 146.21 | 21.86 |
| SBP2-SBP3 | 1352 | 1.18 | 20.24 |
| SBP2-SBP4 | 1352 | -0.13 | 22.66 |
| | | | |
| DBP in P2 | 1352 | 84.98 | 11.42 |
| DBP in P3 | 1352 | 82.10 | 11.44 |
| DBP in P4 | 1352 | 78.58 | 11.36 |
| DBP2-DBP3 | 1352 | 2.88 | 11.89 |
| DBP2-DBP4 | 1352 | 6.40 | 12.28 |
| | | | |
| PP in P2 | 1352 | 61.09 | 17.37 |
| PP in P3 | 1352 | 62.79 | 17.71 |
| PP in P4 | 1352 | 67.63 | 17.74 |
| PP2-PP3 | 1352 | -1.70 | 16.69 |
| PP2-PP4 | 1352 | -6.53 | 18.31 |

### 5.3.1 SBP

The regression dilution bias due to variation in SBP is estimated using the information given in the Table 5.16. After adjusting the regression coefficient by the coefficient of regression dilution using the method explained in section 3.7.1, the baseline and adjusted $\beta$'s are given in the following table.

**Table 5.16: $\beta$s adjusted with coefficient of measurement error**

| | $\beta_0$ | $\beta_1$ |
|--|-----------|-----------|
| Baseline $\beta$s (phase 2) | -4.409 | 0.0133 |
| $\beta$s adjusted for Phase 3 | -5.751 | 0.0225 |
| $\beta$s adjusted for Phase 4 | -6.449 | 0.0272 |

In the above table the first row represents the regression coefficients at the baseline i.e., the regression coefficients to the logistic model fitted to the SBP and MI. This model was fitted using the 1352 subjects who had all the three measurements. The second row represents the coefficients adjusted to the regression dilution bias due to the variation in SBP using the $3^{rd}$ phase SBP values whereas the $3^{rd}$ row represents the coefficients adjusted using the $4^{th}$ phase SBP values. The Figure 5.11 shows the probability of MI when fitted with the baseline $\beta$s and the adjusted $\beta$s in the subsequent phases.



**Figure 5.11: Predicting MI with SBP for all men using baseline and RDB adjusted models**
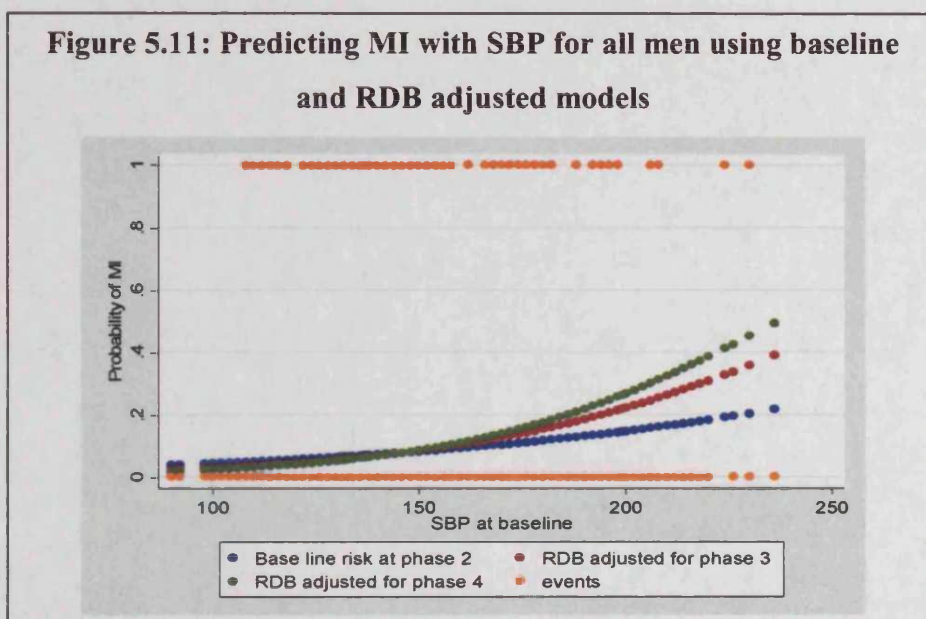
In the Figure 5.11, when the baseline coefficient was adjusted using the coefficient of measurement error, the risk of MI with SBP appears to increase over the subsequent phases. There appears to be a twofold increase in the log odds ratio when adjusted with the $4^{th}$ phase SBP values compared with the baseline risk.

## 5.3.2 DBP

The regression dilution bias due to variation in DBP is estimated using the information given in the Table 5.17. After adjusting the regression coefficient by the coefficient of regression dilution using the method explained in section 3.7.1, the baseline and adjusted $\beta's$ are given in the following table.

**Table 5.17: $\beta$s adjusted with measurement**

|  | $\beta_0$ | $\beta_1$ |
|---|---|---|
| Baseline $\beta$s (phase 2) | -5.016 | 0.0298 |
| $\beta$s adjusted for Phase 3 | -8.083 | 0.0651 |
| $\beta$s adjusted for Phase 4 | -8.487 | 0.0707 |

In the above table the first row represents the regression coefficients at the baseline i.e., the regression coefficients to the logistic model fitted to the DBP and MI. This model was fitted using the 1352 subjects who had all the three measurements. The second row represents the coefficients adjusted to the regression dilution bias due to the variation in DBP using the 3$^{rd}$ phase DBP values whereas the 3$^{rd}$ row represents the coefficients adjusted using the 4$^{th}$ phase DBP values. The Figure 5.12 shows the probability of MI when fitted with the baseline $\beta$s and the adjusted $\beta$s in the subsequent phases.



**Figure 5.12: Predicting MI with DBP for all men using baseline and RDB adjusted models**
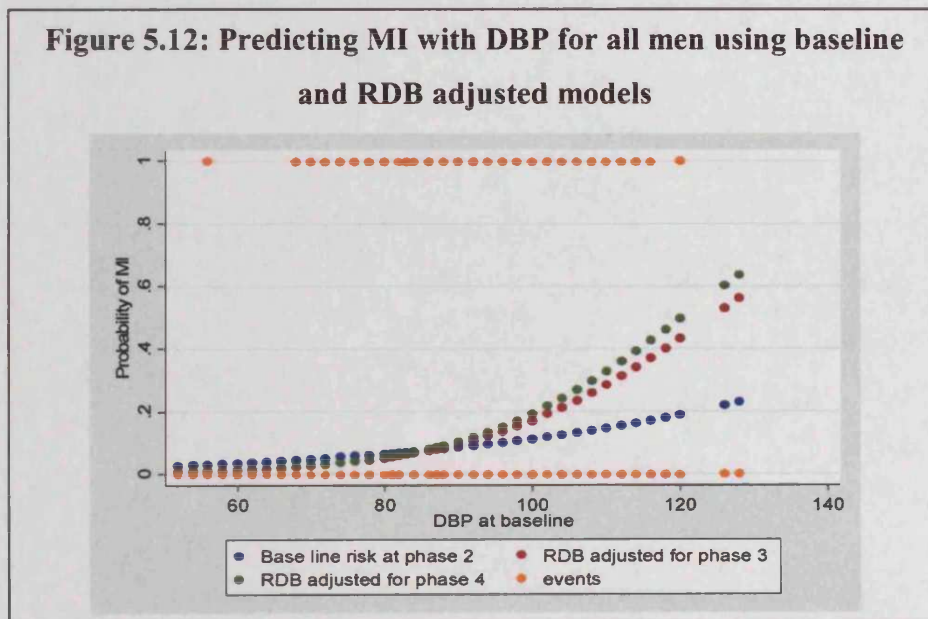
In the Figure 5.12, when the baseline coefficient was adjusted using the coefficient of measurement error, the risk of MI due to DBP appears to increase over the subsequent phases. There appears to be a twofold increase in the log odds ratio when adjusted with the 4$^{th}$ phase DBP values compared with the baseline risk.

## 5.3.2 PP

The regression dilution bias due to variation in PP is estimated using the information given in the Table 5.18. After adjusting the regression coefficient by the coefficient of regression dilution using the method explained in section 3.7.1, the baseline and adjusted $\beta$'s are given in the following table.

**Table 5.18: $\beta$s adjusted with measurement**

|  | $\beta_0$ | $\beta_1$ |
|---|---|---|
| Baseline $\beta$s (phase 2) | -3.062 | 0.0101 |
| $\beta$s adjusted for Phase 3 | -3.589 | 0.0187 |
| $\beta$s adjusted for Phase 4 | -3.829 | 0.0226 |

In the above table the first row represents the regression coefficients at the baseline i.e., the regression coefficients to the logistic model fitted to the PP and MI. The second row represents the coefficients adjusted to the regression dilution bias due to the variation in PP using the $3^{rd}$ phase PP values whereas the $3^{rd}$ row represents the coefficients adjusted using the $4^{th}$ phase PP values. The Figure 5.13 shows the probability of MI when fitted with the baseline $\beta$s and the adjusted $\beta$s in the subsequent phases.



Figure 5.13: Predicting MI with PP for all men using baseline and RDB adjusted models
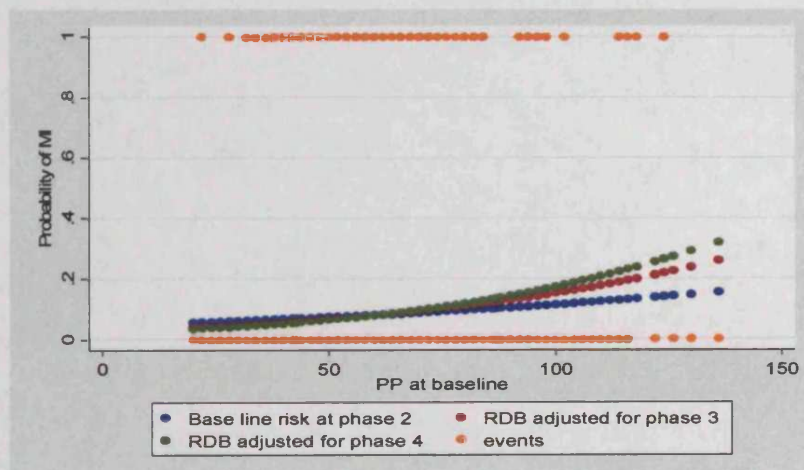
In the above figure, when the baseline coefficient was adjusted using the coefficient of measurement error, the risk of MI due to PP appears to increase over the subsequent phases. There appears to be almost twofold increase in the log odds ratio when adjusted with the 4$^{th}$ phase PP values compared with the baseline risk.

## 5.4 Discussion

### 5.4.1 SBP

The odds ratio of having an MI for every 10mmHg increase in SBP is about 1.19 in Caerphilly. The Framingham study [113] from the literature review showed a 41% greater risk of VAD for men with an increase of 1 standard deviation of SBP. In the CaPS, the odds of MI increases by 50% with an increase of 1 standard deviation of SBP. The range of age in the Caerphilly analysis is 45-59 whereas the range of age in the Framingham analysis is 36-64 and the results shown by them were adjusted for age. Furthermore, the outcome variable used in the Framingham analysis was VAD whereas the outcome of this analysis is MI. These could be the possible reasons for the effect of SBP on risk to be higher in Caerphilly men. A Swedish study showed that the odds of heart failure for men with 145<SBP<176 and SBP>175 compared to the SBP<146 are 1.31 (95% CI: 1.13 – 1.53) and 1.92(95% CI: 1.57 – 2.35) respectively [114]. For the same groups when compared in the CaPS, the odds were 1.27 (95% CI: 0.94 – 1.73) and 3.15 (95% CI: 2.15 – 4.6). It appears that the odds of MI in CaPS are higher then the odds of heart failure in the Swedish study for the high SBP group. However, the outcome variables in the CaPS and the Swedish study are slightly different; this could possibility explain some of the difference in the results. From the analysis of stroke and SBP in the CaPS, the odds of stroke increases by 14% for every 10mmHg increase in SBP. The Framingham study from the literature review has used the outcome variable as CVD, which combines the MI and stroke. Therefore, a direct comparison cannot be made. However, the odds of stroke in the CaPS for an increase of 1 SD are 35%. Other studies discussed in the literature review have also used as their outcome variable either CHD or VAD, making it difficult to compare the results with those of CaPS.

The odds of MI appear to show a twofold increased risk when adjusted with the 3$^{rd}$ and 4$^{th}$ phase repeated measures. The Oxford study based on the Framingham data

and Whitehall data showed that the baseline measurements underestimate the effect of SBP on risk by one third in the first decade [119]. Whereas, in the CaPS analysis, when the baseline is adjusted for the $3^{rd}$ and $4^{th}$ phase repeated measurements, which is technically a decade, there appears to be slightly more than twofold increase in the odds of MI. There were several repeated measurements carried out in the Framingham study to measure the regression dilution due to variation in SBP more accurately then the Caerphilly analysis. The CaPS analysis is a bit crude as it is based on only three measurements.

## 5.4.2 DBP

A Study of 11,150 male physicians aged 40-84 years who were free of heart disease at recruitment showed an increased risk of VAD by 1.46 [95% CI: 1.26 – 1.69] for every 10mmHg increase in DBP [3]. The increased odds of MI with DBP in Caerphilly men is 1.35 [95% CI: 1.19 – 1.51] per 10 mmHg increase in DBP. Benetos A, et al [116] showed that the risk of CVD for men with DBP<90 compared with men having DBP between 90-99 and DBP>99 an RR of 1.31 (95% CI: 0.94 – 1.83) and 1.60 (95% CI: 1.16 – 2.20). The CaPS results showed an increased odds of 1.67 (95% CI: 1.21 – 2.30) and 2.43 (95% CI: 1.66 – 2.47) for MI when same groups of DBP are considered for the analysis. However, the results in Benetos A, et al [116] were adjusted for various other cardiovascular risk factors. In addition, the outcome of the study was CHD mortality, rather than MI and stroke events as in CaPs. Furthermore, the study sample had a mean age of 52 with a standard deviation of 11, whereas, in the CaPS, the mean age is 56.8 with an SD of 4.5. Finally, those results were adjusted for hypertensive treatment. This could probably explain some of the differences in the results. The odds of stroke with DBP appear to be 20% with an increase in DBP by 10mmHg. The studies discussed in the literature review used CVD as an outcome which includes stroke. Therefore, a direct comparison cannot be made. Regression dilution due to the variation in DBP in the CaPS analysis showed that a single measurement of DBP underestimates the odds of MI by more than 100% when adjusted for $3^{rd}$ and $4^{th}$ phase measurements. However, in the analysis of Framingham and Whitehall study data, Clarke et al suggest an underestimation by 50% [119]. As explained in the case of SBP, the regression dilution due to the variation in DBP effect is about 100% in the case of Caerphilly study whereas the Clake et al study

estimated it to be around 50%. Their results are more trustworthy due to larger numbers and more repeated measurements.

### 5.4.3 PP

Pulse pressure in the CaPS showed an increased odds of MI and stroke by 18% and 13% respectively for an increase of 10 mmHg. Sesso et al. showed a hazard ratio of 1.23 (95% CI: 1.09 – 1.40) for a 10 mmHg increase in PP [122]. However, their outcome was CVD rather then MI or stroke and this may explain the smaller effect found in the Caerphilly analysis. The odds ratio of a 10mmHg increase in PP in the Caerphilly analysis is 1.18 [95% CI: 1.09 – 1.26] whereas the odds reported by an American study [120] were 32%. Several factors, including difference in age groups and lifestyle may explain the reason for higher odds in the American study. The adjustment of measurement error and random fluctuations show that one measurement of pulse pressure underestimates the risk of MI. when the repeated measurements were used; there was a almost a twofold increase in odds. However, these results should not be taken seriously as this twofold increase is due to the modelling measurement error and random fluctuations in PP and also, there very few repeated measurements available in the Caerphilly data.

Although PP is thought to reflect arterial elasticity, it does not appear to be a better predictor of VAD in Caerphilly than either systolic or diastolic pressure. While high PP almost certainly corresponds to high SBP and hence an increased risk, a low PP could occur with normal SBP and high DBP, which might also be associated with a higher risk. So the interpretation of association between PP and MI or stroke is not entirely clear.

## 5.5 Conclusions for Blood pressure

A brief literature review has shown that SBP, DBP and PP are all important risk factors for cardiovascular disease. Most of the literature review showed that the risk of MI and stroke is significantly increased for increased levels of blood pressure. Some studies have shown that the pulse pressure is more strongly associated with the risk of CVD than the SBP and DBP. But as discussed in the discussion of PP, but there is not a strong empirical case for preferring PP to either systolic or diastolic pressure as risk factors for MI or stroke.

According to the Italian meta analysis [118], for every 10mmHg increase in DBP, the odds of MI are approximately equal to odds of MI for an increase of 20mmHg SBP. In Caerphilly analysis, for every 20mmHg increase in SBP, the odds of MI increases by 34%, whereas the odds of MI for every 10mmHg increase in DBP is 27%. If the standard deviation increase is taken into consideration, for every 1 SD increase in SBP, the increased odds of MI increase by 40% and for every 1 SD increase in DBP, they increase by 34%. This shows that SBP is strongly associated with the risk of MI then DBP as in the study of Benetos et al [116].

# 5.6 Literature review for Cholesterol levels

Cholesterol is a fatty acid which is essential to the healthy functioning of cells as it strengthens and regulates the function of cell membranes. Cholesterol is transported around the body packaged in lipoproteins in the blood. When there are high levels of cholesterol in the blood, it is not only used to repair and maintain cells, but it is also deposited in comparatively large amounts in the arterial wall causing "hardening of the arteries", also known as atherosclerosis. The deposited cholesterol impairs the arterial wall's elasticity. The cholesterol deposit, also called atheroma, can begin to occupy the lumen of the artery and so narrow space available for blood flow. Although the atheroma is an anomalous deposit, it is still living tissue, which requires a blood supply. As the atheroma grows, its blood supply can become inadequate leading to necrosis. The dead area of atheroma is vulnerable to the turbulence and pressure within the artery and may be dislodged exposing damaged tissue. This in turn causes a blood clot to form in the artery as a haemostatic response occurs to repair the damaged tissue.

The narrowing of the arterial lumen can cause heart disease by gradually occluding a coronary artery and so depriving the tissue it supplies of blood. This leads to the death of heart muscle, which is a heart attack. A heart attack can also occur when the atheroma in an already narrowed artery ruptures, causing a haemostatic response. The blood clot formed by the haemostatic response occludes the already narrowed artery.

Not all types of cholesterol pose the same vascular risk. High density lipoprotein (HDL) cholesterol is 'good cholesterol' as it protects against heart disease by helping remove excess cholesterol deposited in the arteries. High levels seem to be associated with low incidence of coronary heart disease whereas low density lipoprotein (LDL) cholesterol is considered "bad cholesterol" because LDL cholesterol makes up most of the cholesterol which is deposited in arteries.

Most epidemiological studies have used total cholesterol (TC) as an index of vascular risk. One American study with 4066 old men and women with a follow-up of 4 years has shown that men with elevated total cholesterol levels ($\geq$ 6.20 mmol/L) increased the risk for death from coronary heart disease by a RR of 1.57[95% CI:1.06 – 2.04] compared to the men with the cholesterol levels between 4.16 – 5.19 mmol/L [125].

The Seven Countries Study showed a RR of 1.3 for every 1 mmol/L increase in TC for having a fatal MI in older men without pre-existing CHD [126]. Read et al. showed an increased risk of 1.16[95% CI: 1.06-1.27] for CHD incidence for an increase in total cholesterol by 0.5 mmol/L [127]. The Zutphen Elderly Study on Dutch men has showed RR = 1.40 (95% (CI) 1.07-1.83) for every 1 mmol/L increase in TC for fatal CHD [128]. The Finland, Italy and the Netherlands Elderly (FINE) study in year 2000, showed a combined relative risk of heart disease mortality for the total population of the FINE Study was 1.17 (95% confidence interval = 1.06-1.29) for each 1.00 mmol/L increase in total cholesterol[129]. A meta analysis which included all the above studies and several others showed that the overall RR for CHD incidence associated with 1.0 mmol/L increase in TC is 1.28 (CI 1.17-1.39) [130].

Most of the studies which investigated associations between VAD and TC also estimated associations with HDL. In the Finnish cohort of the seven countries study, HDL showed a protective effect with an odds ratio 0.2 (95% CI 0.1-0.8) for fatal MI for an increase of 1 mmol/L of HDL [131]. In the FINE study, HDL was protective without any adjustment for Finnish subjects, whereas the Italian and Dutch subjects showed HDL was protective for fatal MI after adjustment for SBP, BMI, smoking and alcohol consumption [132]. HDL cholesterol was not associated with mortality from coronary heart disease in the Zutphen Elderly study with a relative risk of 0.80 (95% CI 0.60-1.08) for the incidence of heart disease, corresponding to a 0.26 mmol/L increase in HDL [133].

Studies evaluating the relationship between cholesterol levels and the risk of stroke have reported different results. The Framingham study found no connection between the levels of cholesterol and incidence of stroke [7]. Two different studies have reported an increased risk of stroke with high levels of TC [9, 134]. There are some other studies which showed no association between stroke and increased levels of cholesterol [135-137]. The combined analysis of cohort studies showed no significant association between the increased level of serum cholesterol and stroke rate, except for patients younger than 45 years [138].

## 5.7 Analysis of Total Cholesterol

Total Cholesterol (TC) was measured in all four phases of the Caerphilly cohort. The distributions of TC in $2^{nd}$, $3^{rd}$ and $4^{th}$ phases are given below in Figure 5.14.

**Figure 5.14 – Histograms of TC in $2^{nd}$, $3^{rd}$, and $4^{th}$ phases**
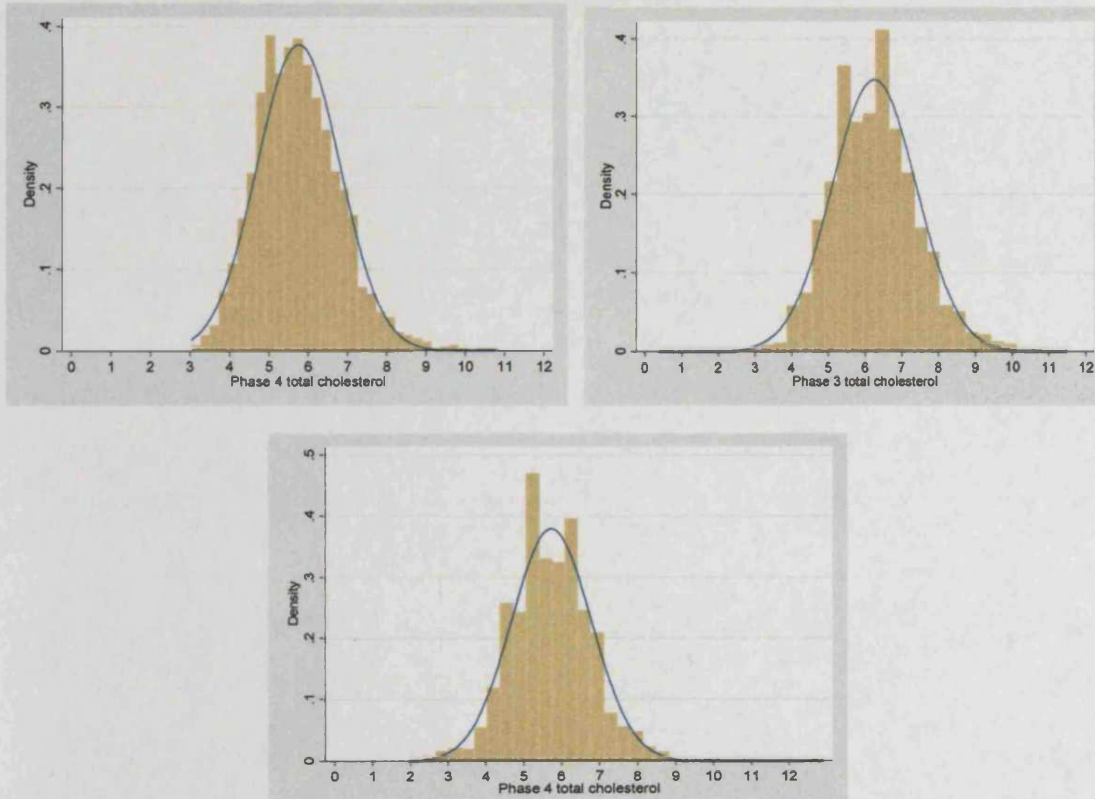


For the comparison purposes, all the three measurements of total cholesterol are plotted on the same X axis scale. However, it appears that there are some outliers in the $3^{rd}$ and $4^{th}$ phase total cholesterol data and hence their distributions slightly vary compared with the $2^{nd}$ phase TC measurements. After excluding missing values in the TC data, the total number of observations considered used in all 3 phases are given below Table 5.19 along with the standard deviations.

**Table 5.19 – Summary statistics of TC in 2$^{nd}$, 3$^{rd}$ and 4$^{th}$ phases**

| Variables | Obs | Mean | Std. Dev | Min | Max |
|-----------|-----|------|----------|-----|-----|
| TC in Ph 2 | 1606 | 5.70 | 1.04 | 3.24 | 10.35 |
| TC in Ph 3 | 1564 | 6.21 | 1.12 | 0.40 | 11.50 |
| TC in Ph 4 | 1406 | 5.69 | 1.05 | 2.00 | 12.90 |

There were 1190 men with cholesterol measured on all three occasions. The means and standard deviations of these men were very similar to those given in the above table (values listed in table 5.24). To find the relationship between TC and MI or stroke, as in the previous chapter, Lowess smoothing was used to visualise the relationships and then decide the modelling strategy. The relationship between MI and cholesterol levels will be explored separately for MI and stroke.

## 5.7.1 Total cholesterol and MI

### 5.7.1.1 Linear logistic regression

The Lowess smoothing plot with a bandwidth of 0.8 is shown in Figure 5.15 along with the linear and fractional polynomial models fitted to the relationship.



Figure 5.15: Lowess smoothed plot of MI and TC
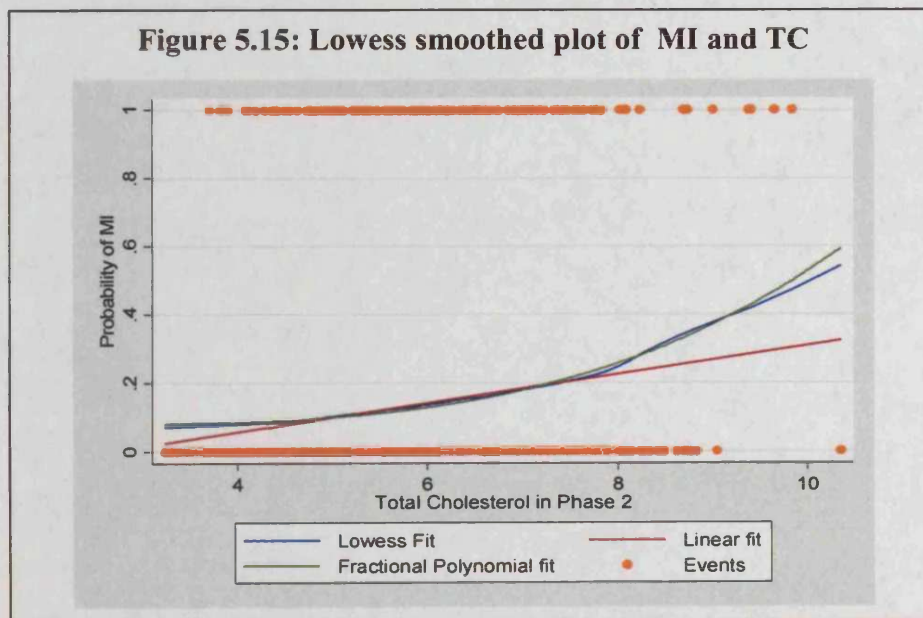
The fractional polynomial model appears to be a better model visually but linear logistic regression is statistically a more appropriate model since the fractional

104

polynomial model's gain over the linear model is 0.98, which is statistically not significant. The regression coefficients are given in Table 5.20.

**Table 5.20 – Logistic regression of baseline TC and MI**

| MI | Coef | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| TC | 0.356 | 0.071 | 0.000 | 0.217 | 0.495 |
| Const | -3.989 | 0.432 | 0.000 | -4.836 | -3.143 |

Substituting these values in the equation, the regression equation is as follows.

$$\text{Logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.356TC - 3.989 \quad\text{.................(5.19)}$$

where $p$ is the probability of MI.

The odds ratio for every 1mmol/L increase in TC is 1.43 [95% CI: 1.24 1.64] for having an MI shows that TC is a very significant risk factor for MI.

## 5.7.1.2 Cox's Proportional hazard model

CPHM also showed the hazard ratio of 1.39 [95% CI: 1.22 – 1.58] for every 1 mmol/L increase in TC for MI. The model can be rewritten in the form of following equation.

$$\log\left(\frac{h\{(t),(TC)\}}{h_0(t)}\right) = 0.33TC \quad\text{.................................. (5.20)}$$

The gain in the fractional polynomial models was 1.338 with 4 degrees of freedom which again shows no statistical difference between the linear and fractional polynomial models.

## 5.7.2 Stroke and TC

To visualise the relationship between, TC and stroke for all men, the Lowess smoothing plot with a bandwidth of 0.8 is shown in Figure 5.16 along with the linear model fitted to it.



**Figure 5.16: Lowess smoothed plot of stroke and TC for all men**
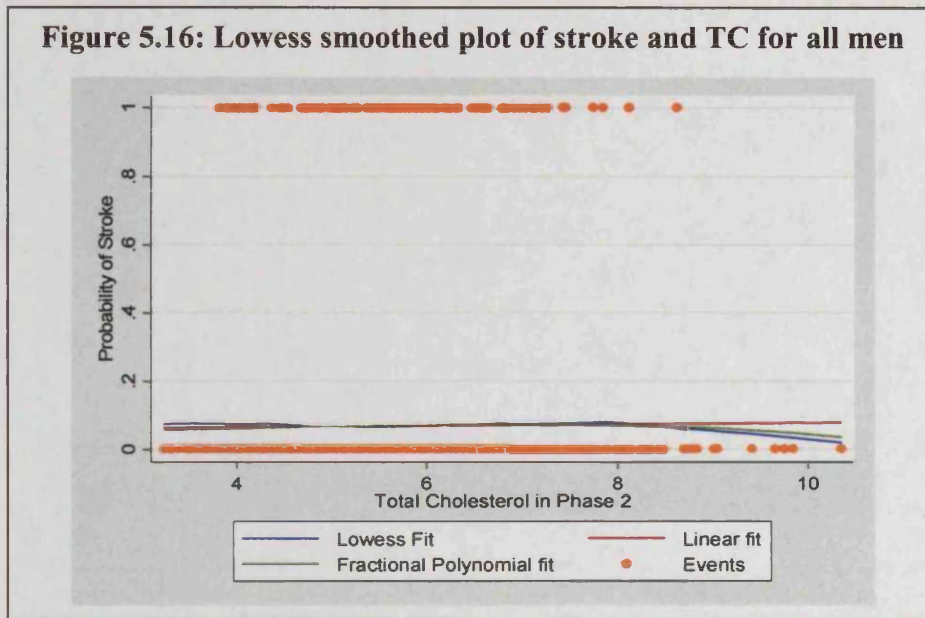
The slope of the relationship between stroke and TC for all men appears to be very close to zero, i.e., the risk appears to be constant for all men irrespective of the increase in TC for all men. This result follows some studies from literature review showing that an increase in TC does not predict stroke. The regression coefficients are given in Table 5.21.

**Table 5.21 – Logistic regression of baseline TC and stroke**

| Stroke | Coef | Std Err | P>|z| | 95 % CI | |
|--------|------|---------|-------|---------|---|
| TC | 0.031 | 0.094 | 0.740 | -0.154 | 0.217 |
| Const | -2.760 | 0.550 | 0.000 | -3.839 | -1.681 |

Substituting these values in the equation, the regression equation is as follows.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.0314TC - 2.7602 \ldots\ldots\ldots(5.22)$$

where $p$ is the probability of stroke.

These results show that there is no significant association between total cholesterol and the risk of stroke. If fractional polynomial regression were to be used to model this relationship, the gain is 0.36 which is statistically not significant. CPHM also showed similar results with no significance of total cholesterol as a risk factor for stroke.

## 5.8 Analysis of High Density Lipoprotein Cholesterol (HDLC)

The literature review suggested that an increase in HDLC reduces the risk of MI whereas the risk is not significant for stroke. HDLC was measured only in phase 2 of the cohort. The distibution of HDLC is given in Figure 5.17.



**Figure 5.17 – Histogram of HDLC in 2$^{nd}$ phase**
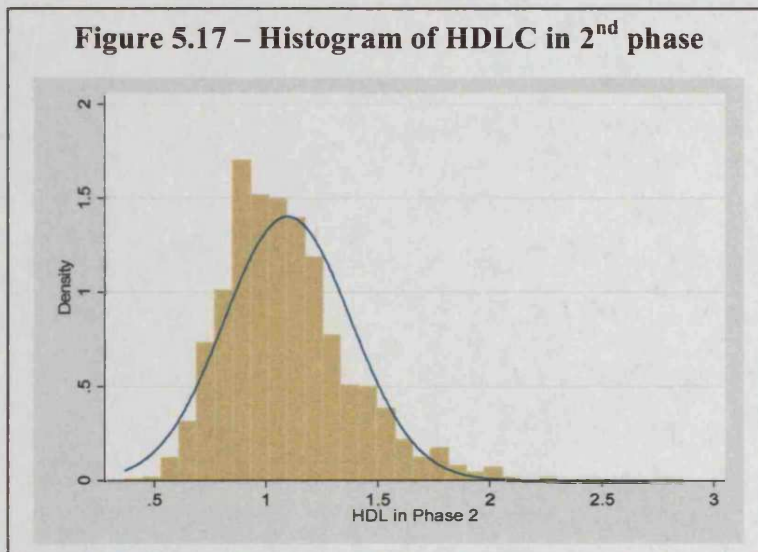
From the above histogram, the distribution of the HDLC appears to be positively skew. After excluding the missing values in the HDLC data, the total number of observations considered is 1426 with a mean 1.097 and standard deviation of 0.284.

To find the relationship between HDLC and MI or stroke, as in the case of TC, Lowess smoothing was used to visualise the relationships and then decide the modelling strategy.

There were two observations with HDLC $\leq 0.5$ and ten observations with HDLC$\geq$2 as these were causing the relationship to appear non-linear, they were excluded from the analysis.

### 5.8.1 HDL and MI

To visualise the relationship between, HDLC and MI, Lowess smoothing plot with a bandwidth of 0.8 is shown in Figure 5.18. The same figure also shows the linear model fitted to the relationship.



**Figure 5.18: Lowess smoothed plot of MI and HDLC**
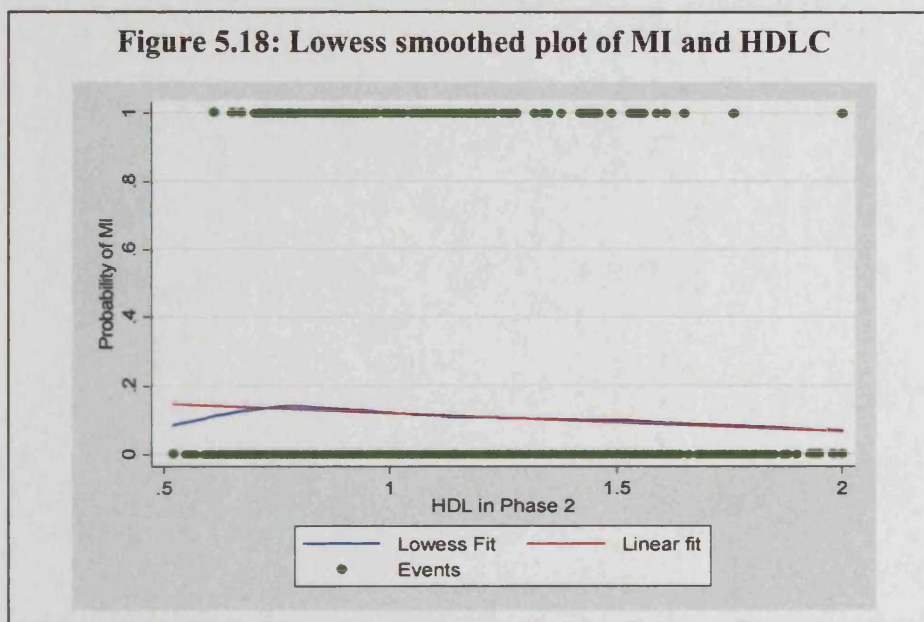
The relationship between MI and HDLC appears to be linear except for low values of HDLC. If the fractional polynomial regression is used to model this relationship, the gain obtained is 1.31 which is statistically not significant to show that factional polynomial model is better than the linear model. The logistic linear regression coefficients are given in Table 5.22.

**Table 5.22– Logistic regression of baseline HDLC and MI**

| MI | Coef | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| HDLC | -0.538 | 0.328 | 0.101 | -1.180 | 0.105 |
| Const | -1.432 | 0.358 | 0.000 | -2.134 | -0.729 |

Substituting these values in the equation, the regression equation is as follows.

$$\text{Logit}(p) = \ln\left(\frac{p}{1-p}\right) = -0.538HDLC - 1.432 \dots\dots\dots\dots(5.23)$$

where $p$ is the probability of MI.

The odds ratio for an increase in 1 unit of HDLC is 0.539 [95%CI : 0.280 – 1.038]. The association is not statistically significant but the results are consistent with an odds ratio well below 1. It is worth noting, however, that a difference of 1 is very large, as the range of HDLC values is from 0.5 to 2. The Cox's proportional hazards model also showed similar results which were not significant (Hz ratio = 0.57 p = 0.07, SE = 0.18, 95% CI = 0.31, 1.04).

### 5.8.2 Stroke and HDLC

To visualise the relationship between, HDLC and, Lowess smoothing plot with a bandwidth of 0.8 is shown in Figure 5.19.



Figure 5.19: Lowess smoothed plot of stroke and HDLC
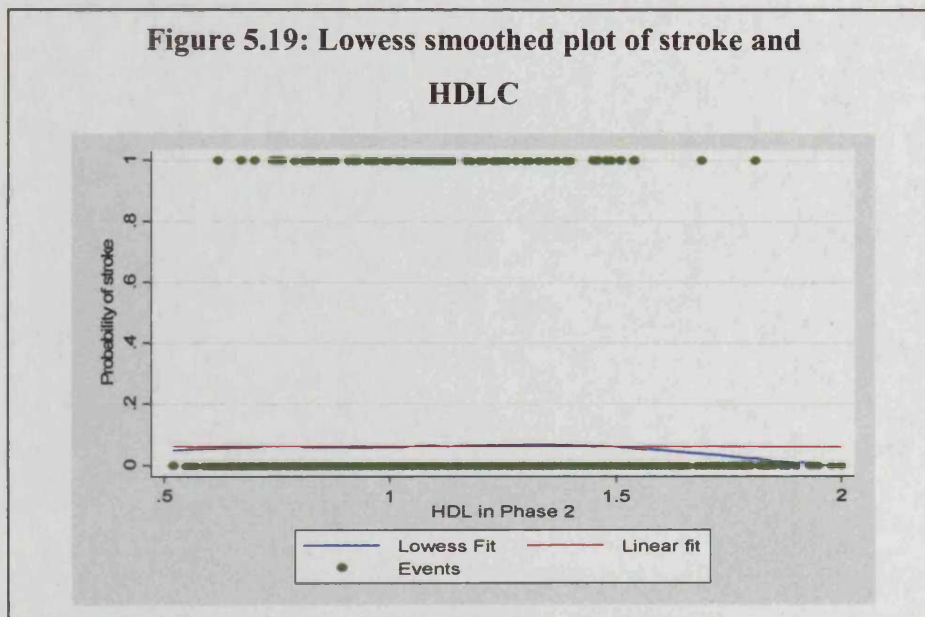
The linear model fitted to this relationship is also shown in the above figure. The relationship between stroke and TC appears no association between the risk of stroke and increase in HDLC. The regression coefficients are given in Table 5.23.

**Table 5.23 – Logistic regression of baseline TC and stroke**

| MI | Coef | Std Err | P>|z| | 95 % CI | |
|---|---|---|---|---|---|
| HDLC | 0.016 | 0.409 | 0.969 | -0.817 | 0.785 |
| Const | -2.662 | 0.458 | 0.000 | -3.560 | -1.764 |

Substituting these values in the equation, the regression equation is as follows.

$$\text{Logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.016 HDLC - 2.662 \dots\dots\dots\dots(5.24)$$

where $p$ is the probability of stroke.

The odds ratio for every 1mmol/L increase in HDLC is 0.98 [95% CI: 0.44 – 2.19]. Results show that there is no association between the risk of stroke and the increase in HDLC. Cox's proportional hazard model also showed similar results which were not significant (Hz ratio = 0.91 p = 0.84, SE = 0.36, 95% CI = 0.41, 1.98).

## 5.9 Regression dilution bias for cholesterol levels

Total cholesterol was the only cholesterol measurement that was measured in the subsequent phases of the Caerphilly cohort. There were 1190 men who had all the three measurements. The information required to measure the coefficient of measurement error is given in the Table 5.24.

**Table 5.24: Summary statistics for regression dilution bias for TCHOL**

| Variable | observations | Mean | Standard deviation |
|---|---|---|---|
| TCHOL in P2 | 1190 | 5.69 | 1.02 |
| TCHOL in P3 | 1190 | 6.20 | 1.10 |
| TCHOL in P4 | 1190 | 5.69 | 1.05 |
| TCHOL2-TCHOL3 | 1190 | -0.51 | 0.83 |
| TCHOL2-TCHOL4 | 1190 | 0.01 | 0.93 |

The coefficient of measurement error and random fluctuations in TC is estimated using the information given in the above table and the adjusted regression coefficients are given below. These are estimated using the method explained in the section 3.7.1.

**Table 5.25: $\beta$s adjusted with measurement**

|  | $\beta_0$ | $\beta_1$ |
|---|---|---|
| Base Line $\beta$s (phase 2) | -4.400 | 0.372 |
| $\beta$s adjusted for Phase 3 | -5.469 | 0.560 |
| $\beta$s adjusted for Phase 4 | -5.884 | 0.710 |

Using the adjusted $\beta$'s, the probability of having an MI event as a function of TC is estimated and shown in figure 5.20 below.



**Figure 5.20: Predicting MI with TC for all men using baseline and RDB adjusted models**
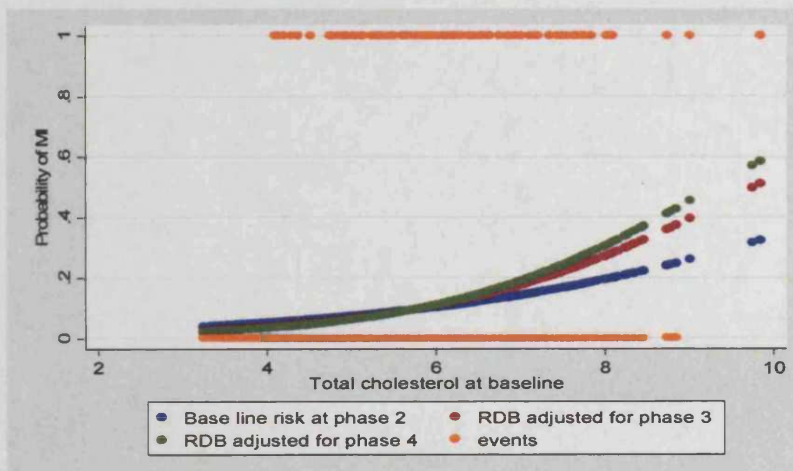
From the above figure, there appears to be an almost twofold increase in the odds of an MI when the coefficients at baseline are adjusted with the coefficient of measurement error and random fluctuations for 4th phase.

## 5.10 Discussion for cholesterol levels

The literature review showed TC as a very significant risk factor for MI which the Caerphilly results confirm (for an increase of 1mmol/L OR = 1.43(95% CI: 1.24-1.64). The Caerphilly odds ratios appear to be fractionally higher then the other studies. This might be because of the difference in the populations. For example, Corti M et al [139] had a sample which included men and women whereas Tervahauta M et al [131] had a sample of elder men without previous heart disease. In the Caerphilly analysis, there are no women included and the age group in the Caerphilly analysis is of middle-aged men whereas the age group in the other studies differ from that of Caerphilly.

The odds ratio for men without prevalent IHD is 0.539 [95% CI: 0.280 − 1.038] which is statistically not significant to show that the risk of MI decreases significantly for an increase of 1mmol/L of HDLC. The Zutphen Elderly study [133] showed that there is no significant association between IHD and HDLC. All the remaining studies in the literature review showed a significant decrease in the risk of IHD for an increase of 1mmol/L increase in HDLC. The difference in the different populations and their age group may be the cause for the results to be non-significant. In addition, the Caerphilly analysis may be underpowered.

Cholesterol levels seem to have little effect on stroke. This is in contrast to the findings of several duties that were reviewed. It may be that there is a small effect of cholesterol on stroke, that was not detected in the Caerphilly data, but there was no evidence of this in the analysis.

The coefficient of measurement error and random fluctuations for the TC estimated using the 3rd and 4th phases of repeated measurements seems increase the odds of MI at baseline by 1.7. This gives an idea of the measurement errors and random fluctuations that could be expected for total cholesterol. This could true to LDL and HDL measurements because of the high correlations between all the three cholesterol levels.

## 5.11 Literature review for Body Mass Index (BMI)

Obesity is a condition of excess body fat. It is the most common form of malnutrition in the Western world. Obesity is an independent risk factor in predicting IHD [106]. It also increases the risk for high blood cholesterol, high blood pressure, and diabetes and, hence, for diseases for which these conditions are risk factors (diabetes, coronary heart disease, high blood pressure etc.). Obesity thus contributes to premature mortality.

In the Framingham Heart Study [140] the relative risk of coronary heart disease in obesity ($Kg/m^2$) was greater than twofold for having CHD with one unit increase in BMI after correction for other known risk factors. A study by Fang et al. [106] showed a hazard ratio for obese($\geq 30kg/m^2$) of 1.32 (CI=1.13–1.53) compared with not obese ($\leq 25kg/m2$) in 9790 subjects aged 25 to 74 years and followed-up for cardiovascular mortality (17 year follow-up data). A meta analysis [141] on BMI and mortality of 6 observational studies showed that hazard ratios for CVD are 1.57 and 1.48 for $25 \leq BMI < 30$ and $30 \geq BMI$ respectively compared to the baseline group $18.5 \leq BMI < 25$.

An American study with 21,414 male physicians participating with a follow-up of 12.5 years, compared participants with BMI less than 23, with those with BMI at least 30. The obese had an adjusted relative risk of 2.00 (95% confidence interval [CI], 1.48-2.71) for total stroke [142]. A prospective study of 7402 healthy middle aged men with 28 years follow-up compared men with low normal weight (BMI, 20.0 to 22.49 kg/m2), and men with BMI >30.0 kg/m2. The later group had a multiple adjusted hazard ratio of 1.93 (95% CI, 1.44 to 2.58) for total stroke [143]. This short literature review suggests that BMI is a significant risk factor for MI and stroke.

## 5.12 The Caerphilly analysis (BMI)

The literature review has shown that BMI is a significant risk factor for CHD. It was measured in all four phases of the Caerphilly cohort. As the analysis starts from the $2^{nd}$ phase, the distributions of BMI in $2^{nd}$, $3^{rd}$ and $4^{th}$ phases are given below in Figure 5.21.



Figure 5.21 – Histograms of BMI in $2^{nd}$, $3^{rd}$, and $4^{th}$ phases
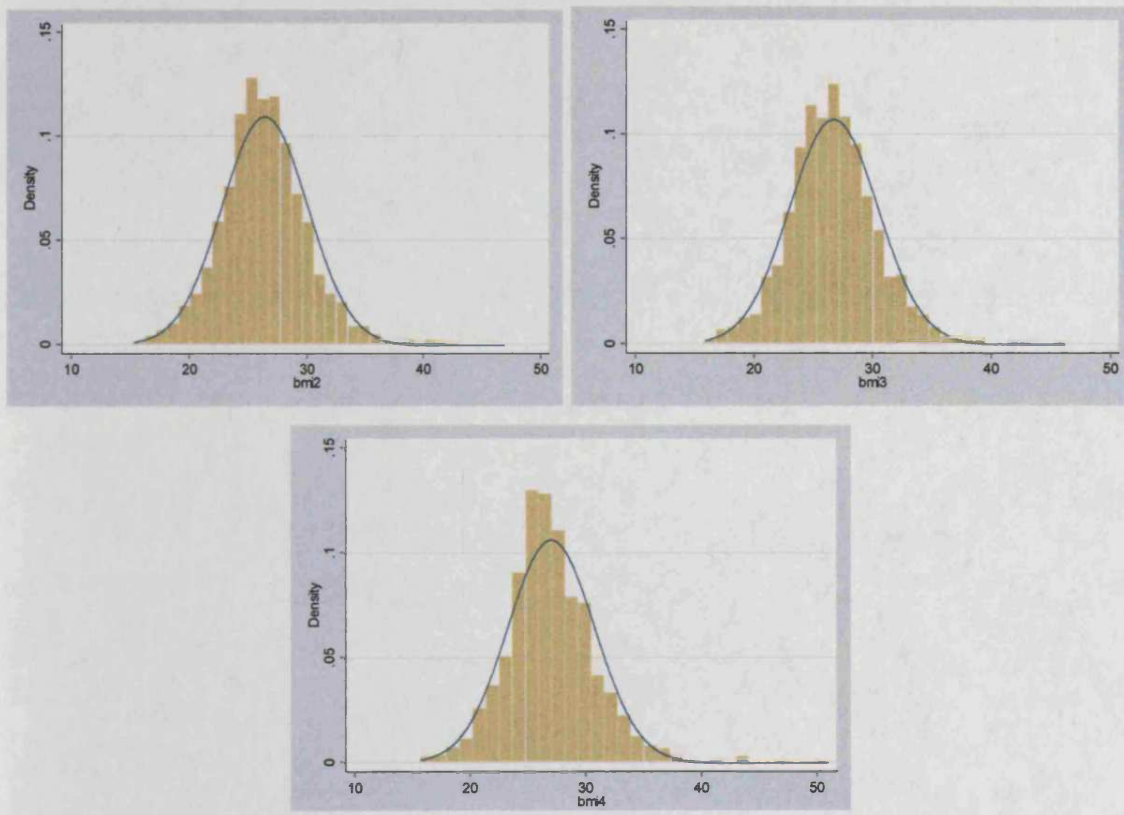
The above histograms of BMI from $2^{nd}$, $3^{rd}$ and $4^{th}$ phases of Caerphilly cohort appear to be approximately normally distributed. After excluding the missing values in the BMI data, the total number of observations used in all 3 phases is given below in Table 5.26, along with the standard deviations.

Table 5.26 – Summary statistics of BMI in $2^{nd}$, $3^{rd}$ and $4^{th}$ phases

| Variables | Obs | Mean | Std. Dev | Min | Max |
|-----------|------|-------|----------|-------|-------|
| BMI in Ph 2 | 1833 | 26.37 | 3.64 | 15.33 | 46.95 |
| BMI in Ph 3 | 1546 | 26.70 | 3.77 | 15.88 | 46.23 |
| BMI in Ph 4 | 1221 | 27.01 | 3.85 | 15.74 | 51.00 |

There were 1177 men with all three measurements. Their summary statistics can be seen in Table 5.29. For these men there was a very slight rise in BMI between phases 2 and 4.

To find the relationship between BMI and MI or stroke, as was done in the previous chapters, Lowess smoothing was used to visualise the relationships and then decide the modelling strategy.

## 5.12.1 MI and BMI

### 5.12.1.1 Linear logistic regression

There were seven observations with BMI more than 40 and 3 with BMI less than 16. These were omitted from the analysis, as they appeared to unduly affect the underlying model. Figure 5.22 shows the relationship between MI and BMI along with the linear model fitted to the data.



**Figure 5.22: Relationship between BMI in phase two and**
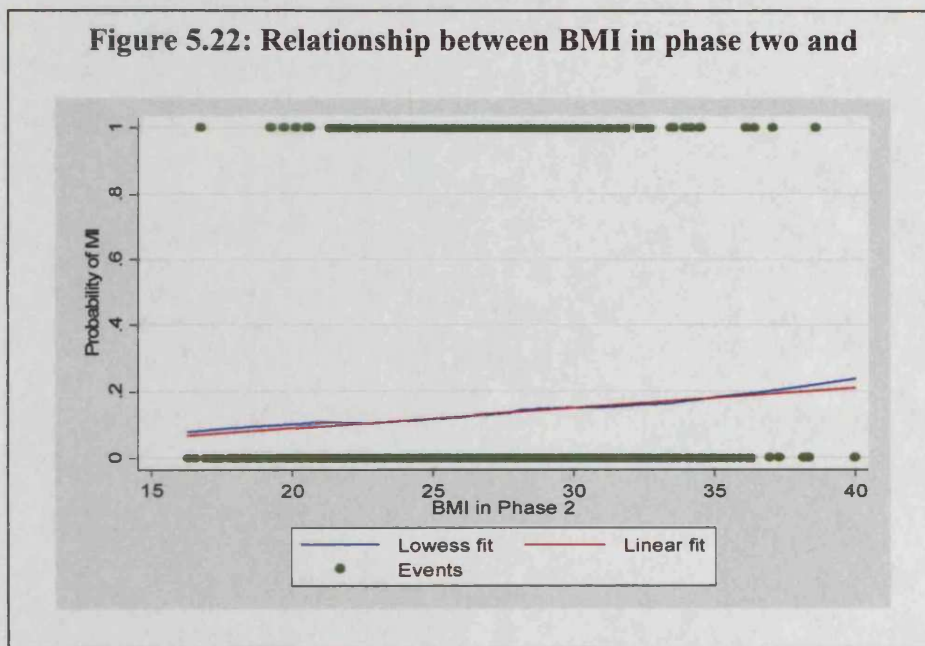
This relationship appears to be linear. The gain in deviance of the fractional polynomial model over the linear model is 0.032; that is not significant suggesting the linear model to be adequate. The coefficients in the linear model are given in the Table 5.27.

**Table 5.27– Logistic regression of baseline BMI and MI**

| MI | Coef | Std Err | P>\|z\| | 95 % CI | |
|---|---|---|---|---|---|
| BMI | 0.053 | 0.022 | 0.014 | 0.011 | 0.096 |
| Const | -3.338 | 0.591 | 0.000 | -4.498 | -2.178 |

Substituting these values in the equation, the regression equation is as follows.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.053BMI - 3.338 \quad\ldots\ldots\ldots\ldots\ldots \quad (5.25)$$

where $p$ is the probability of MI for men without prevalent IHD.

The odds ratio for an increase in one unit of BMI is 1.048[95% CI: 1.01 – 1.09] which is statistically significant to show the risk of MI increases as BMI increases.

## 5.12.1.2 Cox's proportional hazard model

CPHM showed results a better significance compared to the linear logistic model (p=0.012). The Cox's proportional hazard model is shown in the form of an equation as follows.

$$\log\left(\frac{h\{(t),(BMI)\}}{h_0(t)}\right) = 0.045BMI \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5.26)$$

When fractional polynomial model was used, the gain observed was 0.089 which is statistically not significant to show that linear and fractional polynomial models were different.

## 5.12.2 BMI and Stroke

To give a preliminary idea of the relationship between BMI and stroke, the Lowess smoothed plot is shown in Figure 5.23. A linear model along with the fractional polynomial model is also fitted.

**Figure 5.23: Relationship between BMI in phase two and stroke**
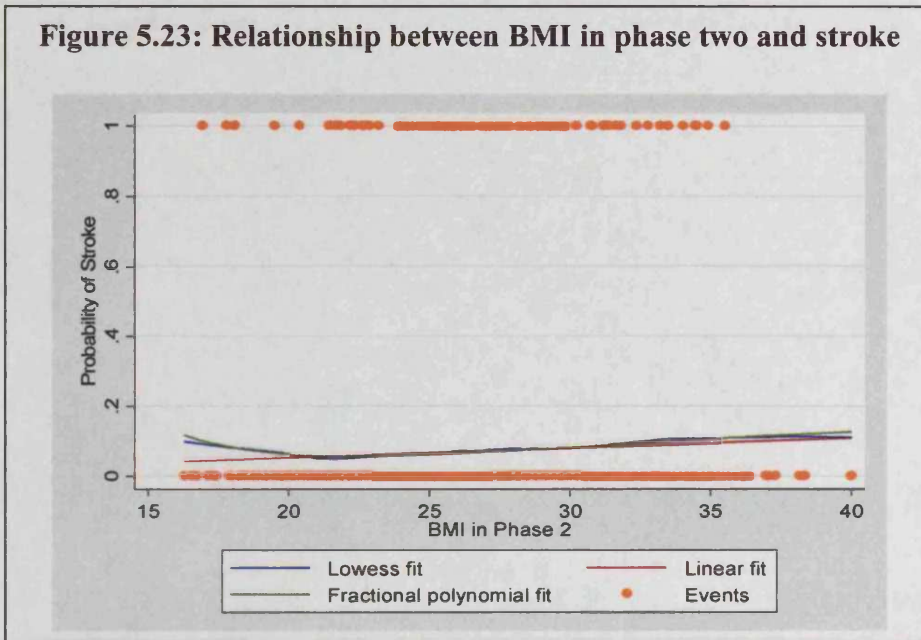


This relationship appears to be non-linear at the lower end of the BMI range. With an improvement of deviance equal to 1.03, fractional polynomial model is not statistically significantly better then the linear model and so the linear logistic model was be used to estimate the coefficients of the relationship. The coefficients of the linear logistic model are shown in the Table 5.28.

**Table 5.28– Logistic regression of baseline BMI and Stroke**

| Stroke | Coef | Std Err | P>|z| | 95 % CI | |
|--------|------|---------|-------|---------|---|
| BMI | 0.040 | 0.028 | 0.156 | -0.015 | 0.095 |
| Const | -3.617 | 0.757 | 0.000 | -5.102 | -2.133 |

Table 5.28 can be represented in an equation as follows

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.040 bmi - 3.617 \dots\dots\dots\dots \quad (5.28)$$

where $p$ is the probability of having a stroke.

The odds ratio is 1.04 with 95% confidence interval as [0.99, 1.10]. These results show that the there is no increased risk of stroke with increase in BMI. The Cox's proportional hazard model also showed no significant results (Hz ratio = 1.03 p = 0.23, SE = 0.03, 95% CI = 0.98, 1.08).

## 5.13 Regression dilution bias for BMI

BMI was also measured in subsequent phases of the Caerphilly cohort. There were 1459 men who had all the three measurements. Seven men with BMI more than 40 were excluded from the analysis. The information of 1452 men required to measure the regression dilution bias is given in the Table 5.29.

**Table 5.29: Summary statistics for regression dilution bias for MI and BMI**

| Variable | observations | Mean | Standard deviation |
|----------|--------------|------|--------------------|
| BMI in P2 | 1177 | 26.52 | 3.64 |
| BMI in P3 | 1177 | 26.78 | 3.75 |
| BMI in P4 | 1177 | 27.02 | 3.81 |
| BMI2-BMI3 | 1177 | -0.37 | 1.46 |
| BMI2-BMI4 | 1177 | -0.50 | 1.76 |

The coefficient of measurement error in BMI is estimated using the information given in the above table and the regression coefficients adjusted with the coefficient of measurement error are given in the following table.

**Table 5.30: $\beta$s adjusted with measurement**

| | $\beta_0$ | $\beta_1$ |
|---|-----------|-----------|
| Base Line $\beta$s (phase 2) | -3.550 | 0.042 |
| $\beta$s adjusted for Phase 3 | -3.655 | 0.046 |
| $\beta$s adjusted for Phase 4 | -3.709 | 0.048 |

In the above table the first row represents the regression coefficients at the baseline i.e., the regression coefficients to the logistic model fitted to the BMI and MI. The second row represents the coefficients adjusted to the regression dilution bias due to

the variation in BMI using the 3$^{rd}$ phase BMI values whereas the 3$^{rd}$ row represents the coefficients adjusted using the 4$^{th}$ phase BMI values. The Figure 5.24 shows the probability of MI when fitted with the baseline $\beta$s and the adjusted $\beta$s in the subsequent phases.



Figure 5.24: Predicting MI with BMI for all men using baseline and RDB adjusted models
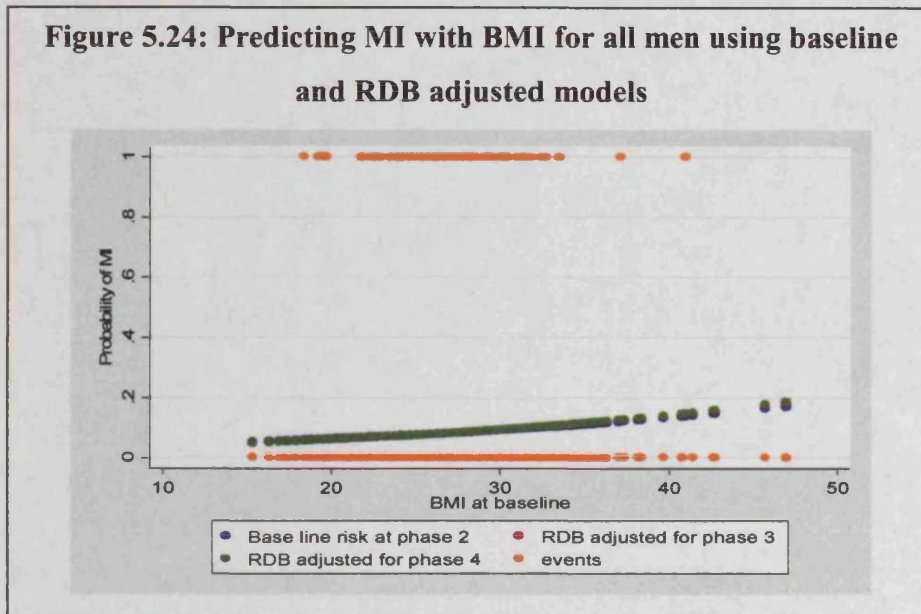
These results and the above plot show that the lack of measurement error means the adjusted values are similar to the unadjusted ones (Table 5.30).

## 5.14 Discussion for BMI

For every 5 kg/m2 increase in BMI for all men, the odds ratio is 1.30 [95% CI: 1.11 – 1.54] for MI. When this result was compared with the different studies from the literature review, the risk seems to be comparable. When the men with 25≤BMI<30, and 30≥BMI compared with men with 18.5≤BMI<25 as a reference group, the odds ratios were 1.40 and 1.67 [141]. The outcome variable in the Meta analysis was CVD whereas the outcome variable here is MI. This may be the reason for the results being slightly different. In spite of slight variations in the results and a huge variations in the samples, both (the Caerphilly and the Meta analysis) showed significantly increased risk when the later two groups were compared with the group with 18.5≤BMI<25. The association of BMI and stroke appears to be not significant in the Caerphilly analysis whereas, one of the studies in the literature review suggested that BMI is associated with an increased risk of stroke. After adjusting the coefficients of the

119

model based on baseline data using the repeated measures of BMI, as an indication of within subject variability, it appears that BMI is a reliable measure and there is very little change after adjustment.

## 5.15 Overall conclusions for the chapter

This chapter dealt with the biological factors measured in the Caerphilly study in order to measure the risk of MI and stroke associated to these factors. Brief literature reviews were carried out before the analysis of each risk factor to obtain some idea of each risk factor. The objective of this chapter was to measure the association of risk of MI and stroke with these factors and establish clear idea on each factor in order to use that information in the hypothetical simulations.

Some of the studies from the literature review suggested that the risk of MI can be measured more accurately using the pulse pressure [122-124]. However, the Caerphilly data analysis suggests that pulse pressure, although an important risk factor for VAD, is not preferable to SBP or DBP. Cholesterol and obesity were also confirmed as risk factors in the analysis.

The value of this analysis is that each factor has been analysed thoroughly using a variety of strategies. The Caerphilly data have not been analysed as thoroughly previously. Of further importance was the demonstration and estimation of measurement error and random fluctuations with the repeated measurements, which has not been shown previously in the analyses of Caerphilly data.

# Chapter 6

# 6.0 Psychological variables

Most of the psychological measurements were carried out in the 2$^{nd}$ phase of the Study. Since attitudes towards coronary related behaviours were thought to be related to the psychological factors, Health Attitude Inventory (HAI) was analysed as one of the psychological factors. Only GHQ30 & HAI were re-measured in later phases. However, HAI was re-measured using a shortened questionnaire. In this chapter, four psychological variables will be considered; namely depression (measured using GHQ30), trait anxiety (using STAI), anger and Type A behaviour (using JAS) along with HAI.

## 6.1 Psychological distress

### 6.1.1 Literature review

Typically, the term 'psychological distress' covers a broad range of symptoms and disorders that are largely related to depression and anxiety.

Although the nature of psychological distress as measured by the GHQ is unclear, several studies have shown that GHQ scores are associated with an increased risk of CHD [33, 144-146]. GHQ was associated with up to a threefold increased risk of CHD in men, that was sustained even after adjustment for CHD risk factors and other confounders [33]. The Whitehall II study has shown that baseline GHQ score among men with no previous heart disease was associated with an increased risk of CHD (OR = 1.83; 95% CI: 1.5 – 2.3) [33].

In CaPS, in a validation study of 97 men, of the 39 men who scored above the standard threshold >4, 20 were found to have psychiatric symptoms by clinical interview, of whom 17 showed symptoms of depression [33]. It may be concluded that the psychiatric caseness, as assessed in CaPS, has a large depression component. There is a larger literature relating depression to VAD [33].

Several studies showed no evidence of a clear association between MI and depression. Vogt et al [147] in 1994, with a sample size of 2573 men and women, showed that

depression has no association with CHD ( RR = 0.94 [95% CI: 0.70 – 1.28]) depression was measured with the Lagner mental health Index. A college alumni health study in USA [148] with 5053 men showed no association between physician-diagnosed depression and fatal CHD (RR = 1.20 (95% CI: 0.53 – 2.71]) when adjusted for age, smoking BMI, exercise, alcohol, hypertension. The National Health and Nutrition Examination Study [149] showed that association of the risk of fatal CHD with depression (measured with the General Well-Being Schedule) was statistically not significant for either black or white men when adjusted for age, smoking, cholesterol, BMI, exercise, hypertension, diabetes, replacement hormones and education.

The Epidemiologic Studies Depression Scale (CES-D) was used to measure depression in the NHANES I study. This showed the RR of CHD incidence among depressed men was 1.71 (95% CI, 1.14-2.56) compared with non depressed men when adjusted for age [145]. In a multivariate analysis, with all standard risk factors included for adjustment, the men who reported clinical depression were at significantly greater risk for CHD (RR = 2.12; 95% CI: 1.24-3.63 ) [150].

It is recognized that up to 30% of stroke victims suffer from depression [151]. Several recent epidemiological studies reported positive associations between psychological distress and CHD risk. However, fewer studies have examined the association with stroke, and the results of these were inconsistent or those studies had their own limitations. A prospective US study in an elderly population found no positive association between depressive symptoms and stroke mortality after adjustment for known risk factors [152]. However, in the Alameda County study, stroke mortality was increased in individuals with depressive symptoms after controlling for possible confounders [153]. A study comparing three studies in the United States showed that the risk of stroke in elderly hypertensive men and women who reported high levels of depressive symptoms was more than twice that of non-depressed hypertensive men [154]. A paper published on the Caerphilly study showed that depression (GHQ score) is a predictor of fatal ischemic stroke but not of nonfatal ischemic stroke or Transient Ischemic Attack (TIA) [155].

## 6.1.2 Statistical analysis

The GHQ30 was administered in the 2nd, 3rd and 4th phases of the Study. The scoring of this questionnaire was discussed in chapter 2. The following Figure 6.1 shows the distributions of psychological distress scores measured in these phases.

**Figure 6.1 Histograms of psychological distress scores in 2nd, 3rd, and 4th phases**
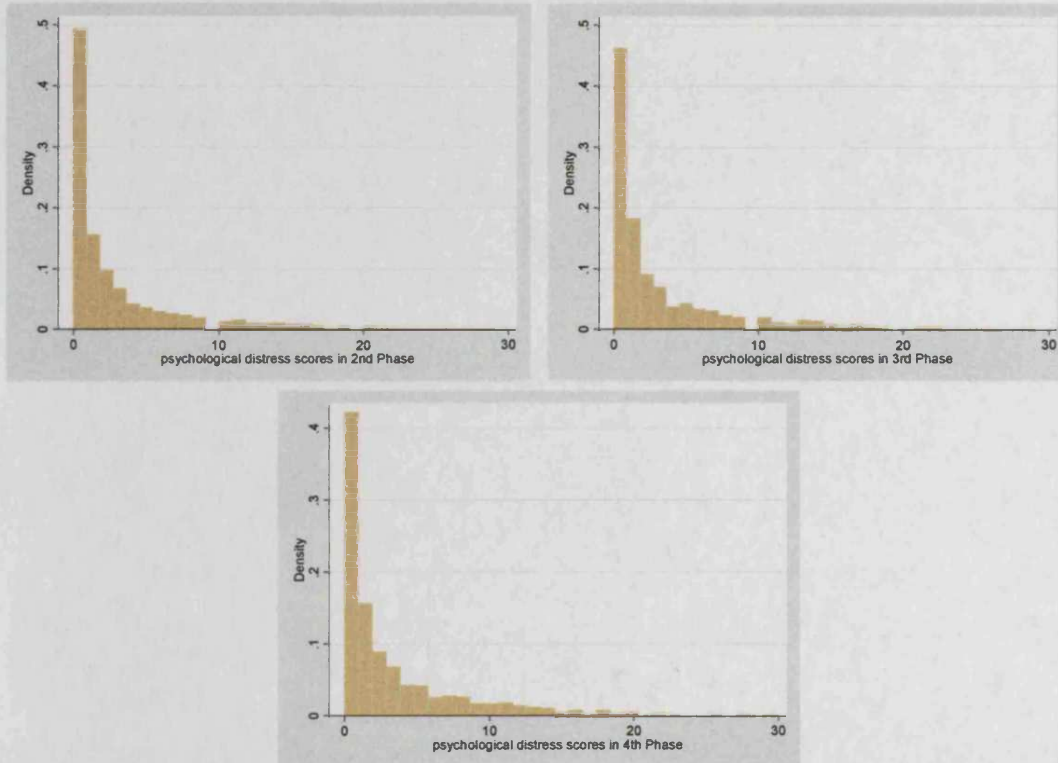


From the above histograms, it is clear that GHQ is very positively skew. After excluding the missing values, the total number of subjects who were included in the analysis for the 2nd phase was 1724. When the data was recoded to a binary variable to use as an indicator variable for depression, 490 subjects were identified as psychologically distressed (PD). This was done according to Stansfeld et all [33]. The summary statistics of the PD Indicator (PDI) of the three phases is given in the Table 6.1.

**Table 6.1 – Summary statistics of Psychological Distress Indicator in $2^{nd}$, $3^{rd}$ and $4^{th}$ phases**

| Variables | Obs | Mean | Std. Dev | Min | Max |
|---|---|---|---|---|---|
| PD in Ph 2 | 1724 | 0.196 | 0.397 | 0 | 1 |
| PD in Ph 3 | 1375 | 0.215 | 0.411 | 0 | 1 |
| PD in Ph 4 | 1204 | 0.221 | 0.415 | 0 | 1 |

From the above table, it could be observed that between 20-22% of subjects were psychologically distressed. There were 947 subjects present in all three phases, 21-23% of them with psychological distress.

### 6.1.2.1 Psychological Distress Indicator and MI

When men with PD (PDI = 1) are compared with men without PD (PDI = 0) for having an MI the odds ratio for having an MI was 1.18 [95% CI: 0.84 – 1.67]. These results are estimated using logistic regression.

Survival analysis also showed very similar results compared to the logistic regression. The hazard ratio for having an MI for men with PDI=1 compared with PDI=0 is 1.21 [95% CI: 0.88 – 1.66].

### 6.1.2.2 Psychological Distress Indicator and Stroke

When men with PD are compared to men without PD for the risk of stroke, there was no significant difference. The odds ratio and hazard ratio of stroke for men with PD compared to men without PD are 0.83 [95% CI: 0.51 – 1.34] and 0.84 [95% CI: 0.53 – 1.35] respectively.

### 6.1.3 Regression dilution bias

Repeated GHQ30 scores are available. However, the nature of the measurement itself makes it difficult to measure the regression dilution bias. Figure 6.1 showed the histograms of the GHQ30 scale in different phases, and it is clear that the distribution of GHQ30 is not normal. The basic assumption in measuring the error in measurement and random variations is to assume that the independent factor follows a normal distribution but this does not apply here.

However, the association of risk of MI with the GHQ score was measured using a psychological distress indicator (PDI) which is binary. This binary variable can be derived for the subsequent measurements of GHQ30 and can be used for estimating the regression dilution bias. Though, this method may produce a very crude estimate of the measurement error and random variation, nevertheless it may show some light on the these errors of the GHQ30 in particular and of psychological risk factors in general.

There were 116 men who had PDI as zero in phase 2 and 1 in phase 3. Similarly there were 103 men with PDI = 1 in phase 2 and PDI = 0 in phase 2. This shows that there were some men who changed from being depressed to not depressed and vice versa from phase 2 to phase 3. All those men with no repeated measurements were excluded from the analysis.

The required information for estimating the adjusted odds ratio is given in the Table 6.2. The method for estimating the adjusted odds ratio is straight forward and explained in detail elsewhere [156]. The formula for estimating the new odds ratio is given below.

**Table 6.2: Tabulated values of MI indicator and PDI in phase 2 & 3**

| | PDI in Phase 2 | | |
|---|---|---|---|
| MI in Phase 2 | 0 | 1 | TOTAL |
| 0 | $a_1$=744 | $b_1$=168 | 912 |
| 1 | $c_1$= 76 | $d_1$= 25 | 101 |
| TOTAL | 823 | 193 | 1013 |
| | PDI in Phase 3 | | |
| MI in Phase 2 | 0 | 1 | TOTAL |
| 0 | $a_2$=734 | $b_2$=178 | 912 |
| 1 | $c_2$= 72 | $d_2$= 29 | 101 |
| TOTAL | 806 | 207 | 1013 |
| | U = 207/806 V = 193/823 | | |

From the above table, the odds ratio for having an MI for men with $PDI_{phase\ 2}=1$ is OR = 1.4568. The odds ratio is slightly different to the results from 6.1.2.1. This is because of the subjects with missing data were excluded for the analysis. Using the method described in the section 3.7.3 (Case 2), the adjusted OR is given by the following formula.

$$\hat{OR} = \frac{\hat{a_1}\hat{d_1}}{\hat{b_1}\hat{c_1}} = OR \times \frac{\left(1 - \dfrac{a_2 \times (1-U)}{a_1 \times U}\right)\left(1 - \dfrac{d_2 \times (1-U)}{d_1 \times U}\right)}{\left(1 - \dfrac{b_2 \times (1-U)}{b_1 \times U}\right)\left(1 - \dfrac{c_2 \times (1-U)}{c_1 \times U}\right)} \dots\dots\dots\dots(6.1)$$

Then the adjusted odds ratio is

$$\hat{OR} = OR \times 1.215 = 1.7699 \approx 1.77$$

The odds ratio for having an MI among depressed men compared with non-depressed men was 1.46 [95% CI: 0.90 – 2.36]. That the odds ratio is not significant may be due to the very small number of events (MIs were 101 out of 1013 subjects). Risk of MI appears to increase by 21.5% when adjusted by the regression dilution bias.

The other method explained in the statistical methods (section 3.7.3 Case 1) is not used here as there is no literature available on the sensitivity and specificity of the repeated measurements on GHQ30. The only disadvantage with the above method is that it does not give estimates for the standard error but gives an idea of the measurement error and random variations in the psychological depress indicator.

## 6.1.4 Discussion

Several studies have shown that psychological distress or depression are associated with MI, while others have not. The CaPS analysis found no significant association. PD was measured using the GHQ30 questionnaire in the Caerphilly study. Other studies have used different questionnaires to measure depression. It appears that this depression indicator defined by using the GHQ30 questionnaire might not be as good a measure of depression as a dedicated depression scale.

The other important issue that needs to be mentioned here is the regression dilution bias and random variations. Applying a correction factor for these biases increased the odds by 22%. This shows that that there exists some error in measuring psychological variables. One should be very careful in concluding these results because the repeated measurements are separated by four to five years and there is a genuine possibility that subjects' depression levels have changed. This is one of the problems of converting a continuous scale like GHQ30 into a binary outcome. However, it would be important to know there is some measurement error or genuine change in the depression over the period of time. This justifies the need to measure psychological factors more than once to estimate any existing bias.

## 6.2 Anxiety

### 6.2.1 Literature review

Anxiety is generally defined as a psychobiological emotional state or reaction that can be distinguished most clearly from other emotions such as anger or sadness by feelings of tension, apprehension, nervousness, and worry. Changes in anxiety levels also have an effect on different risk factors of CHD that generally include blood pressure, smoking etc [157].

There are relatively few studies relating anxiety directly to VAD. The Normative Aging Study with 498 men showed that there is an increased risk of CHD for very anxious persons [158]. Hagman et al, 1987 followed 5735 men for between 2-7 years and found anxiety was associated with angina but not MI [159]. Haines et al. 1987 followed 1457 men for 10 years and found phobic anxiety to be strongly related to CHD and fatal MI (OR=3.8) [160]. Kawachi et al (1994) followed 33,999 men for 2 years and found phobic anxiety to be associated with CHD and MI [161].

Prospective evidence linking anxiety to stroke could not be found, although high anxiety has been linked to high blood pressure, which is a risk factor for stroke [162].

Analysis of CaPS data linking anxiety to VAD has not been previously conducted.

### 6.2.2 Statistical analysis

Trait anxiety was measured in the second phase of the Caerphilly study using the 20 item trait scale of the State Trait Anxiety Questionnaire (STAI). There were no repeated measurements taken. The following figure gives the distribution of the anxiety scores.

**Figure 6.3 Histogram of Trait anxiety score**
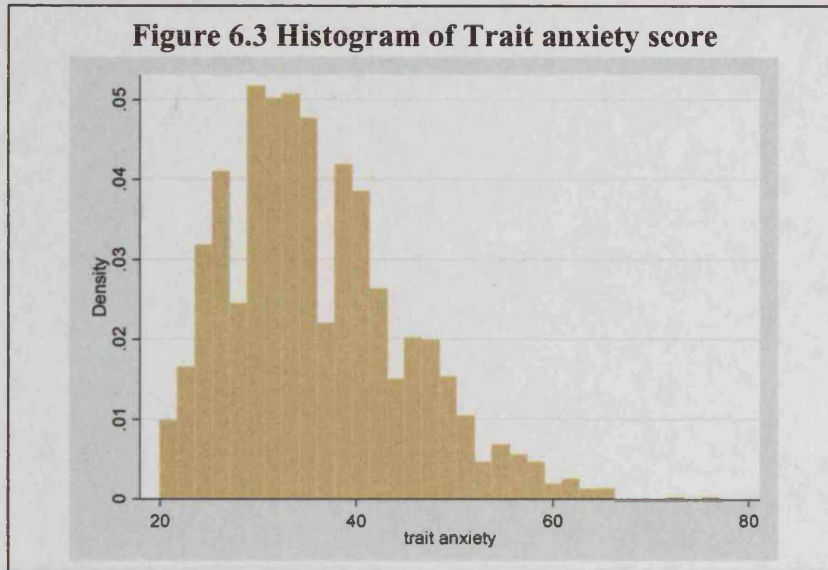
From the above plot, it can be observed that the distribution of trait anxiety appears to be slightly positively skew. After excluding the missing values, the number of subjects included in the analysis was 1838. Summary statistics of the anxiety score are given in the below Table 6.3.

**Table 6.3 – Summary statistics of trait anxiety score**

| Variables | Obs | Mean | Std. Dev | Min | Max |
|-----------|-----|------|----------|-----|-----|
| STAI | 1838 | 35.93 | 8.94 | 20 | 77 |

To identify an appropriate strategy for modelling the relationship and MI or stroke Lowess smoothed plots were examined. The Lowess smoothed plots for the relationship between anxiety score and MI and stroke is given in the Figure 6.4.



**Figure 6.4 Relationship between trait anxiety score and MI, Stroke**
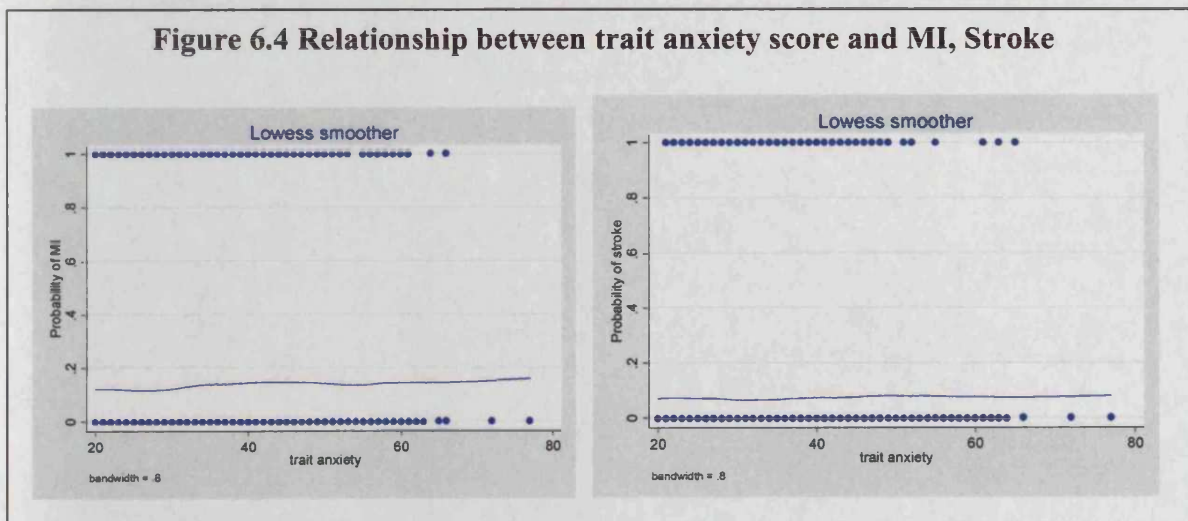
The smoothing parameter used in both cases is 0.8. If the relationship between STAI and MI is examined more closely, there is slight evidence for an increased risk in MI up to a score of 38, beyond which the risk is constant. There is no evidence for any non linearity in risk for stroke. For completeness, anxiety was modelled both as a continuous variable and also as a binary variable with a cut-point at 38.

### 6.2.2.1. Anxiety and MI

The table below (Table 6.4) shows the odds ratios and hazard ratios for having an MI according to anxiety score modelled as a continuous variable.

**Table 6.4: Odds and Hazard ratios of MI with Trait anxiety score**

|  | ODDS RATIO (95% CI) | HAZARD RATIO (95% CI) |
|---|---|---|
| MI | 1.01 [95% CI: 0.99 – 1.02] | 1.01 [95% CI: 0.99 – 1.02] |

From table 6.4 it is clear that there is no convincing evidence of an association of anxiety score with MI either with logistic Cox's proportional hazards models. However, it should be noted that the confidence intervals are very narrow but given then range of STAI (20-80), a one unit change in STAI will be very tiny.

If the analysis is repeated using anxiety as a binary variable where a cut-point of 38 is used to identify high and low anxiety groups, for MI odds and hazard ratios are OR=1.28 [95% CI: 0.98 – 1.69] and HR=1.28 [95% CI: 0.99 – 1.65] respectively. Although not formally significant, the confidence intervals show there is some evidence of an association. The Kaplan-Meier survival plots comparing men with high and low anxiety confirm this conclusion in Figure 6.5. The survival lines for high and low anxiety groups consistently diverge over time. This suggests a very slight effect of anxiety on MI.

**Figure 6.5 Kaplan-Meier MI survival plots comparing men with low and high anxiety**
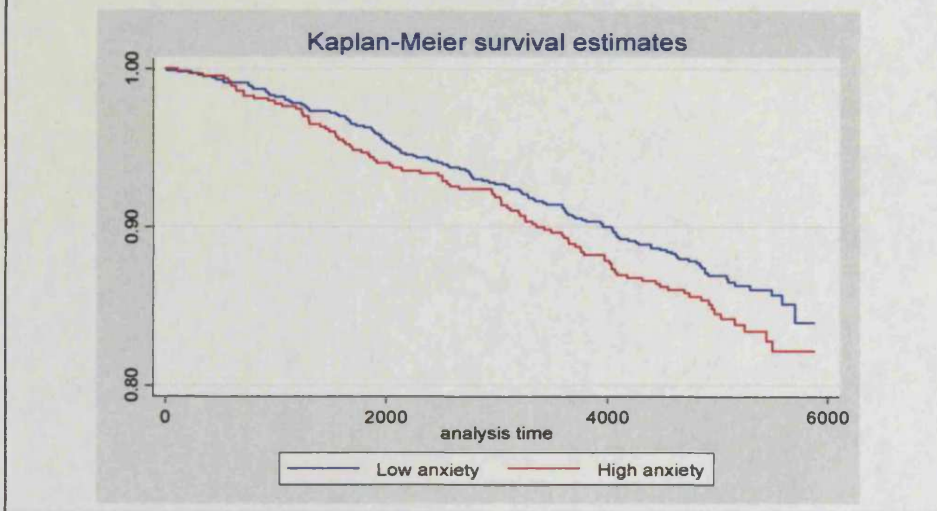
## 6.2.2.2. Anxiety and Stroke

The analysis was repeated for stroke. If anxiety is modelled as a continuous variable the odds and hazard ratios do not show convincing evidence of an association (table 6.5).

**Table 6.5**

**Odds and Hazard ratios of Stroke with Trait anxiety score**

|  | ODDS RATIO (95% CI) | HAZARD RATIO (95% CI) |
|---|---|---|
| STROKE | 1.01 [95% CI: 0.99 – 1.03] | 1.01 [95% CI: 0.99 – 1.03] |

If anxiety is modelled as a binary variable the odds ratio and hazard ratios are 1.28 [95% CI: 0.89 – 1.83] and 1.26 (95% CI: 0.89 – 1.78) respectively. Figure 6.6 shows that there is no consistent divergence over time between high and low anxiety groups for stroke incidence, suggesting that there is no association of anxiety with stroke.

**Figure 6.6 Kaplan-Meier stroke survival plots comparing men with low and high anxiety**
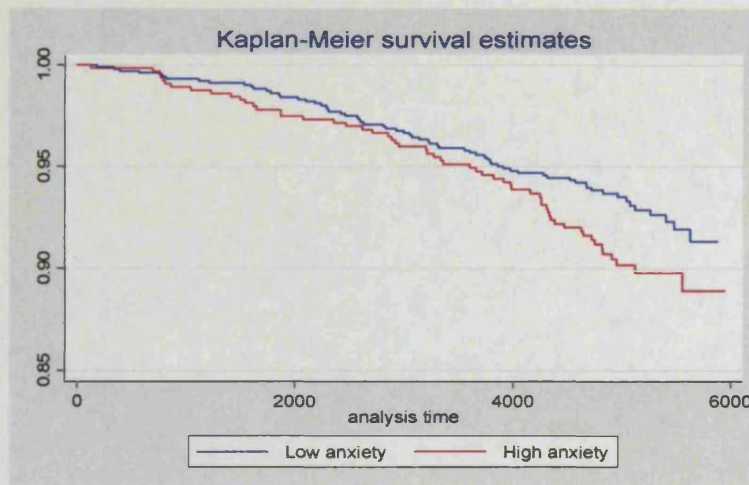


### 6.2.3 Discussion

The evidence for an association of anxiety with CHD is weak in the CaPS data. This is in conflict with several previous studies which found strong associations [159, 161]. Reasons for this difference may be due to differences in the outcome of interest between studies. In the Hagman study anxiety was only associated with angina and not with MI. This is consistent with the CaPS analysis [159]. In the Haines and Kawachi studies anxiety was associated with fatal CHD which included MI [160, 161]. These findings are in contrast to those found in those studies.

This contrast between the CaPS result and those of Haines et al and of Kawachi et al may be due to differences in sample selection. However, all of these studies used middle aged men and excluded men with previous CHD and it is unlikely that differences in sample selection have caused that effect. A further reason might be differences in anxiety measurement between studies. The Haines and Kawachi studies used the phobic anxiety sub-scale of the Crown-Crisp Experiential Index. This sub-scale assesses an anxiety response towards specific situations such as crowds or enclosed spaces, rather than a generalised tendency towards anxiety as is measured in the STAI. It is possible, therefore that a stronger association was found with phobic anxiety than trait anxiety.

132

The evidence from CaPS on anxiety and stroke is that they are not associated. No other relevant literature was found with which to compare these results.

## 6.3 Anger

### 6.3.1 Literature Review

Anger is a psychological response to a perceived unfairness or injustice. It is a complex emotion and although it may be episodic as in responding to specific environmental situations, some people have a greater tendency to anger and experience more extreme anger responses than others. Anger can be expressed actively in terms of behaviour designed to address the cause of the anger; these are typically aggressive verbal or physical displays. Anger can also be expressed passively as in behaviour designed to contain the anger response. A passive response is often expressed in terms of cynicism, distrust, uncooperativeness and ill-will towards the object (usually a person) of the anger. This passive response is often described as a separate aspect of anger called hostility.

The idea that anger may predict VAD is very old [163]. Several studies have supported the suggestion that anger predicts heart disease [164, 165] whereas some showed no significant association between anger and VAD [166, 167]. Everson et al [168] showed that men with hostility scores in the top quartile were at more than twice the risk of all-cause mortality (relative hazards (RH) 2.30, 95% CI: 1.47 – 3.59) and cardiovascular mortality (RH 2.70, 95% CI: 1.27 – 5.76), relative to men with scores in the lowest quartile. This study examined the association between hostility, measured by the eight-item Cynical Distrust Scale [169], and risk for all-cause and cardiovascular mortality and incident myocardial infarction. Another study which used the Cynical Distrust Scale showed that the effects of hostility on health adjusted for standard cardiovascular risk factors showed an odds ratios of 2.3 (1.8 –2.8) for men [170]. A meta-analytic review of several studies showed that hostility is associated with VAD [171]. This study included 15 studies from already concluded systematic reviews and 30 other independent studies to conclude that hostility is an independent risk factor for CHD.

In the Caerphilly study, anger was measured using eleven questions from the Framingham anger scales. A suppressed anger score was constructed using responses to three of those questions [25]. In that analysis, the anger-out scale and the

suppressed anger scale showed an increased risk of IHD [25]. The association of the anger scales with MI and Stroke is examined in more detail here.

## 6.3.2 The Caerphilly analysis

Anger was measured using Framingham anger scales in Caerphilly cohort. The four anger scales used were

1) Anger-in
2) Anger-out
3) Anger-symptoms
4) Anger-discuss

Anger-in scale measures the anger that is not expressed but is 'bottled up' inside. The anger-out scale measures the anger that is expressed on others. The anger-symptoms scale tries to measure the somatic effects of anger such as getting tensed, headache and feeling weak etc. The anger-discuss scale attempts to measure whether the subject tries to discuss his reason for anger with his friends or relatives. In psychometric terms, these scales have had little development and no evidence on validity is available. The distributions of these variables are given in the Figure 6.7.

A brief description of these scales is given in the section 2.3.3.2. The 11 items used for the construction of these scales are given in the appendix 2.

**Figure 6.7: Framingham anger scores**
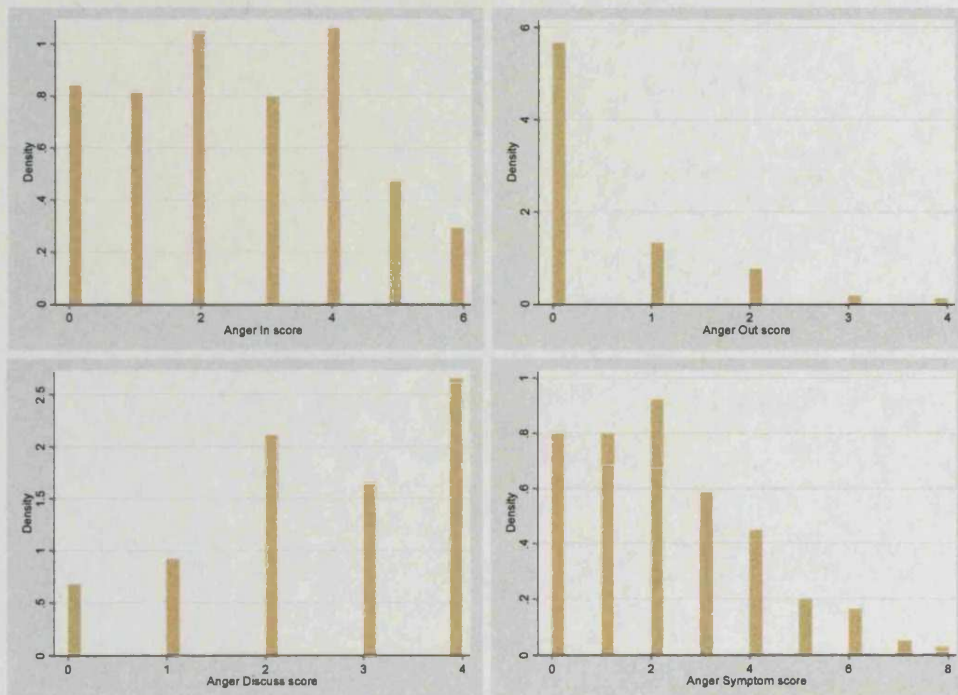
None of the above anger scales are symmetrically distributed. As a preliminary analysis the incidence of MI or stroke was compared for each score within each scale. The following table shows the tabulations of anger scales and MI incidence.

## Table 6.6 tabulating anger scales and incidence of MI

### Anger-in

| MI | 0(%) | 1(%) | 2(%) | 3(%) | 4(%) | 5(%) | 6(%) | Total |
|---|---|---|---|---|---|---|---|---|
| No | 237 (83.5) | 234 (86.4) | 306 (86.4) | 227 (84.7) | 322 (89.7) | 139 (89.1) | 83 (83.8) | 1548 (86.4) |
| Yes | 47 (13.6) | 37 (13.6) | 48 (13.6) | 41 (15.3) | 37 (10.3) | 17 (10.9) | 16 (16.2) | 243 (13.6) |
| Total | 284 | 271 | 354 | 268 | 359 | 156 | 99 | 1791 |

### Anger-out

| MI | 0(%) | 1(%) | 2(%) | 3(%) | 4(%) | Total |
|---|---|---|---|---|---|---|
| No | 1075 (84.7) | 268 (91.5) | 151 (89.4) | 33 (91.7) | 21 (87.5) | 1548 (86.4) |
| Yes | 194 (15.3) | 25 (8.5) | 18 (10.6) | 3 (8.3) | 3 (12.5) | 243 (13.6) |
| Total | 1269 | 293 | 169 | 36 | 24 | 1791 |

### Anger Discuss

| MI | 0(%) | 1(%) | 2(%) | 3(%) | 4(%) | Total |
|---|---|---|---|---|---|---|
| No | 136 (90.1) | 176 (84.2) | 404 (85.2) | 324 (88.8) | 508 (85.8) | 1548 (86.4) |
| Yes | 15 (9.9) | 33 (15.8) | 70 (14.8) | 41 (11.2) | 84 (14.2) | 243 (13.6) |
| Total | 151 | 209 | 474 | 365 | 592 | 1791 |

### Anger Symptoms

| MI | 0(%) | 1(%) | 2(%) | 3(%) | 4(%) | 5(%) | 6(%) | 7(%) | 8(%) | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| No | 304 (84.7) | 299 (83.5) | 364 (88.4) | 231 (87.2) | 175 (87.5) | 80 (89.9) | 64 (84.2) | 20 (95.1) | 11 (100) | 1548 (86.4) |
| Yes | 55 (15.3) | 59 (16.5) | 48 (11.6) | 34 (12.8) | 25 (12.5) | 9 (10.1) | 12 (15.8) | 1 (4.9) | 0 (0) | 243 (13.6) |
| Total | 359 | 358 | 412 | 265 | 200 | 89 | 76 | 21 | 11 | 1791 |

When the chi square test for trend was performed to check if there is trend in having an MI for different anger scales, the *p* values of 0.27, 0.02, 0.3 and 0.28 for anger-in, anger-out, anger-discuss and anger-symptoms respectively. This suggests that apart from anger-out, none of the scales show associations with MI.

A comparable analysis was conducted for stroke. When the chi square test for trend was performed to check if there is trend in having a stroke for different anger scales, the *p* values of 0.32, 0.66, 0.86 and 0.98 for anger-in, anger-out, anger discuss and anger symptoms respectively. This suggests that none of the scales as associated with stroke. The case frequencies for these calculations are given in table 6.7.

**Table 6.7 tabulating anger scales and incidence of stroke**

**Anger-in**

| Stroke | 0(%) | 1(%) | 2(%) | 3(%) | 4(%) | 5(%) | 6(%) | Total |
|---|---|---|---|---|---|---|---|---|
| No | 269 (92.4) | 264 (94.6) | 341 (94.5) | 248 (90.2) | 342 (92.7) | 151 (92.1) | 90 (90.0) | 1705 (92.7) |
| Yes | 22 (7.6) | 15 (5.4) | 20 (5.5) | 27 (9.8) | 27 (7.3) | 13 (7.9) | 10 (10.0) | 134 (7.3) |
| Total | 291 | 279 | 361 | 275 | 369 | 164 | 100 | 1839 |

**Anger-out**

| Stroke | 0(%) | 1(%) | 2(%) | 3(%) | 4(%) | Total |
|---|---|---|---|---|---|---|
| No | 1200 (92.2) | 283(94.3) | 162(93.1) | 36(94.7) | 24(96.0) | 1705(92.7) |
| Yes | 102 (7.8) | 17(5.7) | 12(6.9) | 2(5.3) | 1(4.0) | 134(7.3) |
| Total | 1302 | 300 | 174 | 38 | 25 | 1839 |

**Anger Discuss**

| Stroke | 0(%) | 1(%) | 2(%) | 3(%) | 4(%) | Total |
|---|---|---|---|---|---|---|
| No | 148(94.9) | 194(92.8) | 446(92.2) | 352(92.6) | 565(92.6) | 1705(92.7) |
| Yes | 8(5.1) | 15(7.2) | 38(7.8) | 28(7.4) | 45(7.4) | 134(7.3) |
| Total | 156 | 209 | 484 | 380 | 610 | 1839 |

**Anger Symptoms**

| Stroke | 0(%) | 1(%) | 2(%) | 3(%) | 4(%) | 5(%) | 6(%) | 7(%) | 8(%) | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| No | 347 (93.5) | 338 (91.8) | 392 (92.9) | 254 (93.4) | 192 (93.2) | 82 (91.1) | 69 (90.8) | 20 (90.9) | 11 (91.7) | 1705 (92.7) |
| Yes | 24 (6.5) | 30 (8.2) | 30 (7.1) | 18 (6.6) | 14 (6.8) | 8 (8.9) | 7 (9.2) | 2 (9.1) | 1 (8.3) | 134 (7.3) |
| Total | 371 | 368 | 422 | 272 | 206 | 90 | 76 | 22 | 12 | 1839 |

Due to the relatively small numbers of MI and stroke events in some of the groups, for further analysis, some groups were collapsed to establish, where possible, groups of comparable size. For the anger-in scale, groups 5 and 6 were combined to form a six point scale. For anger-out groups 1, 2, 3 & 4 were combined to form a binary scale, for anger discuss 0 & 1 were combined to form a four point scale and for anger symptoms 5, 6, 7 & 8 were combined to form a six point scale.

### 6.3.2.1 Anger in

The reference group '0' is compared to the remaining five groups using a logistic model and a Cox's proportional hazard model. The results are given below.

| Table 6.8a logistic regression for MI | | | | | |
|---|---|---|---|---|---|
| | Odds Ratio | z | P>\|z\| | 95% CI | |
| **Anger-in 1** | 0.797 | 0.190 | 0.342 | 0.500 | 1.272 |
| **Anger-in 2** | 0.791 | 0.176 | 0.292 | 0.511 | 1.224 |
| **Anger-in 3** | 0.911 | 0.212 | 0.688 | 0.577 | 1.438 |
| **Anger-in 4** | 0.579 | 0.137 | 0.021 | 0.365 | 0.920 |
| **Anger-in 5** | 0.750 | 0.184 | 0.240 | 0.463 | 1.213 |

| Table 6.8b Cox's Proportional hazard model for MI | | | | | |
|---|---|---|---|---|---|
| | Odds Ratio | z | P>\|z\| | 95% CI | |
| **Anger-in 1** | 0.788 | 0.173 | 0.279 | 0.512 | 1.213 |
| **Anger-in 2** | 0.789 | 0.162 | 0.248 | 0.528 | 1.180 |
| **Anger-in 3** | 0.893 | 0.191 | 0.597 | 0.587 | 1.358 |
| **Anger-in 4** | 0.587 | 0.129 | 0.015 | 0.382 | 0.903 |
| **Anger-in 5** | 0.740 | 0.168 | 0.186 | 0.474 | 1.156 |

When men with anger-in zero are compared with the rest of the groups, it appears that the men with anger-in score of 4 have a lower risk of getting an MI. However, since there was no trend observed across the group for having an MI and the rest of the groups show no significant decreased risk of MI, this is likely to be a chance effect. Therefore, it could be concluded that the anger-in does not appear to be associated with the risk of MI.

For the risk of stroke, a similar analysis was carried out as of MI and the table below shows the odds and hazard ratios when the reference group '0' is compared with the rest of the groups.

| Table 6.9a logistic regression for stroke | | | | | |
|---|---|---|---|---|---|
| | Odds Ratio | z | P>\|z\| | 95% CI | |
| **Anger-in 1** | 0.695 | 0.240 | 0.292 | 0.353 | 1.368 |
| **Anger-in 2** | 0.717 | 0.229 | 0.298 | 0.383 | 1.342 |
| **Anger-in 3** | 1.331 | 0.400 | 0.341 | 0.739 | 2.399 |
| **Anger-in 4** | 0.965 | 0.288 | 0.906 | 0.538 | 1.733 |
| **Anger-in 5** | 1.167 | 0.363 | 0.620 | 0.634 | 2.147 |

| Table 6.9b Cox's Proportional hazard model for stroke | | | | | |
|---|---|---|---|---|---|
| | Hz Ratio | z | P>\|z\| | 95% CI | |
| **Anger-in 1** | 0.686 | 0.230 | 0.261 | 0.356 | 1.323 |
| **Anger-in 2** | 0.709 | 0.219 | 0.266 | 0.387 | 1.300 |
| **Anger-in 3** | 1.291 | 0.371 | 0.374 | 0.735 | 2.267 |
| **Anger-in 4** | 0.934 | 0.268 | 0.812 | 0.532 | 1.640 |
| **Anger-in 5** | 1.120 | 0.334 | 0.704 | 0.624 | 2.010 |

From the above table, it is evident that there is no significant association between the risk of stroke and the anger-in score. This is true for both logistic and Cox's proportional hazard models.

### 6.3.2.2 Anger Out

The distribution of the anger-out score was extremely skew and transformed to a binary scale. As the preliminary analysis showed low anger out score to be associated with higher risk, for this analysis, anger-out scores greater than zero were recoded to zero and scores of zero recoded to one. For this analysis, therefore, a score of 1 indicates low anger-out and 0 indicates high anger-out. The odds ratio and hazard ratio of MI for men with low anger-out scores were OR = 1.75 [95% CI: 1.25 – 2.43] and HR = 1.75 [95% CI: 1.28 – 2.40] respectively relative to men with high anger-out. A low anger-out score does not appear to show any statistically significant increased risk of stroke for angry or annoyed men with OR = 1.34 [95% CI: 0.89 – 2.03] and HR = 1.40 [95% CI: 0.94 – 2.08] respectively.

The Kaplan-Meier survival plots shown in the figure below appears to show that the risk of MI for angry men is consistent over time. Though the results are not statistically significant, the Kaplan-Meier plot for stroke and anger-out suggest there may be a very slight association that is too weak to be detected formally by this study.

141

**Figure 6.8 Kaplan-Meier survival plots comparing men with anger out scores 0 and 1**
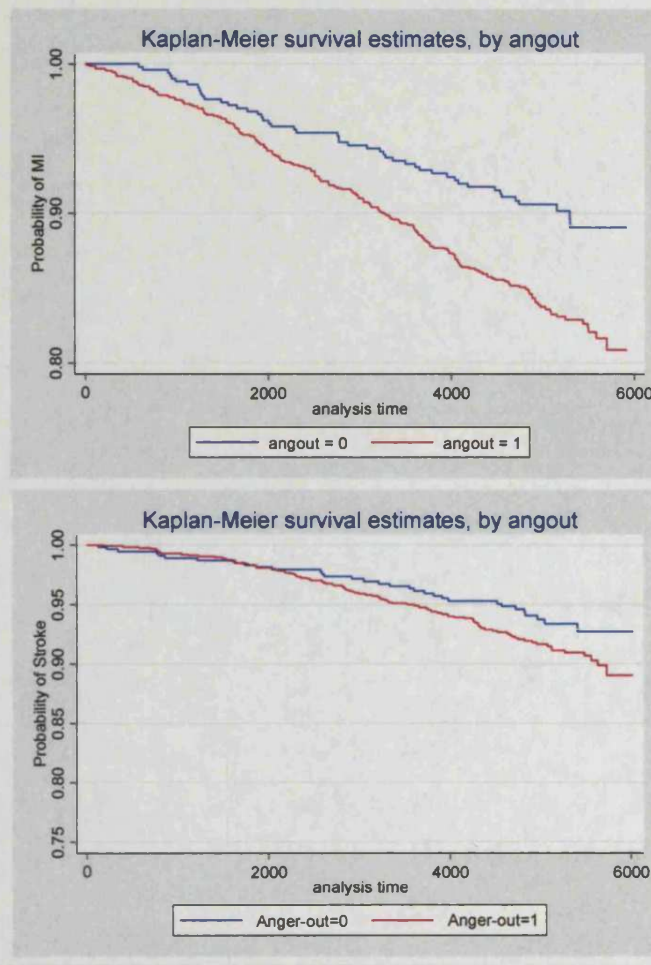
From the above plots, it can be observed that the risk of MI associated with high anger-out group is higher than the risk of MI for low anger-out group. However, the risk of stroke associated with the low and high anger-out score is highly different.

## 6.3.2.3 Anger Discuss

Anger-discuss scores ranged from zero to four. Scores of zero and one were combined to form a larger group. This was made the reference group and compared with the remaining groups for risk of MI. the following table shows the comparisons using logistic and Cox's proportional hazards model. No evidence for an association was found.

| Table 6.7a logistic regression for MI | | | | | |
|---|---|---|---|---|---|
| | Odds Ratio | z | P>\|z\| | 95% CI | |
| **Anger Disc 1** | 1.126 | 0.227 | 0.556 | 0.758 | 1.673 |
| **Anger Disc 2** | 0.823 | 0.187 | 0.389 | 0.527 | 1.283 |
| **Anger Disc 3** | 1.075 | 0.209 | 0.711 | 0.734 | 1.574 |

| Table 6.7b Cox's Proportional hazard model for MI | | | | | |
|---|---|---|---|---|---|
| | Hz Ratio | z | P>\|z\| | 95% CI | |
| **Anger Disc 1** | 1.099 | 0.206 | 0.615 | 0.761 | 1.586 |
| **Anger Disc 2** | 0.813 | 0.173 | 0.330 | 0.536 | 1.233 |
| **Anger Disc 3** | 1.050 | 0.190 | 0.787 | 0.736 | 1.498 |

The analysis was repeated for stroke. Again, no evidence for an association was found.

| Table 6.8a logistic regression for stroke | | | | | |
|---|---|---|---|---|---|
| | Odds Ratio | z | P>\|z\| | 95% CI | |
| **Anger Disc 1** | 1.267 | 0.347 | 0.388 | 0.741 | 2.167 |
| **Anger Disc 2** | 1.183 | 0.345 | 0.565 | 0.668 | 2.094 |
| **Anger Disc 3** | 1.184 | 0.314 | 0.524 | 0.704 | 1.992 |

| Table 6.8b Cox's Proportional hazard model for stroke | | | | | |
|---|---|---|---|---|---|
| | Hz Ratio | z | P>\|z\| | 95% CI | |
| **Anger Disc 1** | 1.251 | 0.330 | 0.397 | 0.745 | 2.099 |
| **Anger Disc 2** | 1.157 | 0.326 | 0.605 | 0.666 | 2.008 |
| **Anger Disc 3** | 1.181 | 0.303 | 0.516 | 0.715 | 1.952 |

## 6.3.2.4 Anger Symptoms

Anger-symptoms ranged from 0 to 8. To achieve more comparable group sizes, the top four groups were combined. Logistic and Cox's proportional hazard models for the risk of MI comparing the lowest anger symptoms group (group'0') with the remainder are given in the table below. There was no evidence of an association.

### Table 6.9a logistic regression for MI

|  | Odds Ratio | z | P>|z| | 95% CI | |
|---|---|---|---|---|---|
| Anger symp 1 | 1.091 | 0.223 | 0.671 | 0.731 | 1.628 |
| Anger symp 2 | 0.729 | 0.155 | 0.136 | 0.481 | 1.105 |
| Anger symp 3 | 0.814 | 0.191 | 0.380 | 0.513 | 1.289 |
| Anger symp 4 | 0.790 | 0.205 | 0.362 | 0.475 | 1.312 |
| Anger symp 5 | 0.695 | 0.187 | 0.177 | 0.410 | 1.178 |

### Table 6.9b Cox's Proportional hazard model for MI

|  | Hz Ratio | z | P>|z| | 95% CI | |
|---|---|---|---|---|---|
| Anger symp 1 | 1.069 | 0.200 | 0.723 | 0.740 | 1.543 |
| Anger symp 2 | 0.744 | 0.147 | 0.134 | 0.505 | 1.096 |
| Anger symp 3 | 0.820 | 0.179 | 0.364 | 0.535 | 1.258 |
| Anger symp 4 | 0.826 | 0.199 | 0.429 | 0.515 | 1.326 |
| Anger symp 5 | 0.701 | 0.177 | 0.159 | 0.427 | 1.150 |

The analysis was repeated for stroke as outcome. There was no convincing evidence of any association. The analysis is presented in the table below.

### Table 6.10a logistic regression for stroke

|  | Odds Ratio | z | P>|z| | 95% CI | |
|---|---|---|---|---|---|
| Anger symp 1 | 1.283 | 0.365 | 0.380 | 0.735 | 2.240 |
| Anger symp 2 | 1.107 | 0.314 | 0.721 | 0.635 | 1.929 |
| Anger symp 3 | 1.025 | 0.330 | 0.940 | 0.545 | 1.928 |
| Anger symp 4 | 1.054 | 0.367 | 0.879 | 0.533 | 2.086 |
| Anger symp 5 | 1.430 | 0.465 | 0.271 | 0.756 | 2.703 |

### Table 6.10b Cox's Proportional hazard model for stroke

|  | Hz Ratio | z | P>|z| | 95% CI | |
|---|---|---|---|---|---|
| Anger symp 1 | 1.233 | 0.338 | 0.445 | 0.721 | 2.108 |
| Anger symp 2 | 1.067 | 0.292 | 0.812 | 0.624 | 1.826 |
| Anger symp 3 | 0.997 | 0.311 | 0.993 | 0.541 | 1.837 |
| Anger symp 4 | 1.044 | 0.351 | 0.897 | 0.540 | 2.019 |
| Anger symp 5 | 1.330 | 0.415 | 0.360 | 0.722 | 2.452 |

## 6.3.2.5 Suppressed anger

Results published on the Caerphilly data identified low anger-out score as a risk factor for CHD [25]. A limitation with this finding is the large proportion of men (70%) with low anger out i.e. who are at greater risk. An attempt was made, therefore, to make the low anger-out score more specific. Individual item analysis on the Caerphilly data showed that out of the 11 Framingham anger score items, only three items were independently associated with incident IHD, namely the two anger-out scale items and one item from the anger-in scale. The wording of these items is given below:

Stem: When really angry or annoyed do you....

| Anger-out item branch: | Take it out on others |
| Anger-out item branch: | Blame someone else |
| Anger-in item branch: | Keep it to yourself |

For these items, low scores for the anger-out items were associated with increased risk whilst a high score for the anger-in item was associated with increased risk. The direction of association of these items suggests a pattern of behaviour where the expression of anger is strongly internalised or, to use other words, suppressed.

For the purpose of further analysis, these items were used as a suppressed anger scale. To be consistent with the scoring used in Gallacher et al [25], the suppressed anger score was calculated using the following formula:

*Suppressed Anger = 2 + first 'Anger-out' + second 'Anger-out' – 'Anger-in' item.*

All these items have three responses (likely, somewhat likely and not too likely). The psychology behind this suppressed anger scale is discussed elsewhere [25]. The distribution of scores for the suppressed anger scale ranged from 0-6 is given in Figure 6.8.

**Figure 6.8: Bar Chart of Suppressed anger score**
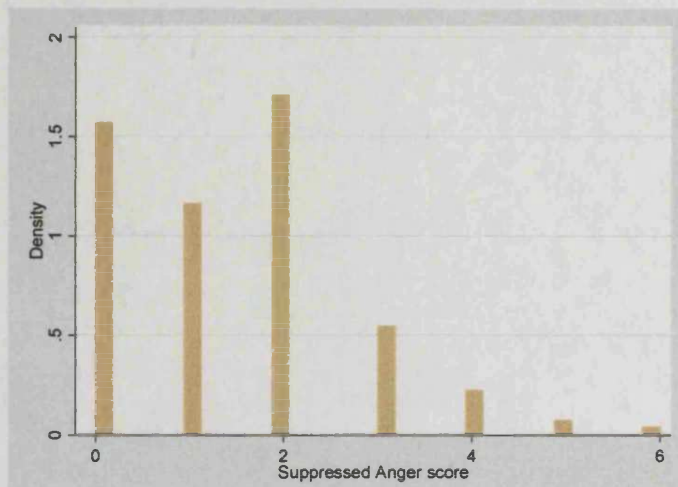


After excluding the missing values, the number of subjects included in the analysis was 1792. The cross tabulation of suppressed anger score and MI is given below.

Table 6.11 – Frequencies of suppressed anger score

| Suppressed Anger score | NO MI | MI | Total |
|---|---|---|---|
| 0 | 450 | 81 | 534 |
| 1 | 347 | 44 | 391 |
| 2 | 488 | 84 | 572 |
| 3 | 161 | 23 | 184 |
| 4 | 71 | 4 | 75 |
| 5 | 22 | 1 | 23 |
| 6 | 10 | 3 | 13 |
| Total | 1549 | 243 | 1792 |

Those who scored zero represent the group with most suppressed anger. This represents 30% of the sample. On this basis the suppressed anger score provides a more widely distributed range than the anger-out scale and allows a smaller putatively high risk group to be identified than did the anger-out scale.

For the present analysis, as the low scoring group may be considered to be at highest risk, the findings will be more easily interpreted if the scale is reversed. Men with a suppressed anger score zero were recoded to '1', whilst the remainder were recoded to '0'.

When men with suppressed anger score one are compared with men with suppressed anger score zero for having an MI, the odds and the hazard ratios are OR = 1.29 [95% CI: 0.97 – 1.72] and HR = 1.27 [95% CI: 0.97 – 1.65] whereas for stroke, OR = 1.13 [95% CI: 0.78 – 1.65] and HR = 1.13 [95% CI: 079 – 1.63]. For MI these results show that the increased sensitivity provided by the suppressed anger score is balanced by a reducing strength of association. For stroke these results confirm the null findings of the anger-out scale.

### 6.3.3 Discussion

Of the four Framingham anger scales and the derived suppressed anger scale used in this analysis, the anger-out scale shows strongest evidence of an association with MI. None of the scales show evidence of an association with stroke.

These data are interesting as they extend the analysis of the anger scales from IHD to MI and stroke and they adjust for prevalent heart disease by omitting men with previous disease rather than using a statistical control. That low scoring on the anger-out scale was found to be related to MI is confirms the previous findings, although a weaker association was found for suppressed anger [25].

That low scoring on the anger-out scale was related to MI whereas high scoring on the anger-in scale was not is surprising. As the Framingham anger scales have not been psychometrically developed to any large extent, their items have to be interpreted according to their content. High risk seems to be associated with responses identifying men who are not too likely to "take it out on others" and not too likely to "blame someone else" but very likely to "keep it to themselves", suggesting a group who are more likely to work their anger out on themselves. If these responses are taken at face value, it may be that they indicate a more toxic psychological reaction than just not expressing anger.

# 6.4 Type A behaviour

## 6.4.1 Literature Review

Type A behaviour (TAB) is a term used to characterise people who are driven, hard-working, busy and impatient. It was described as a risk factor in cardiovascular disease in the 1950s by cardiologist Meyer Friedman. It was estimated that Type A behaviour doubles the risk of CHD in subjects with Type A Behaviour (TAB) when compared with subjects without Type A behaviour [172]. TAB can be measured with different questionnaires such as Jenkins Activity Survey (JAS), Bortner, and Framingham type A; the JAS is the most widely used. In the Caerphilly cohort, all these three questionnaires were used to measure TAB.

Earlier studies conducted by Jenkins and Rosenman worked on the Western collaborative group study in US which showed a positive association between TAB and coronary heart disease [22, 172]. Haynes worked on Framingham heart study, showing similar results [173]. There were 2750, 3154 and 954 male subjects in these studies with a follow-up of 4, 8.5 and 8 years respectively. They showed that the risk of MI nearly doubled when men with and without TAB were compared. Several studies conducted after these three studies showed no association between TAB and CHD in men [174, 175].

## 6.4.2 The Caerphilly analysis

The JAS has been used more widely than the other two scales and is used here for the analysis. The distribution of JAS is given in the Figure 6.9.

**Figure 6.9: Histogram of JAS**
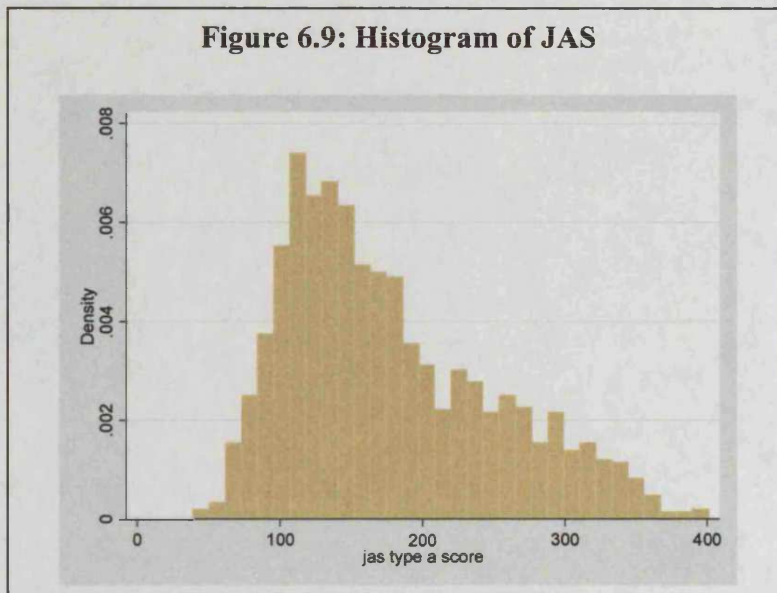


After excluding the missing values, the number of subjects used for the analysis is 1839. The summary statistics of JAS is given in the Table 6.12.

**Table 6.12 – Summary statistics of JAS score**

| Variables | Obs | Mean | Std. Dev | Min | Max |
|-----------|-----|------|----------|-----|-----|
| JAS | 1839 | 1738 | 72.83 | 39 | 401 |

To identify the appropriate modelling strategy for modelling the relationship between JAS score and MI or stroke, a Lowess smoothed plot was used. There were 17 subjects having JAS less than 65, and six with JAS>365 causing the relationship between MI, stroke and JAS to appear non-linear. Various modelling strategies such as linear and fractional polynomial models were applied before exclusion and found no significant differene between fractional polynomial and linear models. Therefore, those 23 subjects were excluded and the following Figure 6.10 shows these different fits along with the Lowess plot.

**Figure 6.10 Relationship between JAS score and MI, Stroke**
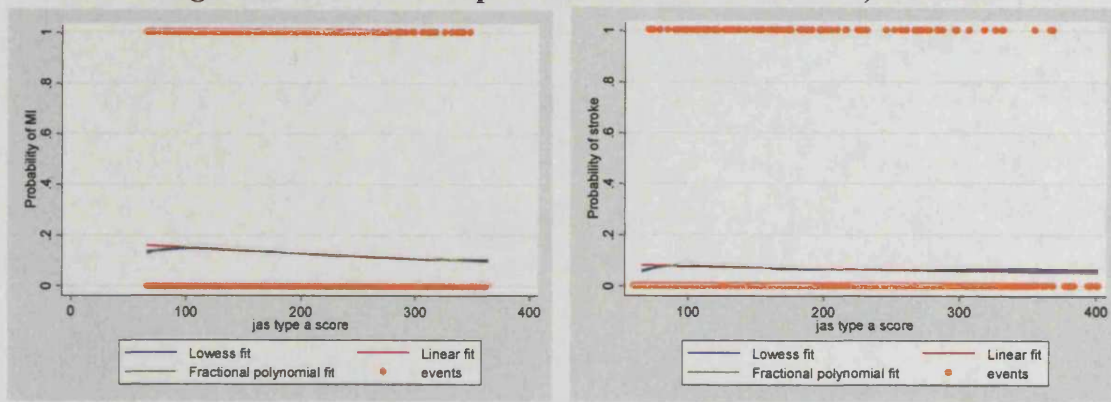
The smoothing parameter used for the Lowess fit is 0.8. The linear model appears to be a very reasonable fit for both MI and stroke with JAS. However, the results show that JAS is not associated with an increased risk of either MI or stroke.

The Table below shows the odds ratios and hazard ratios for having an MI or stroke with the JAS score.

**Table 6.13 Odds and Hazard ratios of MI and stroke with JAS**

|  | ODDS RATIO (95% CI) | HAZARD RATIO (95% CI) |
|---|---|---|
| MI | 0.998 [95% CI: 0.996 – 1.000] | 0.998 [95% CI: 0.997 – 1.000] |
| STROKE | 0.999 [95% CI: 0.996 – 1.001] | 0.999 [95% CI: 0.996 – 1.001] |

the confidence intervals are very small for these results because there range of the JAS data is quite wide ranging from 65 to 365 i.e. the unit change in Type A score is small. Therefore, instead of a unit change, a change in 20 units was estimated. But the results showed no difference (odds of MI 0.98 95% CI: 0.94 – 1.02). There is no increased risk of MI or stroke as JAS score increases.

150

### 6.4.3 Discussion

Some early studies from the literature review have shown that Type A behaviour is associated with heart disease risk whereas this was not found in the Caerphilly study. TAB was also not found to be associated with stroke in Caerphilly.

In most long term studies of TAB with heart disease, the association has grown weaker over time [176, 177]. This may be due to a reduction in TAB with age or reduced validity of the TAB score as time increases, or even a change in behaviour after a first coronary event. In the present analysis this last possibility was taken into account by omitting men with previous disease and analysing for the first MI only.

Previous analysis of the Caerphilly TAB data has shown that Caerphilly also follows the pattern of a stronger association at shorter follow-up periods. [34] However, this analysis suggested that Type A acts as a trigger for IHD in men with already diseased arteries rather than as a cause of atherosclerosis. Although this specific hypothesis was not relevant to this thesis, the analyses conducted here are consistent with those published previously [34].

## 6.5 Health attitudes

### 6.5.1 Literature review for Health attitudes

Heart disease is a condition with a multi-factorial aetiology of both behavioural and biological risk factors. Psychological theory proposes that the determinants of behaviour include attitudes [46]. Attitudes are complex constructs but have emotional, cognitive and behavioural components. As such they are useful for measuring a person's summary assessment. For example, a person's attitude to smoking represents their assessment of the pleasure and health risk that they believe smoking offers. If they believed the pleasure outweighs the risk, their attitude would be positive. The limitations in the attitude construct are that it is so complex and that there are many different theoretical approaches.

In the Caerphilly heart disease, attitudes towards seven coronary related behaviours were measured using the Health Attitude Inventory (HAI) The description, development and application of HAI is discussed in detail elsewhere [47]. The development of the HAI was based on Fishbein and Ajzen's Theory of Reasoned Action (TRA). This theory was chosen as it is relatively straightforward to operationalise.

The TRA proposes that behaviour is influenced by our intention to behave in a certain manner. Intention does not always predict behaviour for many reasons, such as changes in, or unforeseen, circumstances. However, as a general principle, the TRA proposes that intention is a strong predictor of behaviour.

In TRA, intention is determined by attitudes and subjective norms. Attitudes are the combined influence of the person's own beliefs and values about the outcomes associated with the behaviour. Subjective norms are "the person's perception that most people who are important to him or her think he should or should not perform the behavior in question" (Azjen and Fishbein, 1975 [48]). The subjective norm has two constituents; normative beliefs about what others think should be done, and the

152

motivation to comply with those normative beliefs. The relationship of attitude and subjective norms to intention and behaviour is given below:

$$Attitude + Subjective\ Norm = Intention \rightarrow Behaviour$$

Classically, the TRA is described using the formula below where the attitude and subjective norm components are described using their constituent variables:

$$\sum_{i=1}^{i=n} (b_i \times v_i)_{w1} + \sum_{i=1}^{i=n} (nb_i \times mc_i)_{w2} = Intention \rightarrow Behaviour$$

Where:

$b_i$ = beliefs about outcome

$v_i$ = values regarding outcome

$nb_i$ = normative beliefs about what other people want you to do

$mc_i$ = motivation to comply with other peoples wishes.

w1 = regression weight for attitude

w2 = regression weight for subjective norm

In the above equation beliefs ($b_i$) and values ($v_i$) make up the attitude component of the theory whilst normative beliefs ($nb_i$) and motivation to comply ($mc_i$) make up the subjective norm component of the theory

Although this theory is very clear by definition, the theory may be criticised as not thorough because the group of beliefs and values that are attributed to an outcome are unlikely to be as simple as suggested by the theory. Also several strong mathematical assumptions are involved in specifying the relations between variables.

It is unlikely, for example, that beliefs and values really are combined multiplicatively by people in everyday life. It is also unclear whether the use of regression weights

adds to the explanatory power of the theory or obscures a lack of explanatory power. A further criticism is the lack of definition about how the constituent variables should be measured. For example, if beliefs are considered to be some form of probability judgment and measured on uni-polar scales of increasing certainty, should values be measured using bipolar scales, allowing both positive and negative valuation? If these variables are to be combined multiplicatively then different scoring systems will radically affect the distribution of attitude scores. In spite of these criticisms the theory is widely used by psychologists.

Aspects of coronary related behaviour were measured in the HAI were identified by interview of a random sample of 100 men and women selected from the electoral register. The seven coronary risk behaviours investigated were exercise, dairy produce consumption, fried food consumption, smoking, vegetable consumption, whole meal bread consumption and stress. From these interviews between four and six beliefs were identified for each coronary risk behaviour.

Scoring of the HAI has been designed to reflect the meaning behind the judgement made by the subjects and to be relatively simple for the researcher to calculate the score. Scoring is explained briefly here since it is described in detail elsewhere [47].

Beliefs were measured using four point unipolar Likert scales. For example to the question "Is smoking expensive?" the response categories and coding were:

| Definitely no | Probably no | Probably yes | Definitely yes |
|---------------|-------------|--------------|----------------|
| 0 | 1 | 2 | 3 |

Where appropriate, value judgements were assessed using five point bi-polar scales. For example, to the question how much do you like or dislike spending money on smoking the response categories and coding were:

| Dislike Very much | Dislike | Not sure | Like | Like Very much |
|-------------------|---------|----------|------|----------------|
| -2 | -1 | 0 | 1 | 2 |

For several value items unipolar scales were appropriate. For example to the question how much do you fear increasing your risk of a heart attack through smoking the response categories and coding were:

| Not at All | A Little | Moderately | Very much |
|------------|----------|------------|-----------|
| 0 | -1 | -2 | -3 |

There was a single normative belief item for each coronary risk behaviour. This was assessed with regard to wife or closest friend and was scaled and coded in the same way as the belief items. There was a single overall motivation to comply item. This was scaled and scored as follows:

How much do you generally like to do what you wife or closest friend says?

| Couldn't Care less | Not much | Not sure | Like to | Like to very much |
|--------------------|----------|----------|---------|-------------------|
| -2 | -1 | 0 | 1 | 2 |

Altogether the HAI comprised 76 items and a copy is given in the appendix 3.

The raw questionnaire data were processed to provide attitude and social norm data by multiplying the respective variables (beliefs x values and normative beliefs x motivation to comply) for each health risk behaviour and summing the attitudes and social norms for each behaviour. The summed attitude variable and summed social norm variable were then used as exposure variables in the analysis. The distribution of the summed attitudes is given for each coronary risk behaviour below:

Figure 6.11 – Histograms of attitudes towards behaviours in 2<sup>nd</sup> phase
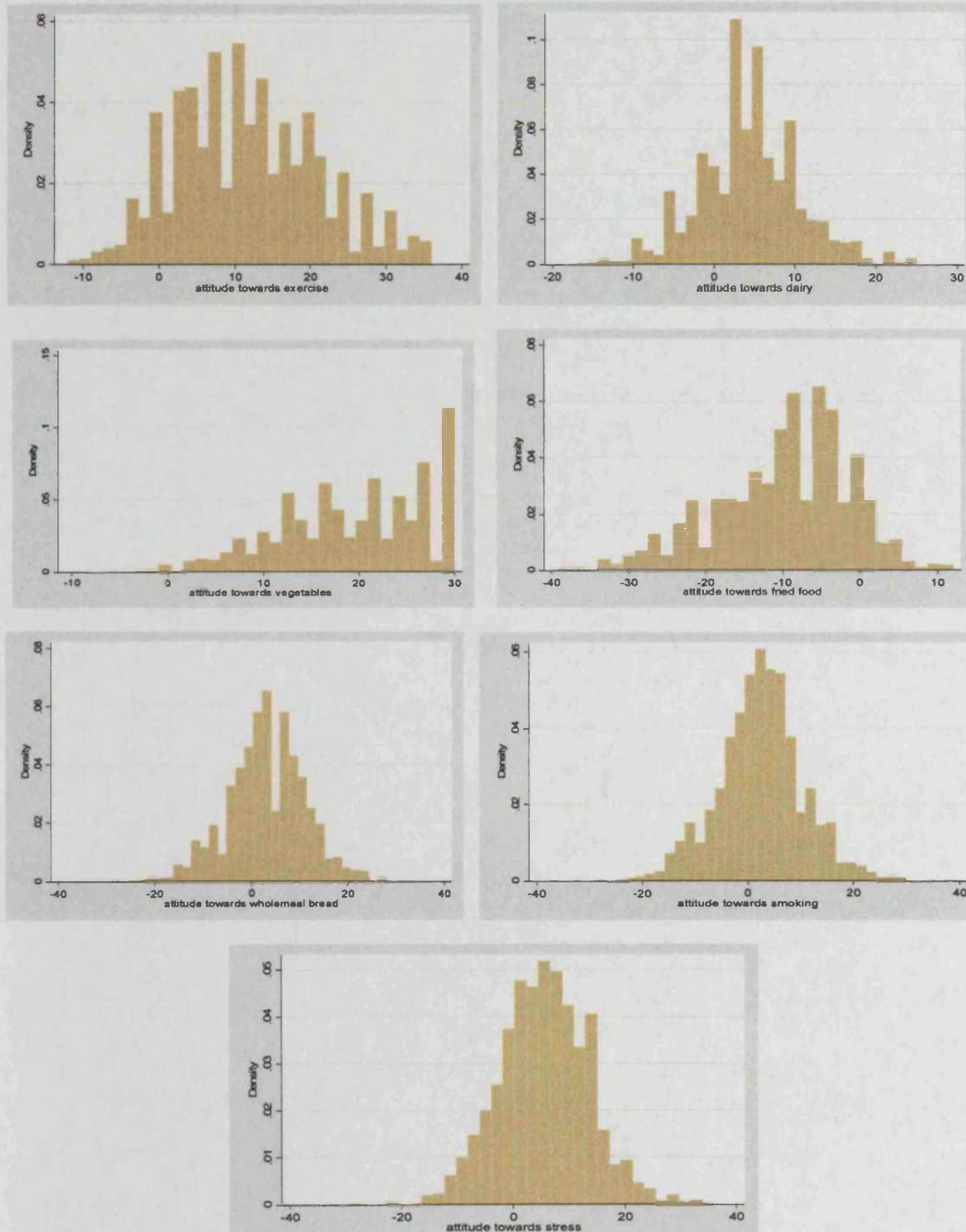
These distributions are helpful for interpreting the attitude data. A high score reflects a positive attitude towards the behaviour and a low score reflects a negative attitude towards the behaviour. The distributions for exercise, dairy produce consumption, vegetable consumption, wholemeal bread consumption and stress were strongly positive i.e. >0. For smoking the summed attitude was more ambivalent and for fried food it was strongly negative. Although the vegetable consumption variable was

strongly negatively skewed, transforming the score would have made interpreting subsequent analyses extremely difficult. The scale was not transformed. The summary statistics for the summed attitudes are given below.

**Table 6.14 – Summary statistics of attitude towards behaviours**

| Attitude | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Exercise | 1482 | 11.68 | 9.35 | -12 | 36 |
| Dairy Produce | 1482 | 4.30 | 6.20 | -17 | 25 |
| Vegetables | 1482 | 19.43 | 7.70 | -10 | 30 |
| Fried Food | 1482 | -9.82 | 8.53 | -39 | 12 |
| Wholemeal Bread | 1482 | 3.10 | 7.92 | -29 | 28 |
| Smoking | 1482 | 2.47 | 8.02 | -29 | 30 |
| Stress | 1482 | 5.53 | 8.02 | -29 | 36 |

The distributions of the summed subjective norms were also investigated. The summary statistics are given below and this is followed by the distribution plots.

**Table 6.15 – Summary statistics of subjective norms in $2^{nd}$ phase**

| Attitude | Obs | Median | Min | Max |
|---|---|---|---|---|
| Exercise | 1482 | 6 | 0 | 12 |
| Dairy Produce | 1482 | 4 | 0 | 12 |
| Vegetables | 1482 | 6 | 0 | 12 |
| Fried Food | 1482 | 4 | 0 | 12 |
| Wholemeal Bread | 1482 | 6 | 0 | 12 |
| Smoking | 1482 | 6 | 0 | 12 |
| Stress | 1482 | 3 | 0 | 12 |

The higher the subjective norm score the greater the influence of the wife or best friend on behaviour. It is clear that in terms of self report, most men considered themselves to be relatively little influenced socially.

**Figure 6.12 – Histograms of subjective norms towards behaviours in 2nd phase**
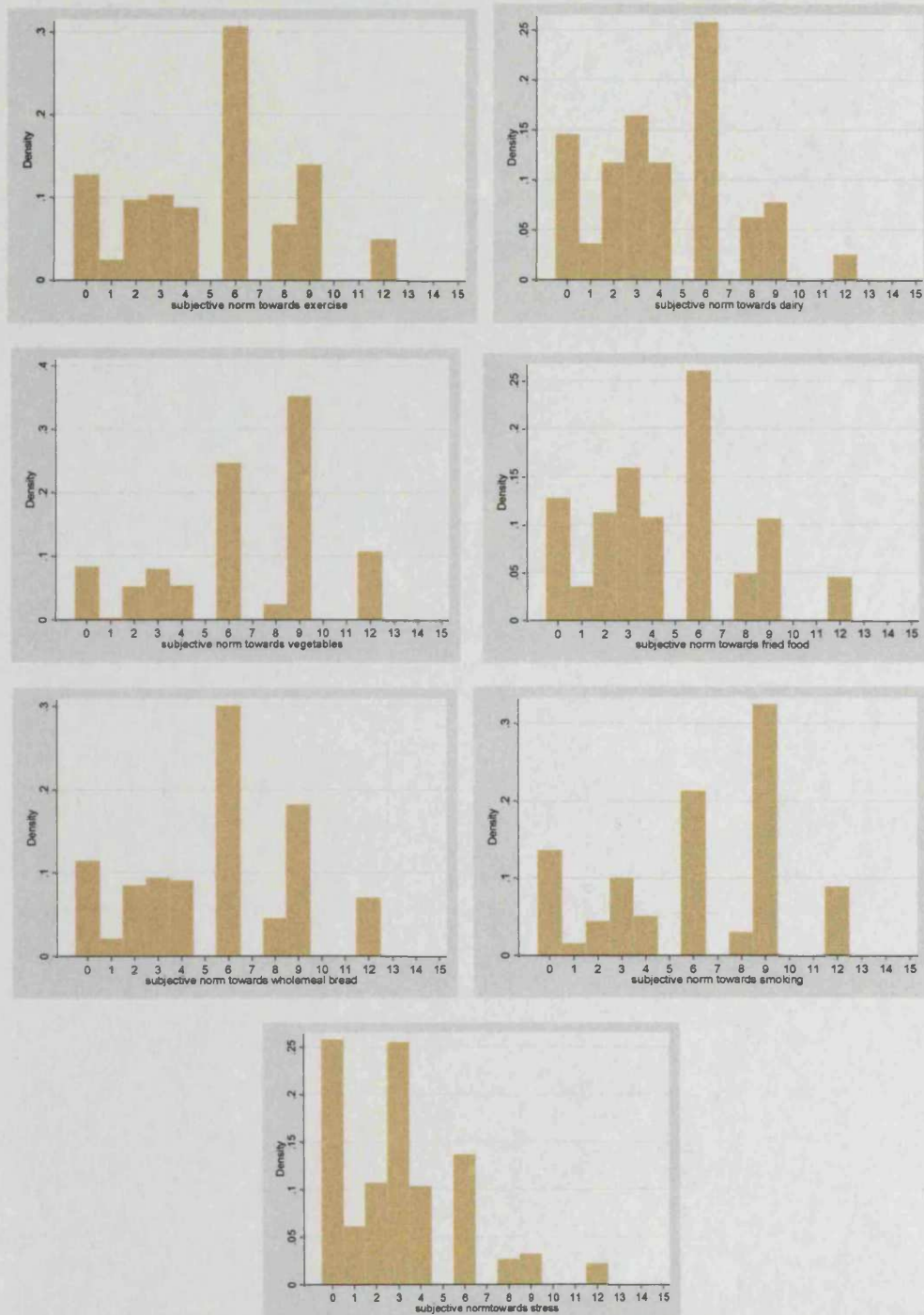


## 6.5.2 Attitudes towards coronary related behaviours and MI

TRA proposes that attitudes predict behaviour. This theory has been applied to the Caerphilly cohort and found some support [47]. The purpose of the present analysis is to show whether attitudes can also be used to predict vascular outcome 17 years later. TRA proposes a linear model and this was assumed for this analysis. Preliminary

158

analysis using Lowess plots was appropriate for the summed attitudes and showed that a linear model was reasonable although not necessarily optimal. The Lowess plots for summed attitudes are shown below.

**Figure 6.12 – Lowess Plots for the relationships between MI and attitudes towards coronary risk behaviours**
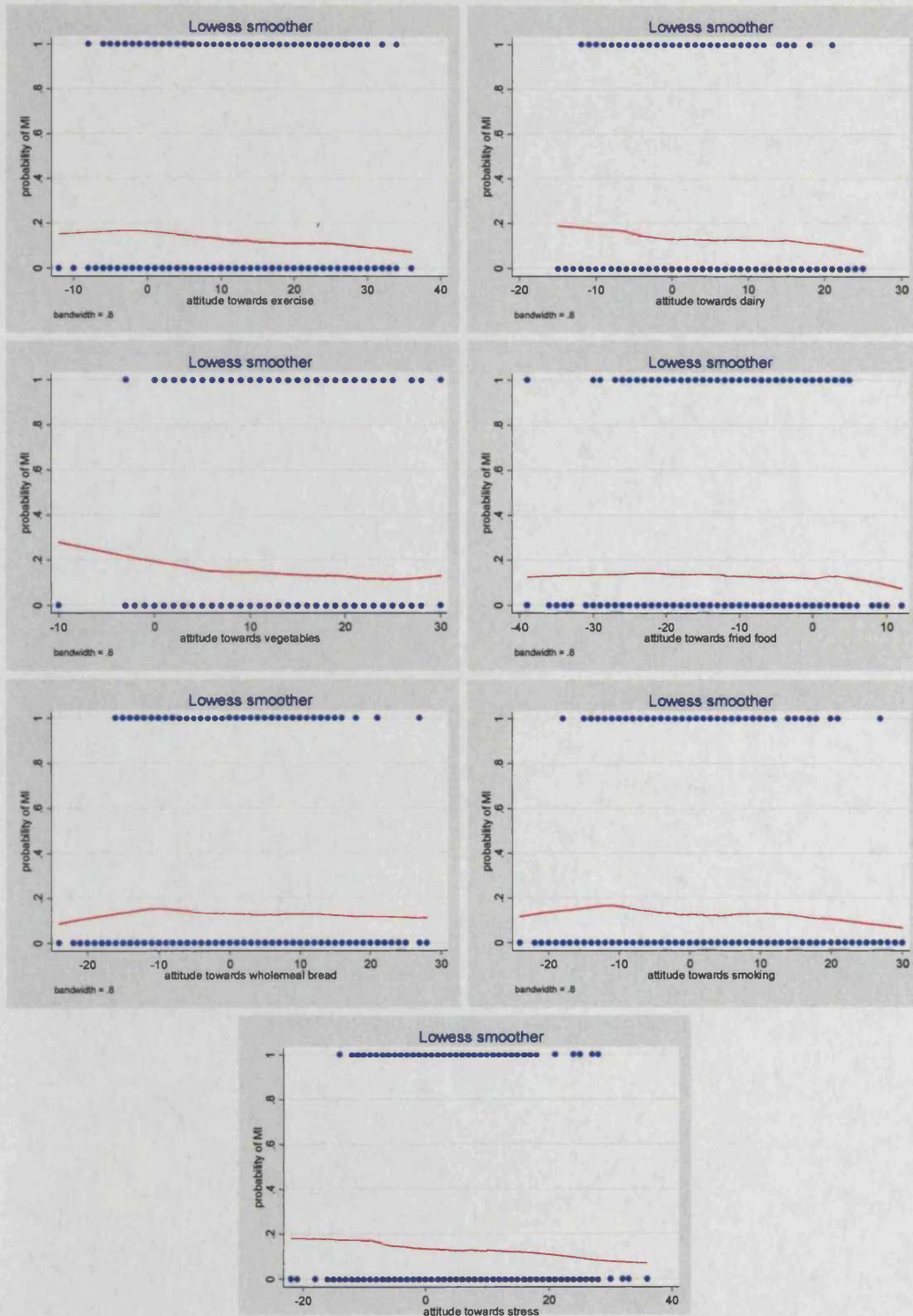
Lowess plots were not appropriate for the subjective norms due to the discrete characteristics of the scales. The full TRA model was tested using a linear combination of the attitude and subjective norm effects.

Logistic regression of MI on summed attitudes and summed subjective norms showed several weak associations. If the associations with attitudes are considered first, a positive attitude towards exercise was associated with a lower risk of MI. Although no other attitudes were related to MI at the formal $p < 0.05$ level, all the odds ratios were less than 1 suggesting that a larger study may have detected weak relationships between MI and some of the other attitudes. Regarding the subjective norms, peer pressure to eat whole meal bread was associated with a reduced risk of MI. No other subjective norms were related to MI at a formal level although, as with attitudes, all the odds ratios were less than 1. If the full TRA model (with each attitude score and subjective norm corresponding to that attitude in the same model), is tested by a linear combination of attitude and subjective norm components, a positive association was found for vegetable consumption and fried food and wholemeal bread.

**Table 6.16 – Logistic regression of Attitudes and subjective norms for MI**

| Coronary risk behaviour | Attitudes and social norms | Odds ratio for a unit change (p, 95% CI) | Linear combination (p, 95% CI) |
|---|---|---|---|
| Exercise | Attitude | 0.98(0.04, 0.96 – 0.99) | 0.96(0.09, 0.91, 1.01) |
| | Subjective Norm | 0.98(0.38, 0.93 – 1.03) | |
| Dairy produce | Attitude | 0.99(0.63, 0.97 – 1.02) | 0.96(0.13, 0.91 – 1.01) |
| | Subjective Norm | 0.97(0.23, 0.91 – 1.02) | |
| Vegetables | Attitude | 0.99(0.28, 0.97 – 1.01) | 0.95(0.04, 0.91 – 0.99) |
| | Subjective Norm | 0.96(0.12, 0.92 – 1.01) | |
| Fried food | Attitude | 0.99(0.13, 0.97 – 1.00) | 0.94(0.05, 0.88 – 1.00) |
| | Subjective Norm | 0.95(0.09, 0.91 – 1.01) | |
| Wholemeal bread | Attitude | 0.99(0.32, 0.97 – 1.01) | 0.93(0.02, 0.88 – 0.99) |
| | Subjective Norm | 0.94(0.02, 0.90 – 0.99) | |
| Smoking | Attitude | 0.99(0.24, 0.97 – 1.01) | 0.96(0.12, 0.91 – 1.01) |
| | Subjective Norm | 0.97(0.17, 0.93 – 1.01) | |
| Stress | Attitude | 0.99(0.21, 0.97 – 1.01) | 0.96(0.24, 0.90 – 1.03) |
| | Subjective Norm | 0.97(0.36, 0.92 – 1.03) | |

The analysis was repeated using Cox regression. This analysis confirmed the findings of the logistic regression and appeared to be show similar results.

**Table 6.17 – Cox's Proportional hazard model of Attitudes and subjective norms for MI**

| Coronary risk behaviour | Attitudes and social norms | Hazard ratio for a unit change (p, 95% CI) | Linear combination (p, 95% CI) |
|---|---|---|---|
| Exercise | Attitude | 0.98 (0.02, 0.96 – 0.99) | 0.96(0.07, 0.91, 1.01) |
| | Subjective Norm | 0.98 (0.35, 0.93 – 1.03) | |
| Dairy produce | Attitude | 0.99 (0.74, 0.97 – 1.02) | 0.96(0.13, 0.91 – 1.01) |
| | Subjective Norm | 0.97 (0.19, 0.92 – 1.02) | |
| Vegetables | Attitude | 0.99 (0.21, 0.97 – 1.01) | 0.95(0.03, 0.92 – 0.99) |
| | Subjective Norm | 0.96 (0.09, 0.93 – 1.01) | |
| Fried food | Attitude | 0.99 (0.16, 0.97 – 1.00) | 0.94(0.05, 0.88 – 1.00) |
| | Subjective Norm | 0.96 (0.07, 0.91 – 1.00) | |
| Wholemeal bread | Attitude | 0.99 (0.37, 0.97 – 1.01) | 0.93(0.02, 0.88 – 0.99) |
| | Subjective Norm | 0.94 (0.01, 0.90 – 0.99) | |
| Smoking | Attitude | 0.99 (0.29, 0.97 – 1.01) | 0.96(0.10, 0.91 – 1.01) |
| | Subjective Norm | 0.97 (0.12, 0.93 – 1.01) | |
| Stress | Attitude | 0.99 (0.23, 0.97 – 1.01) | 0.96(0.28, 0.89 – 1.03) |
| | Subjective Norm | 0.98 (0.40, 0.92 – 1.03) | |

If the analysis is repeated looking at individual attitudes and subjective norms, rather than the summed components, i.e., each individual item in the questionnaire, MI associations were only found for exercise and smoking related variables. The associations of MI and individual exercise attitudes were not found to be independent of each other.

**Table 6.18 – Logistic regression of individual attitudes and MI**

| Coronary risk behaviour | Attitudes | Odds ratio for a unit change (p, 95% CI) |
|---|---|---|
| Exercise | Improve health and fitness | 0.95 (0.04 0.90 – 1.00) |
| | Enjoy exercise | 0.92 (0.01, 0.86 – 0.98) |
| | Reduce risk of a heart attack | 0.94 (0.04, 0.89 – 1.00) |
| Smoking | Unpleasant to others | 1.09 (0.01, 1.02 – 1.15) |

### 6.5.3 Attitudes towards coronary related behaviours and stroke

The analysis was then repeated for stroke. No evidence of any association was found using either logistic or Cox regression. All confidence intervals contain one and all p-values well in excess of 0.05 suggesting no association whatsoever between attitude, subjective norms and stroke.

**Table 6.19 – Logistic regression of Attitudes and subjective norms for stroke**

| Coronary risk behaviour | Attitudes and social norms | Odds ratio for a unit change (p, 95% CI) | Linear combination (p, 95% CI) |
|---|---|---|---|
| Exercise | Attitude | 0.99 (0.28, 0.96 – 1.01) | 1.00(0.89, 0.94, 1.07) |
| | Subjective Norm | 1.01 (0.62, 0.95 – 1.09) | |
| Dairy produce | Attitude | 1.00 (0.92, 0.96 – 1.03) | 1.02(0.62, 0.95 – 1.10) |
| | Subjective Norm | 1.02 (0.60, 0.95 – 1.09) | |
| Vegetables | Attitude | 0.99 (0.67, 0.97 – 1.02) | 0.99(0.87, 0.94 – 1.06) |
| | Subjective Norm | 1.00 (0.98, 0.94 – 1.06) | |
| Fried food | Attitude | 1.00 (0.78, 0.97 – 1.02) | 1.06(0.13, 0.98 – 1.15) |
| | Subjective Norm | 1.07 (0.06, 1.00 – 1.14) | |
| Wholemeal bread | Attitude | 1.00 (0.94, 0.98 – 1.03) | 1.04(0.27, 0.97 – 1.12) |
| | Subjective Norm | 1.04 (0.20, 0.98 – 1.11) | |
| Smoking | Attitude | 1.00 (0.80, 0.98 – 1.03) | 1.00(0.87, 0.94 – 1.08) |
| | Subjective Norm | 1.00 (0.94, 0.95 – 1.06) | |
| Stress | Attitude | 1.00 (0.78, 0.98 – 1.03) | 0.99(0.79, 0.90 – 1.08) |
| | Subjective Norm | 0.98 (0.79, 0.91 – 1.06) | |

**Table 6.20 – Cox's Proportional hazard model of Attitudes and subjective norms for Stroke**

| Coronary risk behaviour | Attitudes and social norms | Hazard ratio for a unit change (p, 95% CI) | Linear combination (p, 95% CI) |
|---|---|---|---|
| Exercise | Attitude | 0.99(0.24, 0.96 – 1.01) | 1.00 (0.99, 0.94, 1.07) |
| | Subjective Norm | 1.01(0.68, 0.95 – 1.08) | |
| Dairy produce | Attitude | 1.00(0.98, 0.97 – 1.03) | 1.02 (0.68, 0.94 – 1.10) |
| | Subjective Norm | 1.01(0.69, 0.95 – 1.09) | |
| Vegetables | Attitude | 0.99(0.62, 0.97 – 1.02) | 0.99 (0.76, 0.95 – 1.06) |
| | Subjective Norm | 1.00(0.93, 0.94 – 1.06) | |
| Fried food | Attitude | 1.00(0.80, 0.97 – 1.02) | 1.06 (0.16, 0.98 – 1.16) |
| | Subjective Norm | 1.06(0.08, 0.99 – 1.13) | |
| Wholemeal bread | Attitude | 1.00(0.92, 0.98 – 1.03) | 1.04 (0.33, 0.96 – 1.13) |
| | Subjective Norm | 1.04(0.26, 0.97 – 1.10) | |
| Smoking | Attitude | 1.00(0.79, 0.98 – 1.03) | 1.00 (0.98, 0.91 – 1.10) |
| | Subjective Norm | 1.00(0.92, 0.95 – 1.05) | |
| Stress | Attitude | 1.00(0.78, 0.98 – 1.03) | 0.99 (0.79, 0.91 – 1.10) |
| | Subjective Norm | 0.98(0.69, 0.91 – 1.06) | |

### 6.5.4 Discussion

There are no previous studies looking at the long term prediction of MI or stroke by attitudes towards coronary risk behaviours. This analysis has investigated the TRA at different levels of detail and found only very slight associations between different components of the theory and MI, but no associations for stroke.

Evidence on the TRA itself was mixed. Some of the attitudes were weakly associated with MI individually suggesting that the theory is over rated. However, that there were no independently predictive attitudes to compare these results.

Although the logistic and survival analyses suggested that attitudes towards vegetable consumption, fried food and wholemeal bread may be associated with MI, the associations were small and may be due to the number of significance tests conducted. If the associations were to be real, they are unlikely to be of public health significance due to their slight magnitude.

## 6.6 Overall discussion for the chapter

Psychological factors tend to have an influence on lifestyle factors that in turn show high association with the risk of MI or stroke. Factors such as anger out and suppressed anger to some extent showed statistically significant results in measuring the association of risk of MI but not stroke whereas other psychological risk factors did not show any significant increase in the risk of either MI or stroke.

Adjusting for measurement error and random variations appear to increase the baseline association of psychological depression indicator with MI. This gives a crude idea that the psychological factors will also have some variations and error while measuring and so, it is important to collect them repeatedly.

# Chapter 7

## 7.0 Multivariate analysis

Several risk factors for MI and stroke were discussed in the previous chapters. Some of them were significant predictors of the outcome and some not. However, the overall risk of MI or stroke cannot be determined by any single risk factor alone. The overall risk must be estimated from the combined effects of the risk factors.

Many risk factors for vascular disease are inter-related in that someone with high cholesterol may also be more likely to have high blood pressure. When estimating the overall risk it is the combined effects of the risk factors which is of importance. The combined effect of risk factors is not easy to estimate for two reasons. First, risk factors may be related i.e. someone with high blood pressure may also have high cholesterol. In this instance, the overall risk is be based on the sum of the independent effects of the risk factors, that is to say the addition of their individual effects after adjusting for the effects of all the other risk factors. In this way, multivariate approaches can give more insight into the relationships between the variables and an outcome, rather than when considering individual risk factors in isolation.

Second, risk factors may also act together so that their combined effect is greater than the sum of their individual effects. For example, a man with high levels of anxiety and depression may have a greater risk than would be expected from the sum of the risks calculated in men with anxiety and men with depression. A further difficulty with a multivariable approach to estimating risk is that adjustment cannot be made for risk factors that have not been included in the study. For these reasons, multivariate analysis is a complex process and the results should be interpreted cautiously.

## 7.1 Literature review

Multivariate analysis is a standard procedure in vascular disease epidemiology. It is used in virtually all major studies. For each risk factor, some major studies were reviewed in chapters 4-6. The results that were reviewed were largely based on multi-variable analysis in which the appropriate adjustments were made. This literature review is not repeated here, therefore.

## 7.2 The Caerphilly analysis plan

In this chapter, the overall risk for MI or stroke is modelled using a multivariate risk model comprising biological and behavioural risk factors followed up by another model which also includes psychological factors. Since the effect of standard risk biological and behavioural risk factors is well established, these results will be compared with the Caerphilly analysis briefly. The results of the analysis that also includes psychosocial factors, will be discussed in more detail.

Linear logistic regression and Cox's proportional hazard models were used to determine the odds and hazard ratios of the risk factors in the multivariate models for MI and stroke. Multivariate methods have an advantage of bringing in more information to bear on a specific outcome (MI and stroke for this analysis). They allow one to take into account the continuing relationships among several variables.

Subjects with extreme values for any of the risk factors that were excluded in the individual models of each risk factor are also excluded in this analysis to make the analyses consistent. For the models that did not include attitudes towards behaviours, when all the missing values were excluded, 1716 men were included in the analysis. When the attitudes towards behaviours were also analysed, 1190 men were included in the analysis.

A few variables were not discussed in previous chapters, but have been used in this chapter for adjustment purposes. These variables were not discussed previously as they were largely non-modifiable (apart from diabetes) and so have been included for adjustment purposes only. The following section describes these factors and their univariate relationships with MI and stroke.

## 7.3 Non modifiable risk factors

### 7.3.1 Age

The odds of MI for increase in one year of age is 1.04 ($p$=0.02, 95% CI: 1.01 – 1.07) i.e., the odds of MI increases by 19% for every 5 years increase in age. If the Cox's proportional hazards are modelled, the hazard ratios was 1.23 for every 5 years. This means that age is a very highly associated with MI but unfortunately it is a non-modifiable factor.

The odds of stroke for an increase in one year of age is 1.09 (p<0.005, 95% CI: 1.05 – 1.14) suggesting that age is also a very highly associated with the increased risk of stroke. This means that the odds of stoke increases by 55% for every 5 years increase in age. With the Cox's proportional hazard model, the hazards ratio for an increase of 5 years in age is 1.63.

### 7.3.2 Marital Status (MS)

In CaPS, marital status was recorded using five different groups. These were married, single, widowed, divorced, and separated. Married men were considered as a reference group and compared to the remainder. The logistic and Cox's proportional hazard models are illustrated in the following table.

| Table 7.1a logistic regression for MI | | | | | |
|---|---|---|---|---|---|
| **MI** | **Odds Ratio** | **z** | **P>\|z\|** | **95% CI** | |
| Single | 1.234 | 0.370 | 0.484 | 0.685 | 2.222 |
| Widowed | 1.921 | 0.580 | 0.031 | 1.063 | 3.472 |
| Divorced | 1.555 | 0.533 | 0.197 | 0.795 | 3.045 |
| Separated | 1.357 | 1.491 | 0.781 | 0.158 | 11.678 |

| Table 7.1b Cox's Proportional hazard model for MI | | | | | |
|---|---|---|---|---|---|
| **MI** | **Hz Ratio** | **z** | **P>\|z\|** | **95% CI** | |
| Single | 1.276 | 0.353 | 0.379 | 0.742 | 2.193 |
| Widowed | 2.049 | 0.549 | 0.007 | 1.212 | 3.463 |
| Divorced | 1.575 | 0.488 | 0.142 | 0.858 | 2.891 |
| Separated | 1.298 | 1.302 | 0.795 | 0.182 | 9.263 |

From the above table, it is evident that the risk of MI for widowed subjects is twofold compared to the married men. However, the risks are also greater for other categories although not formally significant. The lack of significance could be due to the small numbers in each category as 87.4% of men were married. If the not currently married categories are collapsed into a single group they have an increased risk of over 50% compared to married men (OR=1.52, 95% CI: 1.05-2.20) and HR=1.57, 95% CI: 1.12-2.20), showing that the marital status is an important risk factor for MI.

There were no stroke events in the separated men making it impossible to compare them with the married men. Furthermore, there were only 24 stroke events (8 each) between the single, widowed and divorced men making the results unrealistic. Therefore, married men were compared with the rest of the men. The risk of stroke increases by a relative odds of 1.58 (p=0.06, 95% CI: 0.99 – 2.51) and a hazard ratio of 1.66(p=0.02, 95% CI: 1.07 – 2.59) for non-married men compared to married men. Though the results are marginally not significant for the logistic model, using the survival model shows an increased risk of stroke for others compared to married men.

### 7.3.3 Social class

Social class was measured in CaPS using the six groups of the Registrar's General classification of occupations. Social class groups 1 and 2 were professionals and managers in their profession while social class group was divided into manual and non-manual occupation depending upon whether they were tradesmen or office workers respectively. Social class groups 4 and 5 were classified for the occupation of semi-skilled manual and unskilled manual workers respectively.

Although this classification has been criticised as being difficult to interpret, it has been widely used in the UK to adjust for socioeconomic effects [178]. In CaPS social class was not found to be related to MI or stroke. After collapsing these different groups to manual and non-manual working groups, there was no evidence of an association with MI (OR = 0.99(95% CI: 0.75 – 1.32), Haz = 1.03(95% CI: 0.79 – 1.35)) and little evidence of an association with stroke (OR = 1.23 (95% CI: 0.84 – 1.79), Haz = 1.27(95% CI: 0.88 – 1.84)). Furthermore, there was no evidence of an association between different social class groups and the risk of MI or stroke.

## 7.3.4 Diabetes

The risk associated with diabetes is modifiable. This was not discussed earlier as it will be used for adjustment purposes only. Diabetes has been shown in several studies to be related to increased risk of heart disease [179, 180] In CaPs, however, diabetes status showed no significant association with MI (p=0.5). This is not the case for stroke and diabetes. There appears to be a fourfold increase in odds of stroke for diabetic men (OR=4.22 (95% CI: 2.26 – 7.87), Haz = 3.82 (95% CI: 2.20 – 6.65)).

## 7.3.5 Electrocardiogram of Left Ventricular Hypertrophy (ECG LVH)

Left ventricular hypertrophy (LVH) is a thickening of the heart muscle of the left ventricle of the heart. This is a marker for heart disease and is measured with electrocardiogram to identify subjects whereas heart muscle has thickened. The odds of MI for men who were positive for ECG LVH was OR=1.30 (95% CI: 0.99 – 1.72), Haz = 1.32 (95% CI: 1.03 – 1.71). Though the results for the logistic model were marginally not significant, the Cox proportional hazard model does show a significant increase in risk for men with positive ECG LVH. In the case of stroke the risk was OR=1.48 (95% CI: 1.04 – 2.12), Haz = 1.53 (95% CI: 1.09 – 2.15) and shows that men with positive ECG LVH have a higher risk of stroke.

## 7.3.6 Discussion

Age, marital status, and ECG LVH appear to be significant risk factors for MI. Along with age, marital status and ECG LVH, diabetes also appears to be a significant factor for stroke. However, social class seems to show no significant association with MI and a possible association with or stroke.

## 7.4. The Caerphilly Analysis for MI

## 7.4.1 MI

### 7.4.1.1 Multivariate logistic and Cox's proportional hazard model

In order to arrive at a comprehensive but parsimonious multivariate model, only variables that were highly inter-correlated were considered for exclusion. SBP and DBP are both major risk factors of CVD. Having both of them in the same model is not necessary because SBP and DBP are highly correlated (correlation coefficient 0.62). In addition, there appears to be no need to include HDL-C in the model because total cholesterol is an addition of various cholesterols including HDL-C.

DBP and HDL-C were excluded from multivariate analysis. This decision was taken after contemplating various risk models on the Caerphilly data. The following table shows odds and hazard ratios of multivariate logistic regression and the Cox's proportional hazard models for the 16 years follow-up (nearly 17 years of follow-up).

### Table 7.2a Multivariate logistic regression for MI

|  | Odds Ratio | z | P>\|z\| | 95% CI | |
|---|---|---|---|---|---|
| SBP | 1.017 | 5.050 | 0.000 | 1.010 | 1.024 |
| T-CHOL | 1.404 | 4.570 | 0.000 | 1.214 | 1.624 |
| BMI | 1.032 | 1.440 | 0.150 | 0.989 | 1.077 |
| S. CLSSS | 0.795 | -1.350 | 0.177 | 0.570 | 1.109 |
| M. STATUS | 1.152 | 0.610 | 0.539 | 0.734 | 1.808 |
| LOGALC | 0.900 | -1.920 | 0.054 | 0.809 | 1.002 |
| Moderate LEI | 0.814 | -1.070 | 0.283 | 0.558 | 1.185 |
| Heavy LEI | 0.623 | -2.420 | 0.015 | 0.425 | 0.914 |
| SMOKING | 2.035 | 4.260 | 0.000 | 1.467 | 2.823 |
| AGE | 1.012 | 0.650 | 0.513 | 0.976 | 1.049 |
| DIABETES | 1.573 | 1.120 | 0.265 | 0.710 | 3.485 |

### Table 7.2b Multivariate Cox's Proportional hazard model for MI

|  | Haz Ratio | z | P>\|z\| | 95% CI | |
|---|---|---|---|---|---|
| SBP | 1.016 | 5.550 | 0.000 | 1.011 | 1.022 |
| T-CHOL | 1.349 | 4.560 | 0.000 | 1.186 | 1.534 |
| BMI | 1.025 | 1.280 | 0.202 | 0.987 | 1.065 |
| S. CLSSS | 0.830 | -1.200 | 0.230 | 0.613 | 1.125 |
| M. STATUS | 1.193 | 0.860 | 0.387 | 0.800 | 1.779 |
| LOGALC | 0.908 | -1.970 | 0.049 | 0.825 | 1.000 |
| Moderate LEI | 0.816 | -1.180 | 0.240 | 0.581 | 1.146 |
| Heavy LEI | 0.633 | -2.580 | 0.010 | 0.447 | 0.896 |
| SMOKING | 1.985 | 4.500 | 0.000 | 1.472 | 2.677 |
| AGE | 1.014 | 0.840 | 0.400 | 0.982 | 1.048 |
| DIABETES | 1.517 | 1.150 | 0.251 | 0.745 | 3.090 |

In the above table SBP, total cholesterol, BMI, age and logalc are continuous variables. Social class was dichotomised to manual and non-manual, marital status to married or not married, smoking to current non-smokers and current smokers. Leisure activity was used a categorical variable with three groups where low leisure activity was compared with moderate and high leisure activity.

The Table 7.2a. can be rewritten in the form of an equation as follows

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.018 SBP + 0.35 TCHOL + 0.026 BMI - 0.236 Social\_Class$$
$$+ 0.132 Marital\_Status - 0.126 LOGALC - 0.239(Moderate\_LEI) - 0.484(Heavy\_LEI)$$
$$+ 0.414 SMOKING\_status + 0.012 AGE + 0.501 DIABETES - 8.014$$
$$\cdots\cdots (7.1)$$

This model implies that, higher blood pressure, higher total cholesterol, lower alcohol consumption and high levels of leisure activity, and higher smoking increases the risk of MI over the period of 16 years.

The remaining risk factors were not significantly associated with MI. Although age, social class, marital status and BMI did not show significant associations, they were retained in the model as they may have affected other risk factors that were related to MI. In the multivariate models shown in table 7.1, the logarithm of alcohol consumption seems to show that the higher the alcohol consumption, lower the risk of MI. However, in the uni-variate analysis, both logistic and survival, it showed no significant results. This might be due to the presence of other risk factors in the model and their correlation with log alcohol consumption. Fractional polynomial regressions were also used for these models and found to show no significant improvement (deviance improvement of 0.8 and 0.9 for logistic and Cox's fractional polynomial models respectively).

In the uni-variate analysis of SBP in the Caerphilly study, for every 10-mmHg increase in SBP, it was shown that the odds increase by about 16% of MI. However, for an increase of 10 mmHg of SBP in the multivariate scenario, there appears to be an 18% increase of MI. Some of the other non-significant risk factors such as BMI and age might have added an effect and thus increased the risk of MI due to SBP. The

170

Framingham study results show that the risk of MI increases by 50% in subjects with SBP>150 compared to the subjects with SBP<150 [113]. In the multivariate Caerphilly analysis, however, there appears to be an increased odds of 62% when subjects with SBP>=150 and SBP<150 are compared. Wilhelmsen et al [114] showed those who have SBP≥176 have RR of 1.92 for an MI event compared to those who had SBP<145 when adjusted for various standard risk factors. In the Caerphilly multivariate analysis, there appears to be more than threefold increase in the odds of MI when subjects with SBP<145 compared to subjects with SBP>175.

For total cholesterol, in the Caerphilly study, there appears to be no difference between the results of the univariate and multivariate analyses in terms of either odds or hazard ratios. When compared to the findings of Corti et al [139] (where the sample was divided into three groups with total cholesterol between 4.16 – 5.19, 5.20 – 6.19 and >6.19), the Caerphilly study showed a significant odds ratio of 1.57 and 2.20 for the high cholesterol groups compared respectively to the low total cholesterol group. This is little different to the results of the univariate analysis (univariate analysis showed an odds ratios of 1.62 and 2.3 respectively). However, both these odds ratios are higher then the results from Corti 's study [139]. There could be several reasons for these results being different between studies, such as different populations, inclusion of females in the Corti study and the length of follow-up. However, it is important to note that there is very little change in the odds ratios even after adjusting for various other factors, implying that total cholesterol is a reasonably independent risk and an important target for a health-related intervention. For an increase in 1 mmol/L, the odds of an MI were increased by 28% according to Anam et al [181] compared to the 40% increased odds in this multivariate analysis. Although Anam study was a meta analysis, it was of men who were older (>65 years) than those in the Caerphilly cohort and this might account for the difference in relative odds between studies.

Alcohol consumption seems to show a protective effect for the risk of MI that is very close to statistical significance (significant for survival analysis). Other studies in the literature review report similar findings [63, 64]. The CaPS univariate analysis did not find this association, however, the multivariate analysis may be more sensitive due to possible masking effects of other variables not being taken into account. When

171

alcohol consumption was used in the multivariate analysis without taking the logarithm, there appears to be a fall in the significance level, suggesting that a logarithmic value of alcohol might be a better measure of MI. Due to the 'J' shaped relationship between alcohol consumption and MI, multivariate fractional polynomial regression was also used. However, this showed no signs of association although the gain in deviance was only 0.3, which is negligible. Reducing alcohol consumption might result in reducing the risk of other diseases but for MI, the effect would not be large.

For the univariate analysis of smokers, three groups of smokers were considered (non-smokers, ex-smokers and current smokers). However, if observed carefully, the risk of MI for ex-smokers and non-smokers appear to be very similar. Therefore, those two groups were collapsed and current smokers were compared with not current smokers (non-smokers). In the univariate analysis, current smokers showed about 1.8 times higher odds of MI compared to non-smokers.

In the Caerphilly multivariate analysis there is a two-fold increased odds for MI for current smokers. This compares favourably with other major studies including Sir Richard Doll's Doctors study and a huge American study, both of which showed roughly a two-fold increased risk of MI compared to non-smokers [84, 86]. The CaPS univariate and multivariate analyses showed similar results, suggesting that smoking is relatively independent of other risk factors and a suitable target for health interventions.

Low levels of leisure inactivity appeared to show an increased risk of MI when compared to the moderate and high leisure activity in the univariate analysis. However, in the multivariate scenario, only high leisure activity showed a significant reduction in risk. The British regional heart study in 1998 reported that the odds of fully adjusted CVD mortality for light activity=0.61(0.36-1.04), moderate activity=0.36(0.16-0.80) and heavy=0.65(0.37-1.14) compared with very little activity [104]. In both the Regional Heart Study and CaPS, high levels of activity were related to increased risk although the outcome of the Regional Heart Study was CVD mortality whereas the outcome in this analysis is first MI.

The remaining risk factors included in the multivariate analysis were not significantly associated with MI or stroke. This does not mean that these factors are not important; it may be that their independent effects were too small to be detected in this study, or that some of their effect may be accounted for in associations found with other risk factors.

To arrive to the multivariate models in the Table 7.2, there were several model selection methods were used. A to regression models with a cutoff probability of 0.10 were used identify the significant factors. However, using this method, some of the important factors such as age and BMI were excluded from the models. Therefore, a residual analysis was performed in order to arrive at the best possible model. The models discussed in the table 7.2 were the models with the smallest residuals.

## 7.4.2 Multivariate models for MI with psychological variables

In the previous section, multivariate models with the standard cardiovascular risk factors were discussed. However, several psychological risk factors were discussed in chapter 6 and it would be interesting to see how they behave in the multivariate scenario.

Table 7.3 illustrates the comparisons between standard cardiovascular risk factors and psychological risk factors. It appears that the men with high anxiety are more likely to be manual workers. Men with high psychological distress are more likely to be smokers and have low leisure activity. However, men with high anger-out are less likely to be married, have low leisure activity, be slightly older in age, and have high systolic blood pressure compared to the men with low anger-out. This might suggest that the subjects with high anger-out are at higher risk for MI once the effects of low leisure activity, current smoking, high blood pressure etc are taken into account.

Table 7.3

Baseline Characteristics of men with and without Negative emotion

Negative emotion status

| | | STAI anxiety | | GHQ anxiety/depression | | Anger expression | |
|---|---|---|---|---|---|---|---|
| | | No (n=929) | Yes (n=787) | No (n=1378) | Yes (n=338) | Low anger out (n=498) | High anger out (n=1218) |
| Manual occupation | n(%) | 562 (61%) | 548*(70%) | 888(64%) | 222(66%) | 314(63%) | 796(65%) |
| Married | n (%) | 828(89%) | 681(86%) | 1,217(88%) | 441(87%) | 458(92%) | 1050#(86%) |
| Current smoker | n (%) | 384(41%) | 362(46%) | 581(42%) | 165*(49%) | 206(41%) | 540(44%) |
| High leisure activity | n (%) | 333(36%) | 265(34%) | 503(37%) | 95*(28%) | 197(40%) | 401#(33%) |
| No diabetes | n (%) | 893(96%) | 764(97%) | 1333(97%) | 324(96%) | 479(96%) | 1178(97%) |
| Age | mean(SD) | 56.9(4.5) | 56.8(4.6) | 56.9(4.5) | 56.4(4.5) | 55.9(4.4) | 57.2#(4.5) |
| Cholesterol | mean(SD) | 5.68(0.99) | 5.74(1.08) | 5.71(1.03) | 5.71(1.05) | 5.8(1.0) | 5.7(1.03) |
| Systolic pressure | mean(SD) | 147.2(22.5) | 146.6(23.3) | 147.4(22.5) | 144.9(24.4) | 143.8(22.5) | 148.2#(22.9) |
| Body mass index | mean(SD) | 26.6(3.6) | 26.2(3.8) | 26.5(3.6) | 26.1(4.0) | 26.6(3.8) | 26.3(3.6) |
| Ln Alcohol | mean(SD) | 4.2(1.5) | 4.3(1.5) | 4.3(1.5) | 4.2(1.5) | 4.3(1.4) | 4.3(1.5) |

Comparison between negative emotions other than anger groups       *       $p <= 0.05$
Comparison between anger groups       #       $p <= 0.05$

STAI is the trait anxiety score dichotomised as discussed in section 6.2.
GHQ30 is the psychological distress indicator dichotomised as discussed in section 6.1.
Anger-out is the Anger-out score dichotomised as discussed in section 6.3.2.2

A multivariate analysis that included standard and psychological risk factors was carried out. To allow comparison of trends over time between logistic and Cox methods, analyses were conducted at five, nine and sixteen years follow-up. For convenience, these time periods were chosen to be comparable with previous analyses of these data for CHD at 5 and 9 years, although the CHD analyses are not the subject of this thesis. After five and nine years, there were 70 and 139 MIs recorded respectively.

AS a first step, the independence of association between psychological risk factors was investigated. The incidence of MI was compared for the different combinations of psychological risk factors at 5, 9 and 16 years and is shown in table 7.4. Of particular interest in this table is the greater incidence in men with high anger-out compared to the low anger-out in the presence of anxiety and psychological distress. At 5 years, the incidence of MI was 3.4% in men with low anger-out and 7.1% men with a high anger-out. At 9 years, the incidence was 4.2% and 14.3% respectively whilst at 16 years the incidence was 12.7% and 16.2% respectively suggesting that the risk of MI appears to be higher in the subjects with high anger out after selecting for high anxiety and the presence of psychological distress. The evidence for other independent effects is not so clear from the layout of table 7.4, but it may be concluded that there is some independence between anger-out and the other psychological risk factors.

**Table 7.4**

**MI Incidence rates at 5, 9 and 16 years follow-up according to negative emotion status**

| Anxiety present | Depression /anxiety present | Anger Expression Style | Number in group | First MI | | |
|---|---|---|---|---|---|---|
| | | | | **0-5 years** | **0-9 years** | **0-16 years** |
| No | No | Low anger out | 176 | 3 (1.7%) | 9 (5.1%) | 10 (5.7%) |
| | | High anger out | 687 | 28 (4.1%) | 58 (8.4%) | 91 (13.2) |
| | Yes | Low anger out | 26 | 0 (0%) | 2 (7.7%) | 2 (7.7%) |
| | | High anger out | 40 | 1 (2.5%) | 2 (5.0%) | 6 (15.0%) |
| Yes | No | Low anger out | 178 | 9 (5.1%) | 13 (7.3%) | 18 (10.1%) |
| | | High anger out | 337 | 14 (4.2%) | 28 (8.3%) | 55 (16.2%) |
| | yes | Low anger out | 118 | 4 (3.4%) | 5 (4.2%) | 15 (12.7%) |
| | | High anger out | 154 | 11 (7.1%) | 22 (14.3%) | 26 (16.9%) |
| Total | | | 1716 | 70 (4.1%) | 139 (8.1%) | 223 (13.0%) |

The next step is to conduct a formal analysis including psychological and standard risk factors. Table 7.5 shows the results of this analysis. In this table the 'simple associations' are those where the association with the psychological variable is adjusted for standard risk factors only and the 'independent associations' are those where the effect of other psychological factors has also been adjusted for. Only for anger-out there is evidence of an association of MI which is not accounted for by the effect of standard risk factors. After 16 years follow-up the hazard ratio of MI in men with high anger-out compare to men with low anger-out was 1.64 (p=0.007). If the effect of the other psychological risk factors is also taken into account the hazard ratio after 16 years was 1.72 (p=0.004).

It would be interesting to see whether the combined effect of the psychological risk factors amplifies their individual effects. In Table 7.6 additive and amplified effect models are compared. In Table 7.6, 'additive effects model' refers to the sum of the independent effects (main effects) for the individual psychological risk factors. The 'amplified effects model' refers to combining both additive and multiplicative effects (main effects and 2 and 3 way interactions) to obtain an estimate of the overall effect of the risk factors. The HR of 2.72 (95% CI: 1.23, 6.01) shown in the table 7.6 corresponds to the amplified effect of high anger-out + high STAI anxiety + High GHQ depression/anxiety + 2 way interactions + 3 way interaction between the factors compared to the amplified effect of low anger-out + low STAI anxiety + low GHQ depression/anxiety + 2 way interactions + 3 way interaction.

**Table 7.5**

**Odds ratios and hazard ratios for MI according to negative emotion status with adjustment for vascular risk factors***

| Model | Negative emotion | MI at 5 years OR (95%CI) | MI at 9 years OR (95%CI) | MI at 16 years HR (95%CI) |
|---|---|---|---|---|
| **Simple associations** | STAI (High Anxiety) | 1.44 (0.84, 2.47) p=0.18 | 1.13 (0.77, 1.66) p=0.54 | 1.19 (0.90, 1.59) p=0.23 |
| | GHQ (Depression/Anxiety) | 1.10 (0.57, 2.10) p=0.78 | 1.05 (0.65, 1.70) p=0.83 | 1.08 (0.75, 1.53) p=0.69 |
| | High anger out | 1.23 (0.65, 2.31) p=0.52 | 1.42 (0.89, 2.26) p=0.14 | 1.64 (1.14, 2.34) p=0.007 |
| **Independent associations** (adjusted for anxiety & depression/anxiety & anger expression) | STAI (High Anxiety) | 1.53 (0.86, 2.71) p=0.15 | 1.19 (0.78, 1.80) p=0.42 | 1.26 (0.93, 1.71) p=0.13 |
| | GHQ (Depression/Anxiety) | 0.97 (0.49, 1.92) p=0.92 | 1.04 (0.62, 1.72) p=0.90 | 1.05 (0.72, 1.53) p=0.80 |
| | High anger out | 1.34 (0.70, 2.56) p=0.38 | 1.48 (0.92, 2.38) p=0.11 | 1.72 (1.19, 2.48) p=0.004 |

* Baseline risk factors adjusted for:

Age,
Social class
Systolic pressure
Total cholesterol
Diabetes (Y/N)

Smoking
Marital status
Body mass index
Alcohol consumption
leisure activity (light/moderate/high)

178

## Table 7.6
### Effects of high negative emotions using additive and amplified models, and a comparison between high and low negative emotion groups; [1] with adjustment for vascular risk factors*

| Outcome | Follow-up period | Additive effects model[2] (Main effects only) | Amplified effects model[3] (Main and interaction effects) | High negative emotion group |
|---|---|---|---|---|
| MI | OR at 5 years | 1.98 (0.68, 5.79) p=0.21 | 4.63 (0.95, 22.54) p=0.058 | 4.60 (0.98, 12.23) p=0.053 |
| | OR at 9 years | 1.82 (0.82, 3.99) p=0.14 | 2.37 (0.96, 5.86) p=0.063 | 2.43 (1.04, 5.71) p=0.040 |
| | HR at 16 years | 2.26 (1.26, 4.06) p=0.007 | 2.72 (1.23, 6.01) p=0.014 | 2.69 ( 1.21, 5.95) p=0.015 |

Notes:

1. Reference group for all analyses is men with low anger out + low STAI anxiety + low GHQ depression anxiety
2. Additive model is the linear combination of: high anger out + high STAI anxiety + High GHQ depression/anxiety
3. Amplified model is the linear combination of: high anger out + High STAI anxiety + High GHQ depression/anxiety + 2 way interactions + 3 way interaction

* Baseline risk factors adjusted for: age, smoking, social class, marital status, systolic pressure, body mass index, total cholesterol, alcohol consumption, diabetes, leisure time activity

At 5 years the risk for men with all three psychological risk factors was HR=1.98 compared to men with no psychological risk factors using the additive model, and HR=4.63 using the multiplicative model. At 9 years the risks were HR=1.82 for the additive model and HR= 2.36 for the amplified model. At 16 years the risks were HR=2.26 and HR=2.72 respectively. The interaction terms between the high anger-out, high depression and high anxiety was not significant (two way interaction or 3 way interaction). Deviances between the additive and the amplified models showed no significant difference between the two models (test was conducted using the method explained in section 3.4.1 at 5% significance level). Deviance gain by the additive model over the amplified model was 2.35 which is not significant.

These results are suggestive of an amplified effect during the earlier years of follow-up, but the size of CaPS appears to be too small to judge this convincingly. A further analysis was conducted using a less sophisticated approach in which the sample was grouped into men with no psychological risk factors, men with high anger out only and their Kaplan-Meier survival curves were plotted in Figure 7.1. Another figure of Kaplan-Meier survival curves of men with no psychological risk factors and men with all three (high anxiety, high psychological distress and high anger-out) is also plotted. These results are shown in the final column of table 7.6. These results more closely follow those of the amplified model than the additive model, suggesting that there may be an amplified effect of these psychological risk factors.



**Figure 7.1: Kaplan-Meier Survival Estimates (in years) according to negative emotion status**
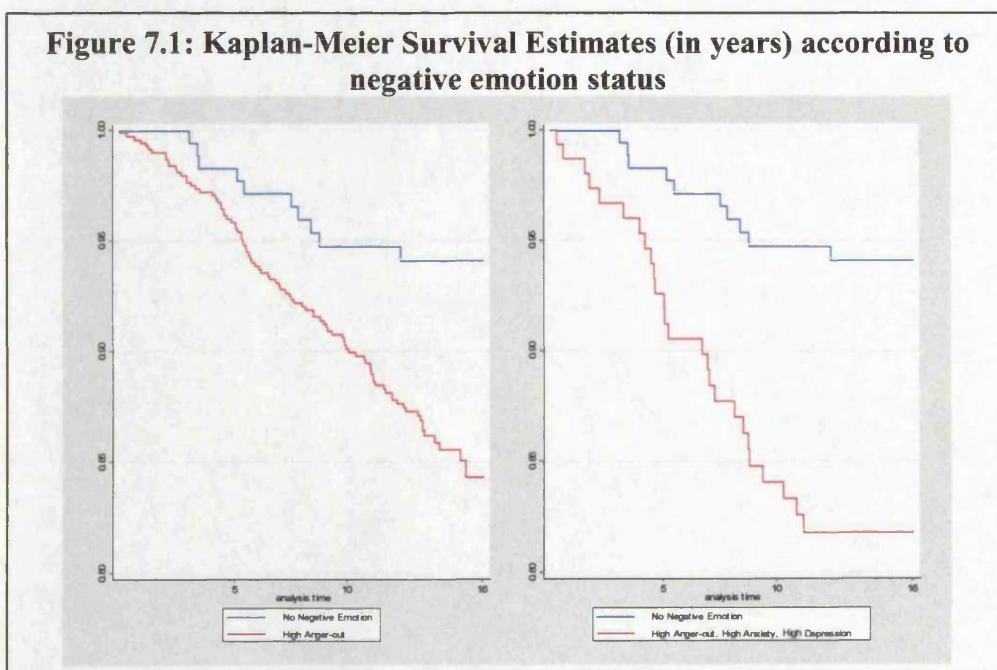
The effect of more than one psychological risk factor may be visualised from the Kaplan-Meier plots in the above figure. In comparison to men with no psychological risk factors (low anger-out, low anxiety and no psychological distress), the left hand plot in Figure 7.1 shows men with high anger out expression only are at higher risk throughout the follow-up period. In plot 2, however, the addition of anxiety and depression to high anger out appears to increase the risk but it is important to note that the additive effect is statistically not different from the amplified model. Therefore, care must be taken to interpret these results carefully because the survival of all groups appears to decrease sharply during the earlier years and due to very few MI evens after 9-10 years.

These results show that psychological risk factors are related to the occurrence of MI. These associations are likely to reflect complex patterns of risk association. There may also be a possible indirect effect of psychological risk factors on lifestyle [182] although major lifestyle risk factors were adjusted for in this analysis. It could also be argued that the presence of negative emotions implies poor health at baseline which has a subsequent effect on cardiovascular disease[183].

The absence of an association of depression with MI in the Caerphilly data may be due to the small numbers of persons in the Caerphilly cohort with high depression. The effect of not expressing anger has been reported previously from the Caerphilly analysis [25]. The distinction between different anger scales such as high anger-out and suppressed anger is used in this analysis is discussed elsewhere [25]. High anger-out has been associated with high cardiovascular risk in several other cohort studies [184-186]. Depression and anger are closely linked emotions, especially in men according to one study [187] and another study suggests that anger and depression interact to affect the risk of peripheral artery disease [188]. Depression and anxiety are also closely linked emotions, with a high co-morbidity in the community [189]. This analysis suggests that the presence of high anger out, anxiety and depression may increase MI risk more than the risk associated with any single psychological factor discussed in the previous chapter.

In conclusion, this analysis suggests that not only the standard biological and behavioural factors risk factors have a role in estimating the risk of MI. These findings show that psychological risk factors too are important and should be legitimately included in risk models such as the Framingham risk model.

Although attitudes and social norms could also have been included in the multivariate analysis, the associations were so small that this analysis was not carried out.

## 7.5 Caerphilly analysis for Stroke

A similar analysis to that described above was conducted for stroke. The first multivariate model using the standard risk factors (as used in table 7.2.a and 7.2.b for MI) were included in the model.

### 7.5.1 Multivariate logistic and Cox's proportional hazard model

The following table shows the multivariate logistic and Cox's proportional hazard models.

### Table 7.7a Multivariate logistic regression for stroke

| | Odds Ratio | z | P>|z| | 95% CI | |
|---|---|---|---|---|---|
| SBP | 1.008 | 0.004 | 0.064 | 0.999 | 1.017 |
| T-CHOL | 0.998 | 0.098 | 0.984 | 0.824 | 1.209 |
| BMI | 1.030 | 0.029 | 0.290 | 0.975 | 1.089 |
| S. CLSSS | 1.001 | 0.219 | 0.997 | 0.652 | 1.537 |
| M. STATUS | 1.546 | 0.408 | 0.099 | 0.921 | 2.593 |
| LOGALC | 1.189 | 0.087 | 0.018 | 1.031 | 1.371 |
| SMOKING | 1.178 | 0.247 | 0.434 | 0.781 | 1.778 |
| AGE | 1.086 | 0.026 | 0.000 | 1.037 | 1.137 |
| DIABETES | 5.194 | 1.868 | 0.000 | 2.566 | 10.511 |
| ECG LVH | 1.766 | 0.472 | 0.033 | 1.046 | 2.983 |

### Table 7.7b Multivariate Cox's Proportional hazard model for stroke

| | Haz Ratio | z | P>|z| | 95% CI | |
|---|---|---|---|---|---|
| SBP | 1.008 | 0.004 | 0.042 | 1.0003 | 1.012 |
| T-CHOL | 1.008 | 0.096 | 0.935 | 0.837 | 1.213 |
| BMI | 1.022 | 0.027 | 0.410 | 0.970 | 1.077 |
| S. CLSSS | 1.034 | 0.215 | 0.871 | 0.688 | 1.554 |
| M. STATUS | 1.580 | 0.388 | 0.062 | 0.977 | 2.556 |
| LOGALC | 1.175 | 0.081 | 0.019 | 1.026 | 1.344 |
| SMOKING | 1.254 | 0.249 | 0.255 | 0.849 | 1.851 |
| AGE | 1.094 | 0.025 | 0.000 | 1.047 | 1.143 |
| DIABETES | 4.516 | 1.396 | 0.000 | 2.464 | 8.276 |
| ECH LVH | 2.160 | 0.531 | 0.002 | 1.334 | 3.496 |

From the above table, both survival and logistic regression show that the risk of stroke is significantly associated with SBP, LOGALC, age, ECG and diabetes status. The remaining risk factors are not significant. Alcohol consumption showed some degree of protection for MI but it significantly increases the risk of stroke. Age and diabetes status did were not associated to the risk of MI significantly but they were highly significant risk factors for stroke in the univariate analysis.

To arrive to the multivariate models in the Table 7.7, there were several model selection methods were used. A to regression models with a cutoff probability of 0.05 were used identify the significant factors. However, using this method, some of the important factors such as SBP were excluded from the models. Therefore, a residual analysis was performed in order to arrive at the best possible model. The models discussed in the table 7.7 were the models with the smallest residuals.

SBP is considered as a major risk factor of MI or stroke and this is supported in the analysis above. The p values in the univariate model (Table 5.4) were very high compared to the multivariate model given in the above table. This may be because he sample used was smaller due to missing values for the other risk factors.

High alcohol consumption appears to show an increased risk of stroke in the Caerphilly analysis with a linear relationship. A meta analysis study in the literature review suggested that that the relationship between alcohol consumption and stroke is not linear [74]. Therefore, a fractional polynomial model was also fitted to this model which showed no significant improvement over the linear model (deviance gain of 0.12). This suggested that as alcohol consumption increases, the risk associated with stroke increases in the Caerphilly men.

Age and diabetes status are also associated with risk of stroke. Diabetes is not discussed in detail here, as it is an important outcome by itself and not covered in this thesis. ECG status appears to show highly significant association with stroke. For the logistic model, with the inclusion of the ECG, SBP showed no significance at 5% level (Table 7.7a). However, in the survival analysis, it even SBP was significant in presence of ECG (Table 7.7b).

An analysis was carried out at 5, 9 and 16 years to see whether there was any independent effect of psychological risk factors for stroke. After adjustment for standard vascular risk factors, irrespective of the combinations of the psychological factors, the risk of stroke was not associated with any psychological risk factors. This may be due to the very low number of stroke events or may be because there is no increased risk of stroke associated with psychological factors. Attitudes and social norms were not found to be related to stroke in a multivariate analysis.

## 7.6 Discussion

Multivariate analysis carried out with psychological risk factors suggest that the risk of MI increases significantly with the presence of anger-out. However, Kaplan-Meier survival plots suggested that the risk of MI is more when all the three negative emotions were present. These results should be carefully interpreted because this analysis is underpowered. Furthermore, it would be naïve to talk about any psychological mechanisms that might explain causal pathways to the risk of MI due to the potential complexity of the mechanisms involved.

This chapter mainly focuses on identifying important risk factors for MI and stroke in a multivariate scenario using logistic regression and survival analysis. It mainly focuses on understanding the effect of psychological risk factors. In the process of understanding the psychological risk factors on the risk of MI and stroke, anger-out proved to be a highly significant risk factor. Alcohol consumption was marginally significant, giving a protective effect for MI and a negative effect for stroke. However, none of these other additional risk factors have been used in risk models, such as the Framingham risk model. These models may be considered to be incomplete and these data suggest that they may be improved by including other risk factors such as anger-out and alcohol consumption. An improved risk model would ultimately enable GPs to identify possible subjects with cardiovascular risk with more certainty and hence reduce the costs incurred to NHS.

# Chapter 8

## 8.0 Risk models for MI and stroke

Suppose that a physician's intention is to assess a patient's cardiovascular risk. This judgement is likely to be based on factors such as age, tobacco use, obesity, blood pressure and other measurements. To estimate a risk, and to identify people at high risk, some form of scoring system is needed. This may estimate an actual risk or simply give a score with associated guidelines, of the form 'score> $\theta$ means a person is at high risk' where $\theta$ is a chosen threshold, or cutoff. It is therefore essential to produce an efficient risk scoring system which can be used to identify people at high risk of MI or stroke.

There are several risk scoring systems that are currently available, including the Framingham risk model, SCORE risk charts and others described below. All these risk models or charts have been derived using data on disease outcome and risk factors from different studies. But why are there so many models? Do some work better than others? Are they population dependent? In this thesis the population of interest is that of men living in Caerphilly. Which is the best scoring method for that population? Is it possible to derive a better one?

Building on the multivariate models derived in Chapter 7, new risk models will be derived and compared with some of the existing ones in their ability to predict those at high risk. This comparison will be effected by identifying a screening process based on a risk scoring system. This will be used to predict which men are at high risk and the sensitivity and specificity of this process will be evaluated. By examining the ROC curve the screening process will be optimised.

The purpose of such a scoring system for this thesis is not merely to identify men at high risk. As was explained in Section 1.3 such a scoring system is essential for the main aim of the thesis, that is evaluating the effects of public health interventions to improve the levels of risk factors. For these purposes a model which quantifies the risk is needed, as the aim will be to model how this risk changes through interventions.

A number of statistical issues arise in tackling this work. The models which have been fitted so far have been either logistic regression or Cox's proportional hazards models. The former models the probability of an event occurring in a given and fixed period of time, and the result is an estimated probability for a given person with specified levels of risk factors. The latter effectively models the relative risk, as the centrepiece of the model is a hazard ratio, comparing the hazard for an individual with certain risk factors with a baseline one. It does not lead directly to a probability of an event and, as well as being useful for a GP, that is essential for the simulation modelling in Chapter 9. There are two approaches that can be used for the survival analysis. One is to try to convert the relative risk estimates into absolute risks by estimating the baseline function, and a number of authors have taken that line. An example of this approach is the PROCAM model [190]. The other is to use parametric survival analysis instead. In this a particular parametric model is used for the hazard function, such an exponential, Gompertz or Weibull model. These lead directly to probabilities of events as functions of time. That is the approach that will be followed here.

After estimating the parameters of these new models, it is important to see whether these models differ in their performance from existing ones. The main means of comparison will be in terms of the predicted number of events compared to the observed number and in the comparison of ROC curves, which reflect the variation in sensitivity and specificity from other existing models. Therefore these different risk scores or absolute risks estimated by different models should be calibrated. Calibrating means choosing a cutoff point $\theta$ for a model which is optimal in some sense, such as maximising the sum of the sensitivity and specificity.

## 8.1 Literature review

There are several risk scores available from different studies discussing multivariate risk models for the risk of CVD [53, 190-195]. The first study which came up with risk equations for predicting the risk of various outcome variables such as stroke, MI, CVD, CHD, fatal CHD and fatal CVD was the Framingham study [53]. These equations demonstrated the potential importance of controlling various risk factors such as blood pressure, total cholesterol, HDL cholesterol, glucose intolerance, and left ventricular hypertrophy. Instead of using Cox's proportional hazards models, non-standard parametric accelerated failure time models were used (section 3.6.4.3). Simplified forms of the Framingham risk functions have been used in New Zealand [196, 197], European [198-200], and UK [201-204] guidelines for prevention of CHD and to target treatment at those at highest absolute risk.

The PROCAM study in 2002 constructed a new risk model using Cox's proportional hazards model with eight risk factors including age, LDL cholesterol, smoking, HDL cholesterol, SBP, family history of premature MI, diabetes mellitus and triglycerides to predict coronary events. They derived a point scoring system based on the beta coefficients estimated from the Cox model to compare it with the Framingham model [190]. Ten-year risk calculations were based on the probability of a cardiovascular event in ten years. It appears that they have used several other risk factors in their model but nothing showed an improvement over the Framingham model.

A recently published study by Hippisley-Cox et al used a similar type of approach to the PROCAM model by using Cox's proportional hazards models to predict the risk scores of MI, CHD, stroke and transient ischemic attack (TIA) [205]. The risk factors used in this model were age, sex, smoking status, SBP, ratio of total and HDL cholesterol, BMI, family history of CHD in first degree relatives aged less than 60, area measure of deprivation and existing treatment with antihypertensive agents. This model was based on over 600,000 subjects registered at 160 UK general practices and followed-up over 10 years. They concluded that their model performs at least as well as the Framingham model and is better calibrated for the UK populations. However, they failed to add anything new to the model.

Using the Joint European societies' coronary risk charts [199, 200], an individual's absolute risk of developing a CHD event over the coming years can be predicted by locating the appropriate box in the charts in relation to the patient's age, gender, smoking status, blood pressure and cholesterol levels. Total cholesterol concentrations have been used rather than the ratio of total cholesterol to HDL-C because HDL-C is not routinely measured across Europe, whereas total cholesterol measurement is available in every European country.

Some studies have compared the Framingham risk equations with their own model [190, 192-195]. Two studies showed that the risk equations given by the Framingham study overestimate the risk of coronary heart disease, one a UK-based study and the other a European study [192, 193]. Other studies have just compared the Framingham risk equations with their studies [190, 194, 195, 206]. Studies that compared the Framingham risk equations with the risk equations derived from their study also tried to improve the model by adding other factors into the model, but the results were not significant. In each of these models, for each combination of variables, a Cox proportional hazards model was fitted and compared to the Framingham model with a Receiver Operating Characteristic (ROC) curve analysis.

Furthermore, the application of one coronary risk model or chart based on, say, high-risk middle-aged North American populations to European populations at different levels of VAD poses several problems. Since the start of the Framingham study, the incidence of VAD has fallen in many countries, resulting in a decrease in absolute risk. Although the risk model based on the Framingham study predicts absolute risk reasonably well in high-risk middle-aged men, it over-predicts in low risk European populations [207, 208]. Menotti *et al* in 2000 performed a systematic re-analysis of 10 year CHD incidence data from the northern and southern European cohorts of the Seven Countries study [209]. The risk factors examined in this analysis were age, SBP, total cholesterol and smoking habits and they concluded that a single European risk chart is inappropriate because of the different VAD incidence rates and distributions of risk factors in different countries.

The joint European task forces were aware of the limitations of the Framingham-based coronary risk functions and charts. A special research project called SCORE

189

(Systematic COronary Risk Evaluation), funded by the European Union, was set up to develop European risk functions for fatal CVD events. The SCORE project is based on 12 European cohort studies representing a wide range of CVD mortality rates [210]. A Weibull regression model was used to predict the absolute risk for 10 years and the results were converted into a simple chart for estimating the risk. There are several advantages of the SCORE function. First, the function was derived from large representative European populations. Second, the risks of fatal CVD and stroke can be derived separately. Importantly, the SCORE function has been designed to be customised to any European country based on national mortality data.

The Sheffield risk and treatment table differed from the initial risk assessment charts produced by the other studies. Its first aim was to determine whether total cholesterol and HDL-C needed to be measured or not [201]. Subsequently a Sheffield table was published to identify individuals at ≥15% CHD risk over the 10 year period [203]. A random sample of 1000 people aged 35-64 years from the 1995 Scottish health survey was used to estimate these risk equations. The joint British societies produced coronary risk prediction charts and an associated computer program 'Cardiac Risk Assessor', which estimated both 10 year CHD risk and cardiovascular risk including stroke, over the same period [202, 204]. These prediction charts were also based on the absolute risk of CHD. An absolute risk of CHD calculated by the Framingham type risk function was used in both cases.

The Framingham risk equations are based on measurements made at baseline on a single occasion and therefore it is appropriate to estimate risk in clinical practice in the same way. A single measurement will be subject to measurement error and, as was seen in Chapters 4,5 and 6, the relationship between a risk factor and the risk of developing VAD is generally steeper when the regression line is adjusted to allow for measurement error, correcting for regression dilution bias. However, if the model is adjusted for regression dilution bias then the model will essentially relate the risk to the true value of the risk factor (say, SBP or total cholesterol). As only a single value is likely to be available this might not be useful.

However there may have been several values observed over a period of years and these could either be used to try to estimate the 'true' value, or potentially more powerfully the repeated values could be built into a dynamic prediction model. A recently-published study using the Framingham data used dynamic random effects modelling with the repeated measurements from 12, 16 and 20 years follow-up data with 8$^{th}$ year data as the baseline [191].

This brief review of the literature has shown that there are several risk models available. Most of them have compared their results with the Framingham models. None of them have included important psychosocial and psychological risk factors such as those discussed in detail in the previous chapters and these could be incorporated in a Framingham or SCORE type risk function. In this chapter the Framingham risk model will be evaluated on the subjects in the CaPS study and a model, using the same risk factors, will be re-derived from these data. A new risk score including psychological and psychosocial risk factors in a risk model will be derived and compared to the Framingham model.

## 8.2 Framingham risk equations for MI

The aim here is to use the Framingham risk model on the Caerphilly data in order to see how well it predicts VAD events in that population. Then the same variables will be used to derive a new model using the Caerphilly data; this will be followed by a comparison of the Framingham model and the new Caerphilly model. The final aim is to improve the Caerphilly model by adding other important factors analysed in the previous chapters.

The logistic regression and Cox's models discussed in the previous chapters suggested that the psychological factor 'anger out', the behavioural factor leisure activity and attitudes towards exercise behaviours improve the efficiency of measuring the risk of MI. However, the improvement of the model using the attitude towards exercise appears to be marginal and so it will not be included in these new Framingham-type models.

## 8.2.1 Classic Framingham risk equations

The classical Framingham equation used HDL and Total cholesterol levels in mg/dl units instead of mmol/L as used in the previous Caerphilly analysis. To make these results comparable, cholesterol levels from the Caerphilly data were converted to mg/dl units by multiplying the cholesterol levels by a conversion factor of 38.6 as 230 mg/dl $\approx$ 5.96 mmol/L. The Framingham risk equation for MI that was estimated with the Framingham subjects is given below.

$$\hat{\mu} = 11.4712 - 0.7965 \times (\log(age)) - 0.6623 \times (\log(SBP)) - 0.2675 \times (Cigs)$$
$$- 0.4277 \times (\log(total\_cholesterol \div HDL\_Ccholesterol)) - 0.1534 \qquad \ldots\ldots\ldots(8.1)$$
$$\times diabetes - 0.1588 \times (ECG - LVH)$$

where $\log(\hat{\sigma}) = \theta_0 + \theta_1 \hat{\mu} = 3.4064 - 0.8584 \times \hat{\mu}$

$$\text{if } \hat{u} = \frac{\log(followup\_time) - \hat{\mu}}{\hat{\sigma}}$$

then $p = 1 - \exp(-\exp(\hat{u}))$

Cigs defined in the above equation represents the smokers and non-smokers coded as a binary variable (1 & 0 respectively). Diabetes is also a binary variable in the equation where diabetic subjects were coded as 1 and the others as zero.

In the above equation, the probability $p$ is the predicted probability of having an MI, given the levels of the risk factors, according to the Framingham model. By summing the predicted probabilities we can obtain an estimate of the total number of events predicted and compare this with the observed number. We can also use the model for screening purposes by defining a cutoff and identifying any subject with a predicted probability above the cutoff as being at high risk. Clearly, the lower the cutoff chosen, the greater the number of men identified as being at high risk correctly.

The total number of MI events in the 1190 men was 95 (8.0%)at 10 years and 143 (12.0%) by 16 years. For this Framingham model, the sum of predicted probabilities, an estimate of the number of events predicted, was 169 at 10 years and 258 at 16

years. This suggests that the Framingham model over predicts MI events by 78% and 80% at 10 and 16 years respectively.

From a screening viewpoint a cutoff in the predicted risk will be chosen and a man whose risk exceeds that will be classed as being at high risk. The following table summarises the various cutoff points at 10 and 16 years of the follow-up.

### Table 8.1 Framingham model for MI

| MI time/ Cut off | 8% | 12% | 15% | 20% | 25% | 30% | 35% |
|---|---|---|---|---|---|---|---|
| 10 yrs false positives | 784 | 573 | 436 | 241 | 97 | 40 | 11 |
| Actual men with events | 86 | 74 | 62 | 45 | 24 | 12 | 3 |
| Tot num of men at higher risk of MI at 10yrs | 870 | 647 | 498 | 286 | 121 | 52 | 14 |
| Sensitivity at 10 years | 0.91 | 0.78 | 0.65 | 0.47 | 0.25 | 0.13 | 0.03 |
| Specificity at 10 years | 0.28 | 0.48 | 0.60 | 0.78 | 0.91 | 0.96 | 0.99 |
| Predictive value 10 yrs | 0.10 | 0.11 | 0.12 | 0.16 | 0.20 | 0.23 | 0.21 |
| 16 yrs false positives | 966 | 836 | 708 | 522 | 344 | 201 | 90 |
| Actual men with events | 141 | 134 | 127 | 107 | 79 | 58 | 30 |
| Tot num of men at higher risk of MI at 16yrs | 1107 | 970 | 835 | 629 | 423 | 295 | 120 |
| Sensitivity at 16 years | 0.97 | 0.94 | 0.89 | 0.75 | 0.55 | 0.41 | 0.21 |
| Specificity at 16 years | 0.08 | 0.20 | 0.32 | 0.50 | 0.67 | 0.81 | 0.91 |
| Predictive value 16 yrs | 0.13 | 0.14 | 0.15 | 0.17 | 0.19 | 0.20 | 0.25 |

Consider the risk of MI over a 10 year period. The Framingham risk model identifies 870 men as having a higher risk than an 8% cut off; these are regarded as being screen positive. The remainder are screen negative. From these, and their known true event status, the sensitivity and specificity and positive predictive value can all be calculated Table 8.1 shows the values of these summary measures for a range of values of the cut off. As the cut off increases, the sensitivity decreases and the specificity increases; the predictive value increases as the cut off increases up to 30% but then decreases. This pattern is similar to that for a 16 year model too. With all these numbers, how does one arrive at a sensible probability cutoff point?

The following plot shows the ROC curves for the 10 and 16 year models. These plots show the sensitivity and 1-specificity plotted against each other.

**Figure 8.1a ROC Curve for the Framingham model at 10 years**



Area under ROC curve = 0.6883
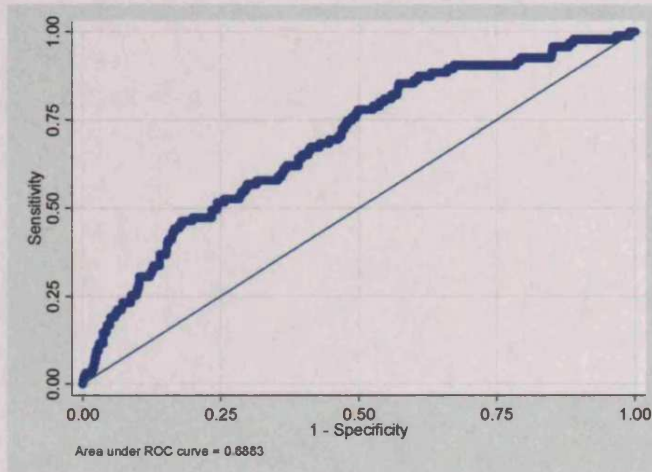
**Figure 8.1b ROC Curve for the Framingham model at 16 years**



Area under ROC curve = 0.6694
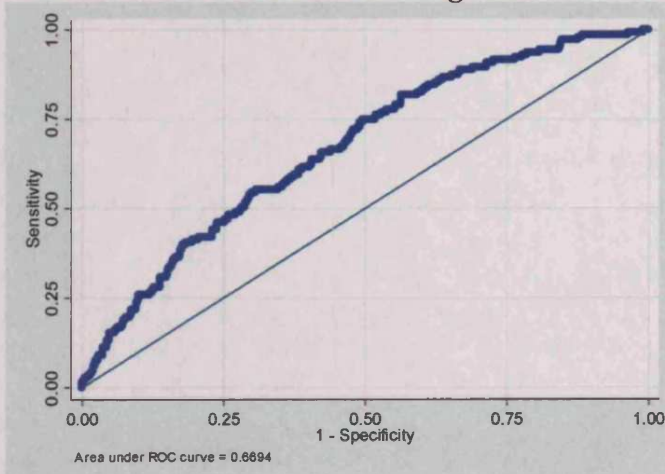
From the above plot, it can be observed that the areas under the curve are 0.6883 & 0.6694 for 10 and 16 year models. The are under the curve is the measure of how useful the screening process is; the higher the value, the better the screening tool. But how does anyone make a sensible judgement of a probability cutoff?

There are several possible ways of choosing a probability cutoff point. The sum of sensitivity and specificity can be maximised and the probability corresponding to that maximum can be taken as a cutoff. The point on the ROC curve closest to the left hand top corner (0,1) can also be used to define a cutoff; these two approaches are likely to yield similar results. Another approach is to make the sensitivity and

specificity as nearly equal as possible   The relative merits of the different approaches depend on the relative costs of false positives and false negatives.  In this case all methods gave similar results and for simplicity the criterion of equal sensitivity and specificity was adopted.

From Table 8.1, it appears that the probability cutoff that should be taken for the Framingham models at 10 and 16 years lies between 15-20% & 20-25% cutoff where the sensitivity and specificity come close to each other. To be precise, the sensitivity and specificity are equal (with common values of approximately 0.61) for cutoff points of 15.4% at 10 years and 23.4% at 16 years.

Using these cutoff points, the percentages of false positives at 10 and 16 years are 88% and 90%. This shows that even with a cutoff point calibrated to the Caerphilly data, there appears to be very high false positive rates.

## 8.2.2 Caerphilly 1 risk equations

Using the Framingham risk equations on the Caerphilly data to predict the number of events showed that there were about 55% more events predicted than occurred over the period of 10 years and about 80% extra over 16.2 years.  A similar model to the classical Framingham model, using the same variables, was fitted to the Caerphilly data. Three different models explained in chapter 3 were estimated and compared. The maximum log likelihood of each model is denoted by '$l$'.

Model 1: $u = \beta' X$,  $\log(\sigma) = \theta_0$,  $l = -1547.8$

Model 2: $u = \beta' X$,  $\log(\sigma) = \theta_0 + \theta_1 \mu$,  $l = -1544.7$

Model 3: $u = \beta' X$,  $\log(\sigma) = \theta_0 + \theta_1 \mu + \theta_2 \mu^2$  $l = -1544.5$

Here $\beta'$ is the coefficient matrix and X is the vector of measurements. Using a likelihood ratio test it was found that models 1 and 2 are highly significant compared to a null model with no risk factors. The linear location accelerated failure time (LLAFT) model (model 1) is rejected in favour of the varying location and dispersion

accelerated failure time (VLDAFT) model (model 2) using a likelihood ratio test [2×(-1547.8 + 1544.7) = 6.2 > $\chi^2_{1,0.05}$]. The new Caerphilly coefficients comparing with the Framingham model's coefficients are given in the following table. Model 3 is statistically no improvement on model 2 and therefore model 2 was adopted as the most appropriate model; this will be referred to as the Caerphilly 1 model below.

**Table 8.2: Comparing Framingham risk equations and Caerphilly equation**

| Coefficients | Framingham Coefficients | Caerphilly Coefficients |
|---|---|---|
| $\theta_0$ | 3.4046 | 1.7485 |
| $\theta_1$ | -0.8584 | -0.5113 |
| $\beta_0$ | 11.4712 | 12.6862 |
| Log(SBP) | -0.6623 | -0.9485 |
| Log(age) | -0.7965 | -0.7734 |
| Log(total-C ÷ HDL-C) | -0.4277 | -0.3360 |
| Smoking (Y/N) | -0.2675 | -0.3287 |
| Diabetes (Y/N) | -0.1534 | -0.2568 |
| ECG-LVH (Y/N) | -0.1588 | -0.1282 |

The false positives, sensitivity, specificity and the predictive values at different probability cutoff values of for the Caerphilly 1 model are given below.

**Table 8.3 Caerphilly 1 model for MI**

| MI time/ Cut off | 6% | 8% | 10% | 15% | 20% |
|---|---|---|---|---|---|
| 10 yrs false positives | 628 | 465 | 340 | 145 | 48 |
| Actual men with events | 81 | 64 | 54 | 34 | 14 |
| Tot num of men at higher risk of MI at 10yrs | 709 | 529 | 394 | 179 | 62 |
| Sensitivity at 10 years | 0.85 | 0.67 | 0.59 | 0.36 | 0.15 |
| Specificity at 10 years | 0.43 | 0.63 | 0.69 | 0.87 | 0.96 |
| Predictive value at 10 yrs | 0.11 | 0.12 | 0.14 | 0.19 | 0.23 |
| 16 yrs false positives | 927 | 809 | 692 | 434 | 273 |
| Actual men with events | 138 | 133 | 127 | 92 | 49 |
| Tot num of men at higher risk of MI at 16yrs | 1065 | 942 | 819 | 526 | 322 |
| Sensitivity at 16 years | 0.97 | 0.93 | 0.89 | 0.64 | 0.34 |
| Specificity at 16 years | 0.11 | 0.23 | 0.34 | 0.59 | 0.74 |
| Predictive value at 16 yrs | 0.13 | 0.14 | 0.16 | 0.17 | 0.15 |

If the numbers of the false positives are compared to the Framingham model, there are fewer in this case.

The predictive value for the 10 year model appears to be increasing as the cutoff increases. But the predictive value for the 16 year model appears to increase till 15% cutoff and decreases after that. The following figure shows the ROC curves for both the 10 year and 16 year models.

**Figure 8.2a ROC Curve for the Caerphilly 1 model at 10 years**



Area under ROC curve = 0.6933
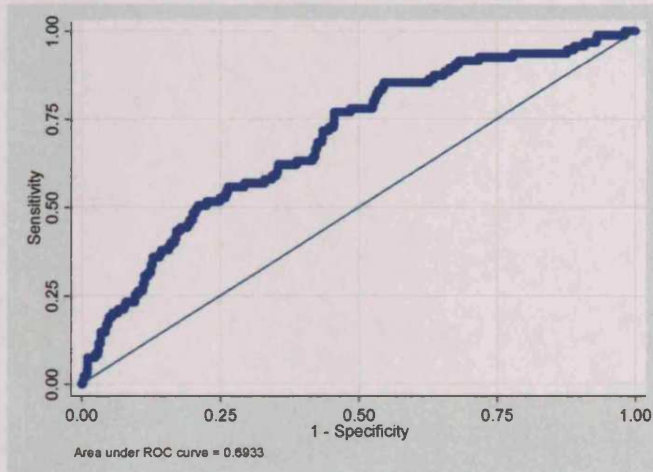
**Figure 8.2b ROC Curve for the Caerphilly 1 model at 16 years**



Area under ROC curve = 0.6733
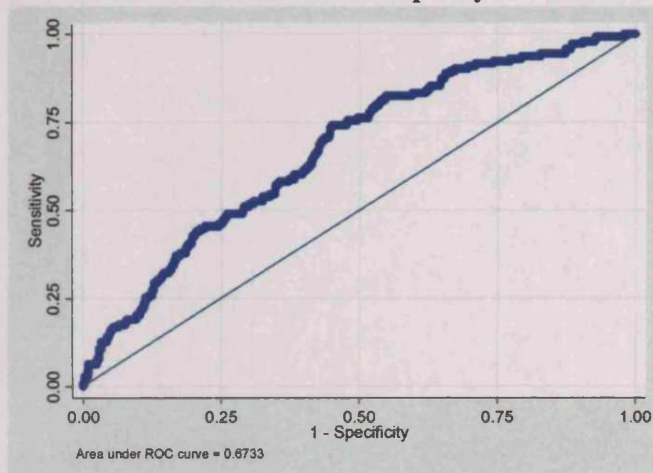
The areas under the curve for the 10 and 16 year models are 0.69 and 0.67, suggesting that the 10 year model is a better model.

Making sensitivity and specificity equal gave cutoffs of 9% and 16% respectively for the 10 and 16 year models; again the common values were approximately 0.61. These cutoff probabilities are lower then the Framingham model but that does not mean that these are better then the Framingham model. This is because the areas under the curve are very similar for both models. To formally compare the two different models, the ROC plots for comparing the Framingham model and the Caerphilly 1 model at 16 years are shown below.

**Figure 8.3 Comparing Framingham model and the Caerphilly 1 model at 16 yrs**
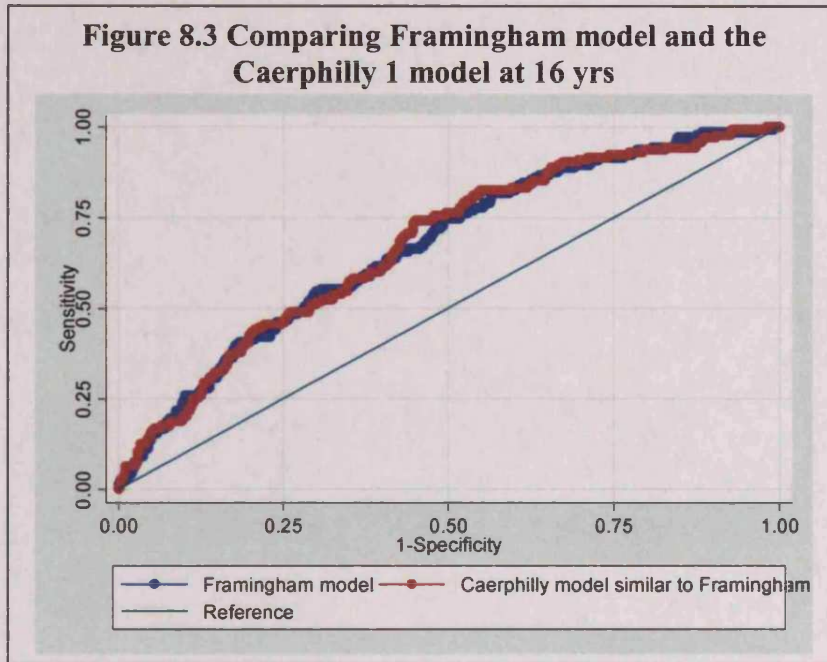
The ROC curve shows that the area under the curve for the Caerphilly 1 model is 0.6733 compared to the 0.6694 of the Framingham model. These models were not statistically significantly different from each other suggesting that the Framingham model is as good as the newly constructed Caerphilly model. Although the risk from the Framingham model is different from that for the Caerphilly 1 model, once an appropriate cutoff point is chosen for each model their performances as screening tools are very similar.

The risk factors that were considered in the above equation were identified as risk factors in the Framingham study. However, from the Caerphilly analysis, there are other highly significant risk factors such as leisure activity and anger out. These could be added to the present model to produce a different model, which might give results that are more accurate than the classical Framingham equations.

### 8.2.3 Caerphilly 2 risk equations

Anger out showed a significant association with the risk of MI along with leisure activity. In addition, attitude towards exercise was significant too but it was of marginal significance and was excluded from this analysis. Therefore, two factors (anger out and leisure activity) were added to the Caerphilly model to see if they show

any improvement in predicting events. This model will be called the Caerphilly 2 model.

Three models were fitted and the maximum log likelihood of each model is denoted by '$l$'.

Model 1: $u = \beta' X,$  $\log(\sigma) = \theta_0,$  $l = -1525.2$

Model 2: $u = \beta' X,$  $\log(\sigma) = \theta_0 + \theta_1 \mu,$  $l = -1525.1$

Model 3: $u = \beta' X,$  $\log(\sigma) = \theta_0 + \theta_1 \mu + \theta_2 \mu^2$  $l = -1525.1$

The following table shows the coefficients of the Caerphilly 2 model with the inclusion of leisure activity and anger out.

**Table 8.4: Comparing Framingham risk equations and Caerphilly 2 model for MI**

| Coefficients | New Caerphilly Model 1 | New Caerphilly Model 2 |
|---|---|---|
| $\theta_0$ | -0.2728 | -0.1822 |
| $\theta_1$ | --- | -0.0201 |
| $\beta_0$ | 18.4210 | 18.0245 |
| Log(SBP) | -1.4801 | -1.4418 |
| Log(age) | -1.3788 | -1.3086 |
| Log(total-C $\div$ HDL-C) | -0.6641 | -0.6416 |
| Smoking (Y/N) | -0.5523 | -0.5329 |
| Diabetes (Y/N) | -0.4198 | -0.4052 |
| ECG-LVH (Y/N) | -0.2071 | -0.1953 |
| Leisure (Low/ med & High) | 0.2597 | 0.2444 |
| Anger out (Y/N) | -0.5357 | -0.5069 |
| Maximised log likelihood | 1524.2 | 1524.1 |

Log likelihood ratio tests shows no improvement of either model 2 or model 3 over the LLAFT model (model 1) and so this simpler model was adopted.

The false positives at different probability cutoff values of for the Caerphilly 2 model are given below.

**Table 8.5 Caerphilly 2 model for MI**

| MI time/ Cut off | 8% | 10% | 15% |
|---|---|---|---|
| 10 yrs false positives | 359 | 306 | 209 |
| Actual men with events | 65 | 62 | 46 |
| Tot num of men at higher risk of MI at 10yrs | 424 | 368 | 255 |
| Sensitivity at 10 years | 0.68 | 0.65 | 0.48 |
| Specificity at 10 years | 0.67 | 0.72 | 0.81 |
| Predictive value at 10 yrs | 0.15 | 0.17 | 0.18 |
| 16 yrs false positives | 512 | 452 | 333 |
| Actual men with events | 111 | 105 | 88 |
| Tot num of men at higher risk of MI at 16yrs | 623 | 557 | 421 |
| Sensitivity at 16 years | 0.78 | 0.73 | 0.61 |
| Specificity at 16 years | 0.51 | 0.57 | 0.76 |
| Predictive value at 16 yrs | 0.18 | 0.19 | 0.21 |

It appears that choosing a cutoff of around 8-10% for 10 years and 8-12% for 16 years of follow-up gives approximately equal sensitivity and specificity. When the numbers of the false positives are compared to the Framingham model and the Caerphilly 1 model, there are fewer in this case. Furthermore, the predictive accuracy for these models is also higher than for the Caerphilly 1 and Framingham models. The ROC curves for the 10 and 16 years Caerphilly 2 model are given below.

**Figure 8.4a ROC Curve for the Caerphilly 2 model at 10 years**



Area under ROC curve = 0.7230
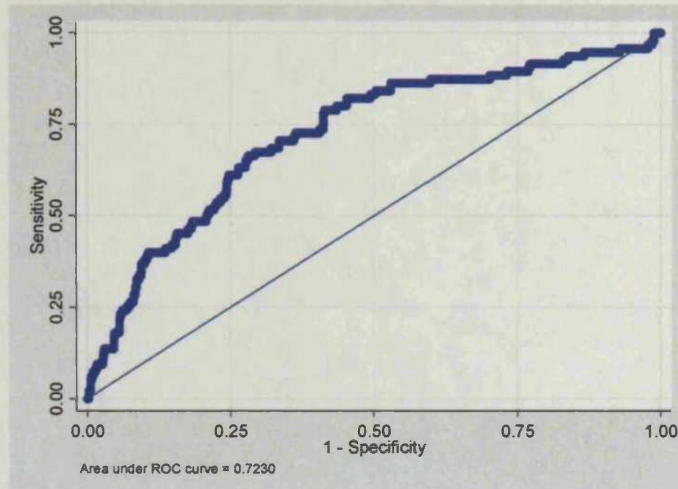
**Figure 8.4b ROC Curve for the Caerphilly 2 model at 16 years**



Area under ROC curve = 0.7032
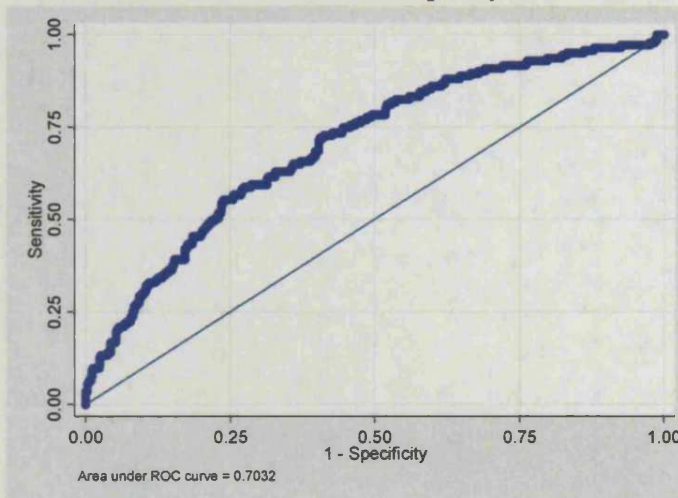
The areas under the curve for the Caerphilly 2 model at 10 and 16 years are 0.72 and 0.7 compared with 0.69 and 0.67 from Caerphilly 1. The equal sensitivity and specificity were approximately 0.68 and 0.64 for 10 and 16 year models with the corresponding probability cutoff values of 8.2% and 13% respectively. These high sensitivity and specificity values along with the bigger area under the curve might suggest that these models are better then the previous Caerphilly 1 and Framingham models.

In order to check, a ROC comparison between Framingham and the Caerphilly models at 16 years was carried out.

**Figure 8.5 Comparing Framingham model and the Caerphilly models at 16 years**
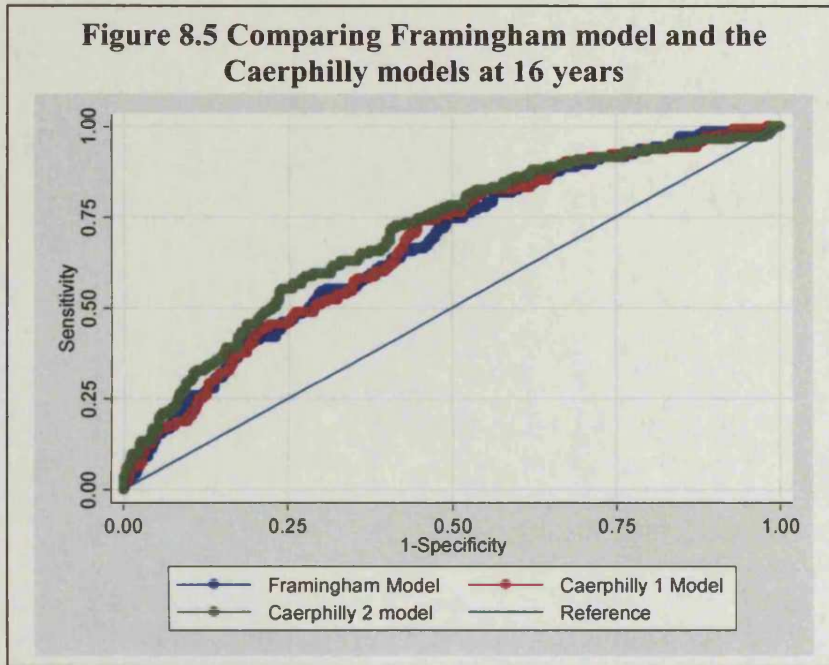
The ROC curve shows that the area under the curve (AUC) for the Caerphilly 2 model is 0.7032 compared to the Framingham model (AUC = 0.6733, p=0.016) and the Caerphilly 1 model (AUC = 0.6694, p=0.022). It shows that the Caerphilly 2 model is a better predictor of MI events than either the Framingham or Caerphilly 1 models..

In order to compare the Caerphilly risk models with the other studies like PROCAM [190], it is required to have similar data for the analysis in the CaPS model. However, few measurements like triglyceride, family history, and fibrinogen are measured in the CaPS and they are not analysed as part of this thesis. Therefore, further direct comparisons cannot be made.

## 8.3 Risk equations for Stroke

### 8.3.1 Framingham risk equation for stroke

The Framingham risk equation for stroke is given below.

$$\hat{\mu} = 26.5116 - 2.3741 \times (\log(age)) - 2.4643 \times (\log(SBP)) - 0.3914 \times (Cigs)$$
$$- 0.0229 \times (\log(Total\_Cholseterol \div HDL\_Cholesterol)) - 0.3087 \qquad .....(8.2)$$
$$\times diabetes - 0.2355 \times (ECG - LVH)$$

where $\log(\hat{\sigma}) = \theta_0 + \theta_1 \hat{\mu} = 3.4064 - 0.8584 \times \hat{\mu}$

if $\hat{u} = \dfrac{\log(followup - time) - \hat{\mu}}{\hat{\sigma}}$

then $p = 1 - \exp(-\exp(\hat{u}))$

Cigs defined in the above equation represents the smokers and non-smokers coded as a binary variable (1 & 0 respectively). Diabetes is also a binary variable in the equation where diabetic subjects were coded as 1 and the others as zero.

In the above equation, the probability $p$ is the predicted probability of having a stroke according to the Framingham model.

Out of the 1190 Caerphilly men analysed, there were 76 stroke events (6.41%) over a 16.1 years of follow-up. For 10 years of follow-up there were 47 events (3.95%).

The following table summarises the results for different probability cutoff points.

### Table 8.6 Framingham model for stroke

| Stroke time/ Cut off | 2% | 4% | 6% | 8% | 10% | 12% |
|---|---|---|---|---|---|---|
| 10 yrs false positives | 909 | 544 | 322 | 218 | 144 | 112 |
| Actual men with events | 44 | 35 | 26 | 18 | 15 | 10 |
| Tot num of men at higher risk of stroke at 10yrs | 953 | 579 | 348 | 236 | 159 | 122 |
| Sensitivity at 10 years | 0.94 | 0.75 | 0.55 | 0.38 | 0.32 | 0.21 |
| Specificity at 10 years | 0.20 | 0.52 | 0.72 | 0.81 | 0.87 | 0.90 |
| Predictive value at 10 yrs | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.08 |
| 16 yrs false positives | 1081 | 893 | 681 | 533 | 413 | 312 |
| Actual men with events | 76 | 71 | 64 | 51 | 41 | 36 |
| Tot num of men at higher risk of stroke at 16yrs | 1157 | 964 | 745 | 584 | 454 | 348 |
| Sensitivity at 16 years | 1.00 | 0.93 | 0.84 | 0.67 | 0.54 | 0.47 |
| Specificity at 16 years | 0.03 | 0.13 | 0.33 | 0.47 | 0.59 | 0.69 |
| Predictive value at 16 yrs | 0.07 | 0.07 | 0.09 | 0.09 | 0.09 | 0.10 |

The false positives in the above table decrease as the cutoff point of the probability increases for both the 10 and 16 year model. It appears that sensible cutoff points lie between 4-6% and 8-10% for 10 and 16 year models respectively if sensitivity and specificity are equal. The ROC curves for both models are given below.

**Figure 8.6a ROC Curve for the Framingham model at 10 years**



Area under ROC curve = 0.6913
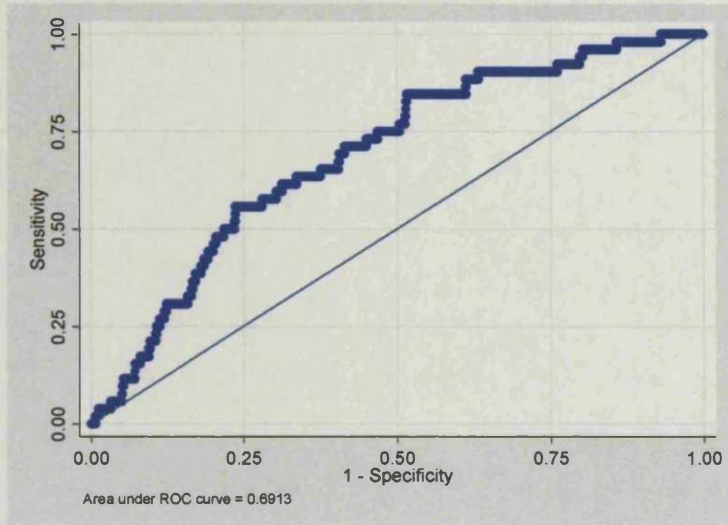
**Figure 8.6b ROC Curve for the Framingham model at 16 years**



Area under ROC curve = 0.6533
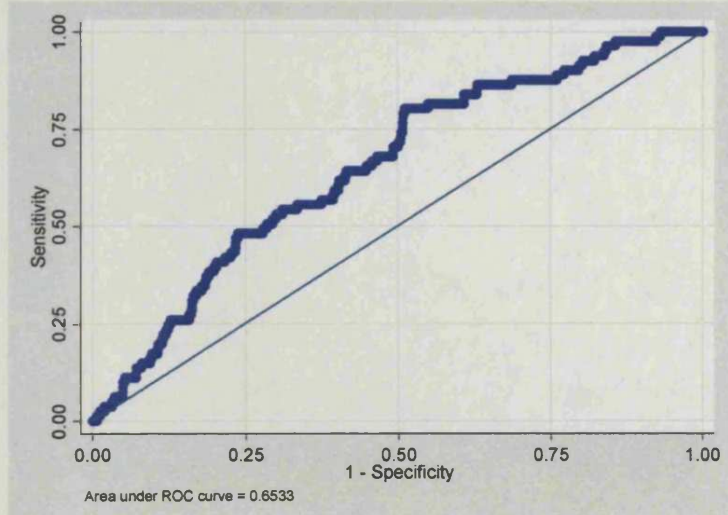
From the above plot, the areas under the curve with the Framingham model at 10 and 16 years are 0.69 and 0.65 respectviely. The corresponding probability cutoffs for the 10 year and the 16 year models are 5% and 9.6% with the common values of sensitivity and specificity close to 0.63 and 0.60 respectively.

## 8.3.2 Caerphilly 1 risk equations

Following the work of Section 8.2.2 a Framingham-type model was fitted to the Caerphilly data  The model did not converge with the presence of $\theta_1$ in the model even after various attempts of modifying the initial values. Therefore, the modelling was done without $\theta_1$ in the model. The new Caerphilly 1 model coefficients, compared with the Framingham model's coefficients, are given in the following table.

**Table 8.7: Comparing Framingham risk equations and Caerphilly equation for stroke**

| Coefficients | Framingham Coefficients | Caerphilly 1 Coefficients |
|---|---|---|
| $\theta_0$ | -0.4312 | -0.4245 |
| $B_0$ | 26.5116 | 26.0767 |
| Log(SBP) | -2.4643 | -0.9047 |
| Log(age) | -2.3741 | -4.2509 |
| Log(total-C ÷ HDL-C) | -0.0229 | 0.1232 |
| Smoking (Y/N) | -0.3914 | -0.1338 |
| Diabetes (Y/N) | -0.2627 | -1.0199 |
| ECG-LVH (Y/N) | -0.2355 | -0.1677 |

From the above table, it can be observed that the estimated $\theta_0$ and $\beta_0$ in Framingham and Caerphilly 1 risk equations are very similar. All the other risk factors show very different coefficients when compared. This could be because of the differences in the populations. It appears that diabetes does not contribute hugely to the Framingham model whereas it does in the Caerphilly 1 model. The logarithm of the ratio between total cholesterol and HDL cholesterol is positively associated with an increased risk in the Framingham model. It shows a decreased risk in the Caerphilly 1 model. The coefficient of log(SBP) is very different in the two models. All these suggest that the models are different.  The following table shows the summary statistics for the new model for a range of cutoffs.

## Table 8.8 Caerphilly 1 model for stroke

| Stroke time/ Cut off | 2% | 4% | 6% | 8% | 10% | 12% |
|---|---|---|---|---|---|---|
| 10 yrs false positives | 857 | 468 | 236 | 123 | 60 | 31 |
| Actual men with events | 45 | 36 | 25 | 15 | 7 | 5 |
| Tot num of men at higher risk of stroke at 10yrs | 902 | 504 | 261 | 138 | 67 | 36 |
| Sensitivity at 10 years | 0.96 | 0.77 | 0.49 | 0.32 | 0.15 | 0.11 |
| Specificity at 10 years | 0.25 | 0.59 | 0.79 | 0.89 | 0.95 | 0.97 |
| Predictive value at 10 yrs | 0.05 | 0.07 | 0.10 | 0.11 | 0.10 | 0.14 |
| 16 yrs false positives | 1100 | 843 | 607 | 454 | 333 | 222 |
| Actual men with events | 76 | 70 | 62 | 51 | 42 | 35 |
| Tot num of men at higher risk of stroke at 16yrs | 1176 | 913 | 669 | 505 | 375 | 257 |
| Sensitivity at 16 years | 1.00 | 0.92 | 0.82 | 0.67 | 0.55 | 0.46 |
| Specificity at 16 years | 0.01 | 0.24 | 0.45 | 0.59 | 0.70 | 0.80 |
| Predictive value at 16 yrs | 0.06 | 0.08 | 0.09 | 0.10 | 0.11 | 0.14 |

The Caerphilly model appears to have reduced the false positives compared to the Framingham model. From Table 8.6, it can be seen that the false positives at the 2% probability cutoff are similar to those in the Caerphilly model in Table 8.8. The number of predicted events at 10 and 16 years using this Framingham model are 53 and 104 respectively. This means, this model over-predicts 13% and 42% compared to the observed number of events

ROC curve analysis was performed to calculate the area under the curve for the Caerphilly 1 risk models at 10 and 16 years.

**Figure 8.7a ROC Curve for the Caerphilly 1 model at 10 years**



Area under ROC curve = 0.7380
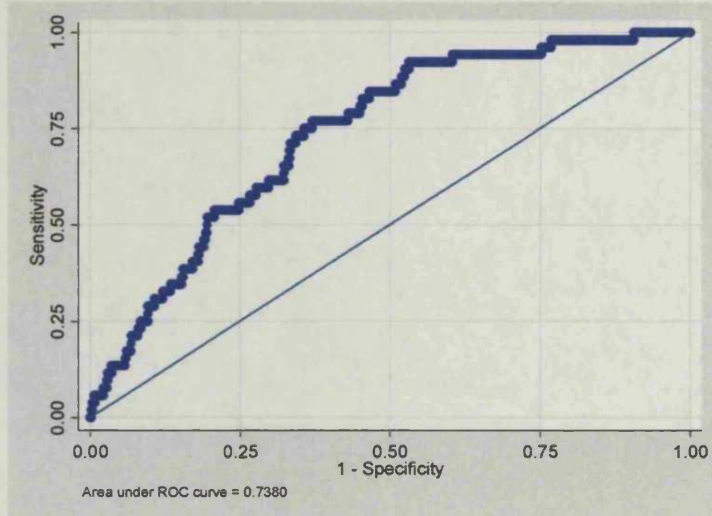
**Figure 8.7b ROC Curve for the Caerphilly 1 model at 16 years**



Area under ROC curve = 0.6908
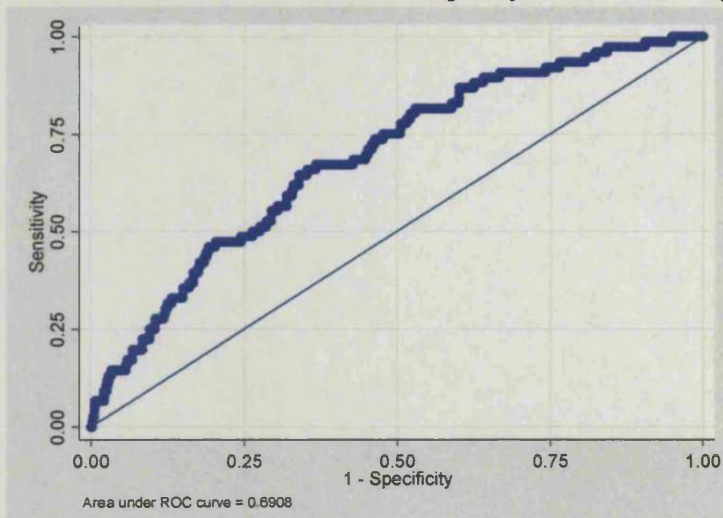
From the above plots, the areas under the curve for the Caerphilly 1 models at 10 and 16 years are 0.74 and 0.69. Compared to the Framingham model (0.69 & 0.65), they appear to be higher. Equal values of sensitivity and specificity were achieved for cutoffs of 0.05 and 0.09 respectively for 10 and 16 year models; the common values were 0.67 and 0.65. The sensitivity and specificity are slightly higher than for the Framingham model

However, superiority of this model cannot be concluded unless a statistical analysis is performed. The following plot shows the ROC curve comparing the Caerphilly model with the Framingham model.

**Figure 8.8 Comparing Framingham model and the Caerphilly 1 model at 16 years**
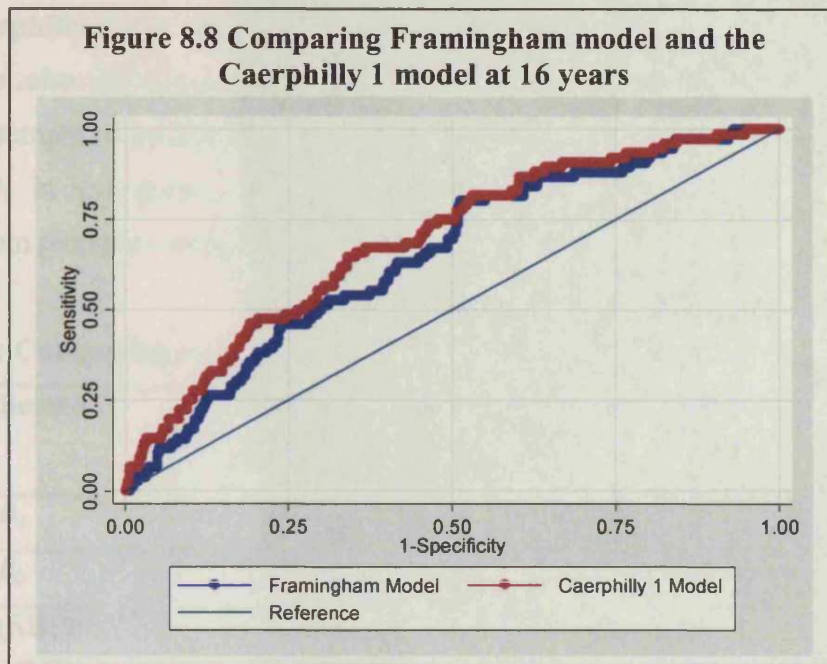
It appears that the Caerphilly 1 model is a better model then the Framingham model visually. The *p* value comparing both the models is 0.062 which is close to the 5% significance level. It is important to note that there are very few stroke events and these models are based on small numbers of events. Therefore, it is hard to conclude definitively whether or not the Caerphilly model is actually different in its predictive ability.

Several other risk factors that were analysed in the previous chapters, including the newly constructed anger-out scores, showed no significant association with the risk of stroke. Smoking status was significantly associated, however. It is important to note that the smoking variable used in the Framingham study was current cigarette smokers versus current non-smokers. However, the CaPS univariate analysis (section 4.2.2.2), showed that the risk for ex-smokers was similar to that of current smokers, even after quitting smoking for more than 10 years. Furthermore, in the univariate and multivariate analyses of alcohol consumption and stroke, alcohol consumption appeared significant (sections 4.1.4 and section 7.4). Therefore, it would be logical to derive a model with a smoking variable defined as never smoked versus ever smoked instead and also add alcohol consumption in the model. The newly constructed model is given below.

## 8.3.3 Caerphilly 2 risk equations

The model also did not converge with the presence of $\theta_1$ in the model even after various attempts of modifying the initial values. Therefore, the modelling was done without $\theta_1$ in the model. The new Caerphilly coefficients comparing with the Framingham model's coefficients are given in the following table.

**Table 8.9: Comparing Framingham risk model and Caerphilly models for stroke**

| Coefficients | Framingham Coefficients | Caerphilly Coefficients | Caerphilly 2 model |
|---|---|---|---|
| $\theta_0$ | -0.4312 | -0.4245 | -0.4221 |
| $\beta_0$ | 26.5116 | 26.0767 | 25.9563 |
| Log(SBP) | -2.4643 | -0.9047 | -0.9316 |
| Log(age) | -2.3741 | -4.2509 | -4.0912 |
| Log(total-C $\div$ HDL-C) | -0.0229 | 0.1232 | 0.1248 |
| Smoking (Y/N) | -0.3914 | -0.1338 | -0.4109* |
| Diabetes (Y/N) | -0.2627 | -1.0199 | -0.9801 |
| ECG-LVH (Y/N) | -0.2355 | -0.1677 | -0.1635 |
| Log (4+ALC) | - | - | -0.1312 |

**\*smoking is defined as never smoked vs rest.**

The new Caerphilly model differs from the other Caerphilly model in the way in which smoking is modelled and in the Caerphilly 2 model, the coefficient of smoking is quite different from that in the Caerphilly 1 model. The remaining coefficients did not change much, however. The sensitivity analysis is carried out for the new Caerphilly model to determine whether it is a better model or not.

## Table 8.10 Caerphilly 2 model for stroke

| Stroke time/ Cut off | 2% | 4% | 6% | 8% | 10% | 12% |
|---|---|---|---|---|---|---|
| 10 yrs false positives | 711 | 339 | 162 | 82 | 39 | 21 |
| Actual men with events | 45 | 28 | 18 | 13 | 7 | 5 |
| Tot num of men at higher risk of stroke at 10yrs | 756 | 367 | 180 | 95 | 46 | 26 |
| Sensitivity at 10 years | 0.96 | 0.60 | 0.38 | 0.28 | 0.15 | 0.11 |
| Specificity at 10 years | 0.38 | 0.70 | 0.86 | 0.93 | 0.97 | 0.98 |
| Predictive value at 10 yrs | 0.06 | 0.08 | 0.10 | 0.14 | 0.15 | 0.19 |
| 16 yrs false positives | 1033 | 696 | 485 | 329 | 230 | 155 |
| Actual men with events | 74 | 66 | 55 | 40 | 27 | 22 |
| Tot num of men at higher risk of stroke at 16yrs | 1107 | 762 | 540 | 369 | 257 | 177 |
| Sensitivity at 16 years | 0.97 | 0.89 | 0.72 | 0.53 | 0.36 | 0.29 |
| Specificity at 16 years | 0.07 | 0.37 | 0.56 | 0.70 | 0.79 | 0.86 |
| Predictive value at 16 yrs | 0.07 | 0.09 | 0.10 | 0.11 | 0.11 | 0.12 |

The false positives of this new Caerphilly model appear to be lower than for the previous Caerphilly 1 model and the Framingham model. The cutoff points for 10 and 16 year models lie between 2-4% and 6-8%, for 10 and 16 years, for equal sensitivity and specificity. The numbers of predicted events at 10 and 16 years using this Framingham model are 49 and 86 respectively. This means that the model over-predicts 4% and 12% respectively compared to the observed number of events. The ROC curve analysis is shown below.

**Figure 8.9a ROC Curve for the Caerphilly 2 model at 10 years**



Area under ROC curve = 0.7810
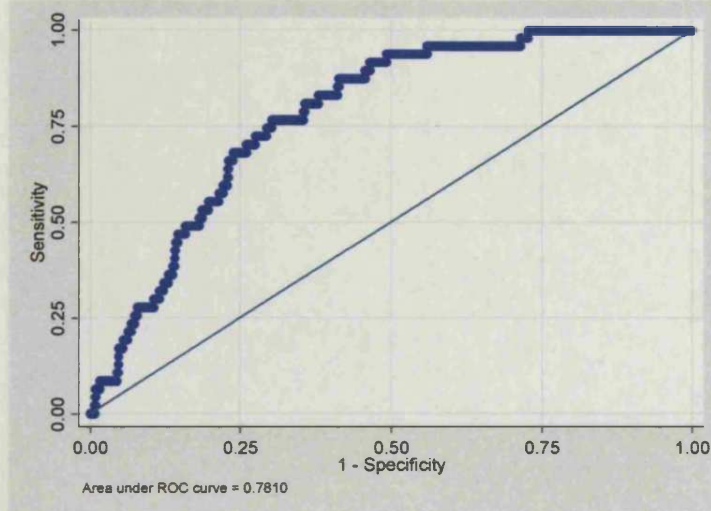
**Figure 8.9b ROC Curve for the Caerphilly 2 model at 16 years**



Area under ROC curve = 0.7095
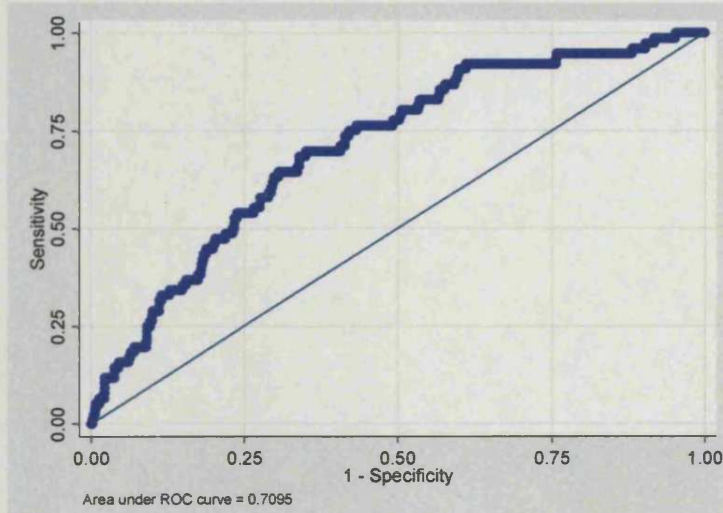
The areas under the curve for the Caerphilly 2 models at 10 and 16 years are 0.78 and 0.71. The 10 year Caerphilly 2 model appears to have a smaller area under the curve compared to the 10 year Caerphilly 1 model. It should also be noted that there were very few stroke events at 10 years. Therefore, the result might not be as powerful as the 16 years model. However, the 16 year Caerphilly 2 model has an area under the curve slightly more than the 16 year Caerphilly 1 model. The probability cutoffs at 10 and 16 years for suggested by the Caerphilly 2 model, when the common values of sensitivity and specificity were close to 0.74 and 0.69, were 0.04 and 0.07 respectively. These cutoffs appear to be lower then the Caerphilly 1 and Framingham models. A formal comparison of the ROC curves is performed below.

213

**Figure 8.8 Comparing Framingham model and the Caerphilly models for stroke**
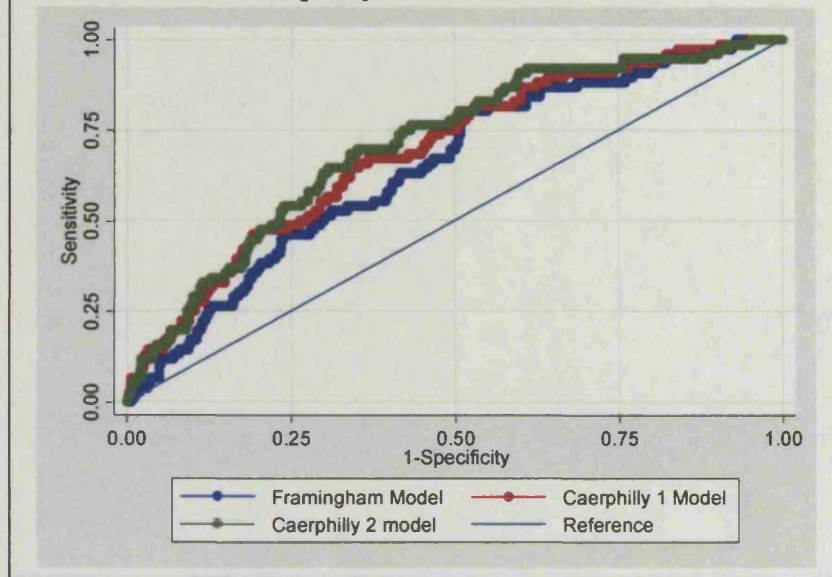
The above plot shows that the Caerphilly 2 model appears to have a larger area under the curve compared to the other two. When the Caerphilly 2 model is compared to the Caerphilly 1 model the p-value of 0.12 showed no conclusive evidence for a difference. However, when the Caerphilly 2 model was compared to the Framingham model, the $p = 0.013$ suggesting that the Caerphilly 2 model is significantly different from the Framingham model. This shows that a model based on the US population is not ideal for predicting the risk of stroke in the UK men.

## 8.4 Discussion

When the classic Framingham risk equations were applied to the Caerphilly cohort to predict the absolute risk for MI and stroke there were many false positives using a cutoff based on the actual prevalence level. Also, when the Caerphilly 1 models were estimated and compared to the Framingham model, they showed no improved fit for either MI or stroke.

However, when the new Caerphilly 2 model for MI was compared to the Framingham model and the Caerphilly 1 model, they were statistically different. In the case of stroke, the Caerphilly 2 model for stroke showed a statistically significant improvement over the Framingham model but not over the Caerphilly 1 model. This new model was constructed by changing the smoking categories from current smokers

versus current non-smokers to never smoked versus the rest. This also shows that the Framingham stroke model is not good for identifying Caerphilly men at higher risk.

The results from this analysis and the other studies indicate that the discrepancies between the false positives predicted by different studies for the risk of MI or stroke have important implications for the primary prevention of the disease. A recent publication by Andrew Beswick et al [28] showed that the various versions of Framingham-type risk equations that are available are not consistent between studies. This analysis actually addresses this basic problem by comparing new Caerphilly models with classic Framingham models. It compared the false positives for each model at different cutoff probabilities. It appears that the Caerphilly 2 models for MI and stroke both predict MI and stroke with fewer false positives at the actual prevalence cutoff point (12% for MI and 6.4% for stroke).

These results must be carefully interpreted because the Caerphilly 1 and 2 models have been tested on the data set on which they were derived, and this is likely to give biased estimates of sensitivity and specificity.

There are several ways to overcome this problem. One of those is to find a different dataset and apply these new Caerphilly equations and the classic Framingham equations to find the difference between their predictions. It might be possible to find a data set with the classical risk factors that were used in the Framingham equation and Caerphilly equation because there are several studies that measured these risk factors but none have included the behavioural and psychological variables.

The other way around this problem is to randomly divide the dataset into two halves, estimate the equations from one half and apply them on the other. After excluding missing values from the Caerphilly data, there were only 1190 men with complete data to estimate these equations. Therefore, the Caerphilly data was split into a 70-30 ratio, the equations were estimated from 70%, and those estimated equations were applied to the other 30%. This exercise did not found a significant difference between the Framingham risk equations and the new Caerphilly equations for either MI or stroke. This might mean that there is actually no improvement with the addition of

psychological factors in the model for predicting MI, but it is very likely that the reduction in the sample size from this split has led to a loss of power.

# Chapter 9

## 9.0 Interventions

### 9.1 Introduction

The risk of MI and stroke was analysed univariately for each individual risk factor in chapters 4-6 followed by a multivariate analysis in chapters 7 and 8. Multivariate analysis discussed the importance of psychological factors in the presence of classical risk factors. The new multivariate Framingham type risk equations were estimated and compared with the Framingham models in the previous chapter. After understanding the effect of various risk factors on the VAD risk, the next step is to identify plausible interventions and assess the effect of the interventions on the risk of VAD which is the main aim of the thesis.

The plan is to achieve this by simulations as it is difficult, expensive and unethical to carry out large scale trials to assess the impact of these interventions. For this purpose, there is a need to know how these interventions might impact on the levels of risk factors. Furthermore, there is a need for a model that would translate the change in the risk factors into an estimate of the number of events saved.

The models developed in the previous chapter will be used to assess how the risk changes through intervention on the risk factors. Therefore, what is needed is to build simulation models which will simulate the effect of realistic interventions, firstly on the levels of the risk factors and then, through the risk models developed in Chapter 8, on the risk itself.

The first objective of this chapter is to develop a model for generating a population with appropriate levels of risk factors. For each relevant risk factor, a literature review on interventions will be carried out. Models for the effect of interventions on each individual risk factor will be built. The effects of these interventions on the level of each risk factor are simulated followed by an estimate of the number of events saved if these interventions were carried out on such populations.

## 9.2 Statistical methods in building simulation models

One of the objectives of this chapter is to identify plausible interventions and simulate intervention scenarios in order to predict their effect. For simulating the interventions, there is a need to simulate populations with realistic levels of various risk factors. Data from the Caerphilly study has been used to understand the nature of the different risk factors and that information is used to simulate the hypothetical population, using a computer program written in C++. The following statistical techniques were used to simulate various datasets.

Logarithms of SBP, DBP, total cholesterol, HDL-C and BMI follow normal distribution more closely then the remaining other data which is evident with the summary statistics described at the start of the analysis of each risk factor in chapters 4 and 5. So these normally distributed data should be generated for the hypothetical populations. Alcohol consumption appears to be is a case of Gamma distribution from the Figure 4.1. Smoking, leisure activity and all psychological variables are discrete and they will have to be estimated using a specific discreet distribution. All these should be simulated.

### 9.2.1 Normal distribution

A random variable X with a probability function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \qquad -\infty < x < +\infty$$

where $\sigma$ is positive, is said to have a normal distribution with parameters $\mu$ and $\sigma$. The expectation and variance of X are $\mu$ and $\sigma^2$ respectively. If a random variable X follows a normal distribution with mean $\mu$ and variance $\sigma^2$, then the random variable Z, defined as

$$Z = \frac{x - \mu}{\sigma}$$

follows a standard normal distribution with mean 0 and variance 1.

In order to generate samples from a normal distribution with parameters μ and $\sigma$, the central limit theorem is employed. This states that if $X_1$, $X_2$, ......... $,X_n$ are $n$ independent and identically distributed random variates with mean μ and SD σ, then the sum $\sum X_i = X_1 + X_2 + ......... + X_n$ approaches a normal distribution as $n$ becomes large. The mean and the variance of this normal distribution are

$$E(\sum X_i) = n\mu$$

$$V(\sum X_i) = n\sigma^2 .$$

The procedure for generating normal variates requires $k$ random numbers $r_1, r_2, ....., r_k$ generated from U(0, 1) Since each $r_i$ is a uniformly distributed random number over a interval [0, 1], we have that

$$E(r_i) = \frac{a+b}{2} = \frac{1}{2} \quad \text{and}$$

$$SD(r_i) = \sqrt{\frac{(b-a)^2}{12}} = \frac{1}{\sqrt{12}} .$$

Using the central limit theorem, the sum of those $k$ random numbers approaches the normal distribution. That is

$$\sum r_i \sim N\left(\frac{k}{2}, \frac{k}{\sqrt{12}}\right),$$

or

$$\frac{\sum r_i - \frac{k}{2}}{\frac{k}{\sqrt{12}}} \sim N(0,1) \text{--------- (1)}$$

Now, consider a normal distribution X with parameters μ and $\sigma$ from which the normal variates are to be generated. Then

$$\frac{X - \mu}{\sigma} \sim N(0,1) \text{------------(2)}$$

Equating (1) and (2) gives

$$\frac{X-\mu}{\sigma} = \frac{\sum r_i - \dfrac{k}{2}}{\dfrac{k}{\sqrt{12}}} \quad ,$$

or

$$X = \sigma\sqrt{\frac{12}{k}}\left(\sum r_i - \frac{k}{2}\right) + \mu \quad .$$

This equation provides a simple formula for generating approximately normal variates with mean $\mu$ and standard deviation $\sigma$. The value of $k$ has to be large, since the larger it is, the closer to a normal distribution. If $k = 12$, the formula has computational advantages and

$$X = \sigma\left(\sum_{1}^{12} r_i - 6\right) + \mu \quad .$$

For example, log(SBP) has an approximately Normal distribution with mean and standard deviation of 4.98 and 0.153 respectively. These values are used in the above equation to generate values of log(SBP) which are then transformed into simulated values of SBP.

### 9.2.3 Gamma Distribution

The gamma distribution with parameters $\sigma > 0$ and $\beta > 0$ has the probability density function

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} \qquad \text{for } 0 \leq x \leq \infty \quad ,$$

$$\text{where } \Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1} e^{-x} dx$$

$\Gamma(\alpha)$ is the complete gamma function. The parameter $\alpha$ is called the shape parameter, and $\beta$ is called the scale parameter. If a random variable $Z$ has a standard gamma distribution with shape parameter $\alpha$ and scale parameter 1, and $X=\beta Z$, then $X$ has a

gamma distribution with $\alpha$ and $\beta$. This means that a method for generating sample values with $\beta=1$ can be extended easily to other values of $\beta$.

Of the distributions that were considered thus far, this is the first distribution that has a parameter that cannot be handled by a simple transformation. The best algorithm for the gamma distribution may be different depending on the value of $\alpha$. If $\alpha = 1$ then a gamma distribution is an exponential distribution.

Cheng and Feast in 1979 [211] used a ratio of uniforms method for a gamma distribution with $\alpha>1$. Although this algorithm is not efficient compared to the one developed afterwards, this would solve the present purpose. The algorithm is discussed below.

### 9.2.3.1 Cheng and Fest algorithm for generating Gamma random variates when shape parameter is greater than 1.

The steps of the algorithm are given below.

1) Generate $u_1$ and $u_2$ independently from U(0,1) and set

$$v = \frac{\left(\alpha - \dfrac{1}{6\alpha}\right)u_1}{(\alpha - 1)u_2}.$$

2) If

$$\frac{2(u_2 - 1)}{\alpha - 1} + v + \frac{1}{v} \le 2, \text{ or if} \frac{2\log u_2}{\alpha - 1} - \log v + +v \le 1,$$

then the random variate $x = (\alpha\text{-}1)v$.

3) Else go to step 1.

An efficient algorithm for values of the shape parameter less than 1 is the acceptance/rejection method described in Ahrens et al [212] and modified by Best et al [213]. The general method for simulating random numbers of particular distributions using the rejection method is illustrated with an example followed by the steps in Best et al [213] algorithm.

### 9.2.3.2 The Best/Ahrens/Dieter algorithm for generating Gamma random variates when shape parameter is less than 1.

1) Initiate $t = 0.07 + 0.75\sqrt{1-\alpha}$ and $b = 1 + \dfrac{e^{-t}\alpha}{t}$.

2) Generate $u_1$ and $u_2$ independently from U(0,1) and set $v=bu_1$.

3) If $v \leq 1$, then

Set $x = tv^{\frac{1}{\alpha}}$

If $u_2 \leq \dfrac{2-x}{2+x}$ or $u_2 \leq e^{-x}$,

then $x$ is a sample value from the gamma distribution.

otherwise,

Set $x = -\log\left(\dfrac{t(b-v)}{\alpha}\right)$ and $y = \dfrac{x}{t}$.

If $u_2(\alpha + y(1-\alpha)) \leq 1$ or $u_2 \leq y^{\alpha-1}$, then $x$ is the sample value from the gamma distribution.

Otherwise go to step 2.

Alcohol consumption follows the gamma distribution with a shape and scale parameters of 0.676 and 0.004 and therefore the above method is used to simulate that data. Figure 9.2 shows Q-Q plot the observed and expected values of alcohol consumption data.

**Figure 9.2: Gamma Q-Q Plot of alcohol consumption (cc/week)**
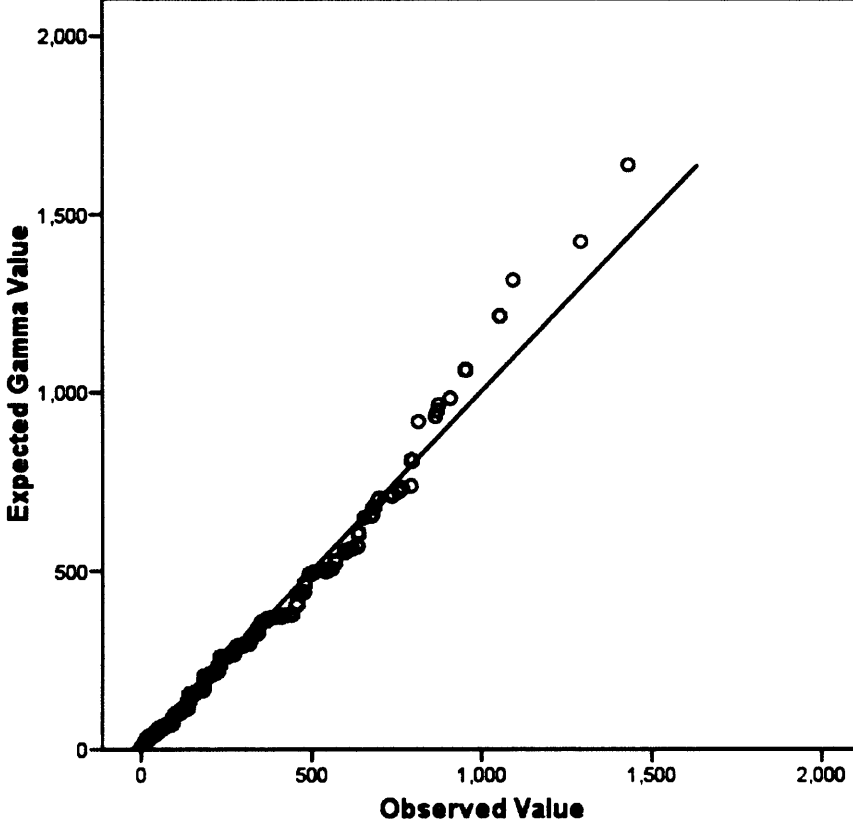
## 9.2.4 Multivariate distribution

In principle the different risk factors for VAD are correlated and therefore there is a need to generate samples for each risk factor including the correlation effects with the other factors. However, the information used to generate the risk factor population are based on the Caerphilly data which shows very low correlations between the risk factors. The correlation coefficients of risk factors with others are given in the following table.

**Table 9.1**

|  | SBP | TCHOL | LOGALC | BMI | Smoking | Leisure | Latent |
|---|---|---|---|---|---|---|---|
| SBP | 1 | | | | | | |
| TCHOL | 0.0515 | 1 | | | | | |
| LOGALC | 0.0583 | 0.0483 | 1 | | | | |
| BMI | 0.1694 | 0.1051 | 0.0149 | 1 | | | |
| Smoking | -0.0017 | -0.0155 | 0.0802 | -0.2181 | 1 | | |
| Leisure act | -0.0565 | 0.0194 | 0.0525 | -0.0549 | 0.0286 | 1 | |
| Latent ang | 0.0617 | -0.0292 | -0.0324 | -0.0152 | 0.0632 | -0.0564 | 1 |

It can be observed that the correlation coefficients of any given factor with other factors are not more than 0.22. Since the Caerphilly data does not provide high correlations between the factors, it was therefore decided to simulate samples individually assuming that there is no correlation between the risk factors. It is important to note that it is easy to generate correlated normal distributions but that it would be harder for some of the other factors as they are either discrete or non-normal distributions.

## 9.2.5 Survival time simulation

Having generated values of the risk factors, the next step is to generate survival times. The distribution of survival times depends on the levels of the risk factors and so a model must be specified for relating survival times to the risk factors. The Weibull model is used in this case as it also includes the case of the exponential distribution (the exponential is a particular case of a Weibull model) and is more robust. The hazard function of an exponential distribution is constant whereas, the hazard function of the Weibull model is very flexible.

The coefficients for the factors that are required for the simulation were taken from the estimated models by STATA. Suppose that $T$ is a survival time, $\beta'$ is the vector of coefficients to the vector of $X$ factors. Then the time could be simulated using the following equation.

$$T = \left( \frac{-\log(U)}{\lambda \exp(\beta' X)} \right)^{1/v}$$

where U is a U(0, 1) random variable, $\lambda = \exp(-\beta_0)$ and $v$ is the reciprocal of the shape parameter estimated in the model. The derivation and the efficiency of this method is discussed by Ralf Bender et al [214].

With this method, the survival time of the hypothetical subjects are simulated. However it is important to note that these survival times are uncensored and correspond to men who do not dropout and are followed until an event occurs. For simulating the events that will occur in a population, these times are appropriate. However, the simulation model needs to be validated against the observed events in the Caerphilly study and a considerable number of the men dropped out of the study before the first MI or stroke event, for a variety of reasons.

To perform the validation, therefore, this dropout process must be simulated and so the dropout process must be investigated. Figure 9.2 shows the times until dropout of

the men who did not have an event and who left the study before the end of year 2000. This can be approximated by a uniform distribution for the purposes of the validation.

**Figure 9.2 Survival times for those who had follow-up times before the end of year 2000**



MI survival times for men with no event and lost follow-up
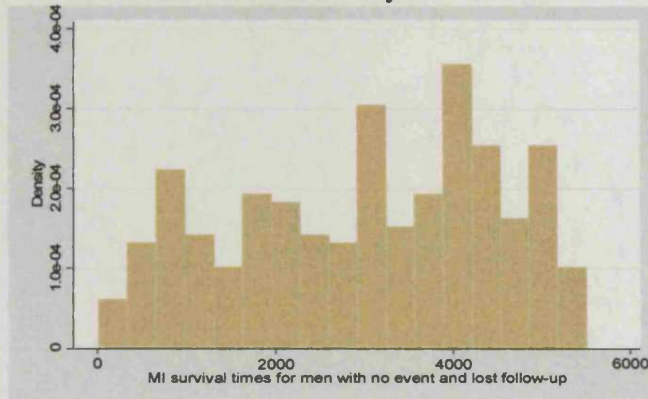
## 9.3 Simulations

The purpose of modelling a hypothetical intervention is to assess its effect on reducing the risk of VAD on simulated subjects, through changes arising in the levels of the risk factors. In order to do this a model must be created which specifies how the intervention will be applied and to whom, what will be the effect of it on risk factors, and how this will change the risk of subjects in receipt of the intervention.

Of importance is to know whom to target i.e., high risk groups could be identified on the whole population targeted. It is also important to know how many people targeted with the intervention will actually take part in the intervention, since undoubtedly some will decline in real life.

In a simulation program, several possibilities of risk factor modification can be achieved but that does not mean that they are plausible. For example, reducing SBP to 120 mmHg for every person is possible in a simulation program but is it possible to reduce blood pressure in such fashion? Certainly not. Therefore, assumptions of the impact of hypothetical interventions should be realistic. In this simulation program, it is assumed that the reduction in the level of a risk factor follows a distribution. For example, assume that an intervention to reduce SBP is carried out aiming to get a reduction of SBP by 20 mmHg. It is not practically possible for the intervention to reduce exactly 20 mmHg in every intervened subject. Some subjects have a higher reduction and some lower. Therefore, it is sensible to assume that reduction follows a statistical distribution for example by assuming that SBP is reduced by 20 mmHg with a standard deviation of 3 or 5.

In a realistic scenario, the risk of MI or stroke does not decrease immediately after the intervention starts. It takes some time for the risk to show an actual reduction after the intervention on the risk factor. For example, suppose a person with an SBP of 180 mm Hg undergoes an intervention to reduce his blood pressure and it actually reduces it by, say, 40 mmHg. Anti-hypertensive drugs act very quickly and this change may occur over a short time period. If the models of Chapter 8 are applied directly then the risk would also change immediately to the level appropriate to a person whose long term SBP was 140. This is probably unrealistic and in practice it may take a

considerable time for the risk to change to its new level. This can be incorporated easily into a simulation model. It should also be noted that the time taken for the risk to reduce after an intervention might be different for each risk factor i.e., the risk might take one year to reduce after a person has an intervention for reduction of SBP but it might take several years for smokers when they quit smoking.

It would be naïve to assume that an intervention does not have side effects. There are bound to be some side effects and subjects who have side effects may either change their medication or quit the intervention program. This must also be incorporated in the simulation.

When an intervention is made to change the value of one risk factor, it may well have an effect on other risk factors. For example an intervention to reduce alcohol consumption may affect other risk factors. This might occur through a direct biological mechanism or by making the subject more aware of the importance of healthy living. In the Caerphilly data, all the risk factors showed very little correlations and therefore, multivariate effects are not considered in this simulation program. This is one of the limitations of the simulation.

Finally, it is important to validate the results of an intervention scenario. For this purpose, there is a need to run a model without an intervention and one with an intervention to estimate the benefits of the intervention.

There could be several intervention scenarios for any risk factor and simulating all possible interventions requires time. Therefore, a brief literature review of interventions on each risk factor is discussed with one or two plausible intervention scenarios, using a flow chart if necessary. However, the following flow chart is a generic one that is common in the various intervention scenarios that could actually be applied to most of the risk factors.

**Figure 9.3: General flowchart of an intervention scenario**
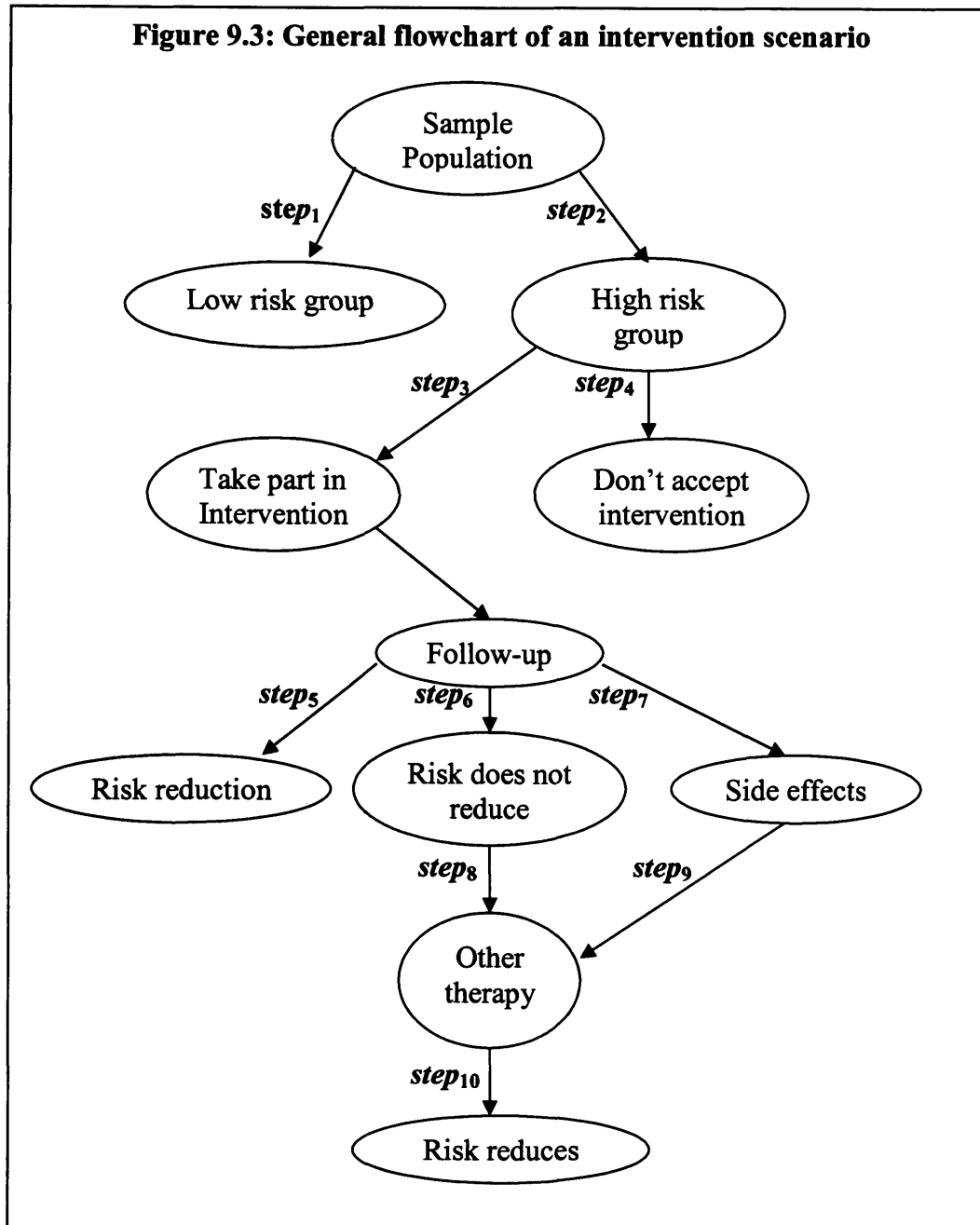
The percentage of the subjects in various steps of the intervention depends on the type of risk factor and intervention. For example, for steps 1 and 2, there could be 40% smokers defined as a high risk group and the rest as a low risk group. In the case of SBP, these percentages could be 30% and 70%. Therefore, there will be several possibilities that could be explored to check the number of events saved by different possibilities at each step. Furthermore, the time taken for the reduction in the risk of MI or stroke to take effect after changing the levels of risk factors in the simulated subjects is also varied. That is, for example, the time taken for the risk of MI to reduce

when they stop smoking may vary between 0-20 years. This is also considered in these simulation models.

The Weibull model will be used to simulate the survival times. Survival times are generated and hence the events in the different time periods can be calculated. Making an intervention on the same subjects by changing the risk factor levels will give a different number of events. Therefore, there will be two set of events, the first where there is no intervention and the second when there is an intervention. Then there will be a comparison between the numbers of events in the simulated sample after the intervention with the number of events without an intervention.

Survival times will be estimated by each individual risk factor separately. A separate model is used for each simulation. After concentrating on individual risk factors, the focus will be changed to multivariate interventions where all the risk factors are used to simulate the survival times and the results will be discussed. Therefore, the number of MI or stroke events estimated by a simulation model for each risk factor may vary depending upon the strength of its association with the outcome. To make the simulation results more robust, a sample of 10000 subjects is generated and repeated 1000 times. The difference between the number with and without an intervention is given with the mean and SD. Models used for each intervention scenario are in the appendix 4.

The above description has concentrated on interventions aimed at reducing the number of MI events, but the same procedure was also used to investigate the effect of interventions to reduce the incidence of stroke.

## 9.3.1 Alcohol consumption

Meta analysis suggests that acute alcohol consumption is a risk factor for heart disease but also that the relationship of risk to alcohol consumption is J-shaped [66]. That Meta analysis had 116,702 subjects from 61 different studies whereas; the Caerphilly study has less than 2000. Though the actual alcohol consumption data showed a 'J' shaped relationship, the results were not significant. Because the relationship was not significant, there is no point in doing the simulation

However, alcohol consumption is a risk factor for stroke and therefore a literature review on the effects of interventions is performed. There are relatively few studies showing the effects of reducing high alcohol intake (Table 9.2). Saltz et al. published a paper which showed evidence that primary intervention by the physicians reduced alcohol consumption for subjects with high alcohol intake [215]. Fleming et al. showed that a brief physician advice for heavy drinkers proved to be effective in reducing alcohol consumption [216]. A comparison of two intensities of psychological intervention for alcohol dependent patients did not show a difference in the alcohol consumption of the two different intensities of the treatments [217] . A web based intervention for people with excessive alcohol drinking proved that it is cost effective and feasible [218]. Most of the literature review on interventions to reduce acute alcohol consumption suggests that interventions are effective and economically                                                                cheap.

**Table 9.2: Intervention studies on the effect of reducing heavy alcohol intake on vascular risk**

| STUDY / OR AIM | SAMPLE SIZE and measuring units. | Results | Variables analysed, outcome variable. | AGE GROUP, LENGTH OF FOLLOW-UP | CONCLUSION |
|---|---|---|---|---|---|
| [215] A Cluster Randomized, Controlled Trial of a Systems Intervention. (2003) | 312 patients with hazardous drinking. | 11.6 [CI, 5.4 to 17.7] units versus 3.8 [CI, 1.9 to 5.7] units of alcohol consumption per day compared before intervention and after. | Levels of alcohol consumption. | Age group not given. 6 months follow-up. | Physicians prompting the heavy drinkers to reduce alcohol consumption may reduce the alcohol intake. |
| [216] Brief Physician Advice for Problem Drinkers: Long-Term Efficacy and Benefit-Cost Analysis. (2002) | 482 men, 292 women were randomly assigned to control and intervention | Subjects in the treatment group exhibited significant reductions ($p - 0.01$). Results were not quantified. | Levels of alcohol consumption. | 18-65, 48 months of follow-up | Long term follow-up shows good results and shows that the brief intervention is cost effective as well. |
| [217] Long-term effect of brief intervention in at-risk alcohol drinkers: a 9-year follow-up study. (2004) | 338 men at the induction time and 247 men after 9 years of follow-up. | Fisher's two-tailed $P = 0.158$, chi-squared $P = 0.137$ comparing before and after alcohol intervention. Results were not quantified. | Levels of alcohol consumption. | Age group not given. 9 years of follow-up. | After 1 year there was a 50% reduction in alcohol intake and after 9 years, results were not significant. |

232

From the above brief literature review on intervening on subjects with high alcohol consumption, it appears possible that a brief intervention to reduce alcohol consumption could be effective for a relatively short period but that after a few years, a subject's alcohol consumption reverts to its previous levels. Nevertheless, it could be observed that the reduction in alcohol consumption appears to be significant at about one year after the intervention and goes back to normal after ten years.

From the literature review, 10cc or 8gms of pure alcohol is considered as one unit of alcohol and 3-4 units per day is recommended [219]. If at most 25 units of alcohol per week are deemed to be acceptable an consumption, it means that those who have more than 250 cc/week should be targeted for alcohol intervention. Two hundred and nineteen men have alcohol consumption more than 250 cc/week in the Caerphilly data (18.4%). The alcohol consumption analysis for stroke showed that the risk starts increasing for men as alcohol consumption increases. Therefore, men with alcohol consumption more than 250 cc/week can be hypothetically intervened. Reducing the alcohol consumption in men with more than 250 cc/week might be a practically possible scenario. Given that 'any' alcohol consumption increases the risk of stroke (this was the case in the Caerphilly analysis, section 4.14), it would be interesting to see what happens if all the men other that teetotallers are hypothetically intervened to reduce alcohol consumption.

Suppose that the actual risk reduction following a reduction in alcohol consumption would take effect after one year of intervention. Furthermore, it was supposed that the reduction in alcohol consumption is 50% on average for all intervened subjects. That means there would be men with a reduction of only 25% and some with 75% but the average reduction in all men is 50%. This information is based from the literature review [217].

The various parameters that are specified in the simulation model are as follows.

- Column 1 – scenario number
- Column 2 – time take for the intervention to take effect.
- Column 3 – condition used to identify subjects at low risk (no intervention is applied on them)
- Column 4 – condition used to identify subjects at high risk
- Column 5 – intervention applied.
- Column 6 – percentage of high risk men intervened.
- Column 7 – Mean and SD of events with no intervention
- Column 8 – Mean and SD of events after 16 years of intervention
- Column 9 – Mean and SD of saved events after 5 years of intervention
- Column 10 – Mean and SD of saved events after 10 years of intervention
- Column 11 – Mean and SD of saved events after 16 years of intervention

On the basis of this information, Table 9.3 shows the different possibilities that could happen, depending on the values of the various parameters of the simulation. The flowchart described in the Figure 9.3 shows the general intervention scenario for any risk factor. For alcohol consumption, there is a possibility that subjects does not accept an intervention to reduce consumption and this is incorporated in the simulation model.

**Table 9.3 Hypothetical Alcohol consumption intervention scenarios for reducing stroke events**

| Scenario number ↓ | Time taken for change | STEP1 (No risk group) | STEP 2 (High risk group) | Intervention | Percentage of intervention | Mean(SD) before intervention | Mean(SD) after intervention (16 years) | Mean(SD) of events saved 5 years | Mean(SD) of events saved 10 years | Mean(SD) of events saved |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 year | ALC≤250 cc/week | ALC>250 cc/week | Reduce ALC by 50% on average | 100% | 689(25.4) | 679(25.1) | -1(17.5) | 3(26.4) | 10(34.5) |
| 2 | 1 year | ALC≤250 cc/week | ALC>250 cc/week | Reduce ALC by 100% | 100% | 690(26.0) | 670(24.8) | 5(16.5) | 14(26.8) | 20(34.7) |
| 3 | 1 year | ALC=0 | ALC>0 cc/week | Reduce ALC by 50% on average | 100% | 690(23.5) | 658(24.0) | -1(17.4) | 18(26.2) | 32(35.2) |
| 4 | 1 year | ALC≤250 cc/week | ALC>250 cc/week | Reduce ALC by 50% on average | 50% | 688(24.6) | 683(24.8) | -1(16.1) | 2(25.9) | 5(34.6) |
| 5 | 2 years | ALC≤250 cc/week | ALC>250 cc/week | Reduce ALC by 50% on average | 100% | 690(25.6) | 681(24.4) | -7(16.8) | -1(26.2) | 9(35.8) |
| 6 | 2 years | ALC≤250 cc/week | ALC>250 cc/week | Reduce ALC by 100% | 100% | 689(25.1) | 671(23.8) | 0(16.8) | 10(26.6) | 18(33.2) |
| 7 | 2 years | ALC=0 | ALC>0 cc/week | Reduce ALC by 50% on average | 100% | 688(25.6) | 652(25.0) | 0(18.7) | 21(28.3) | 36(35.6) |
| 8 | 2 years | ALC≤250 cc/week | ALC>250 cc/week | Reduce ALC by 50% on average | 50% | 687(26.5) | 682(25.7) | -3(16.9) | 2(26.3) | 5(34.2) |

For every 10,000 men approximately 689 stroke events occurred in the hypothetical population in 16 years with no intervention. Compared to the event rates in Caerphilly (6.4% after 16 years), the simulated stroke events are almost equal (6.9% from the simulation).

In the eight scenarios that are shown in the above table, the most plausible intervention scenario appears to be the fourth one. It can be seen that the maximum number of events that could be saved by making everyone quit drinking is only 36, which is about 5%. Even with an unlikely intervention scenario, it appears that the intervention on reducing alcohol consumption seems to have very little effect on reducing cases of stroke. There might be other benefits of reduced drinking, but reducing alcohol consumption to reduce the risk of stroke or MI does not seem to be an effective way of allocating resources.

The uncertainty in these estimates can be calculated using the standard deviations. Since the standard deviations were calculated using the 1000 simulations so a confidence interval for the mean number of events saved after 16 years, say, the 3$^{rd}$ scenario is given by

$$32 \pm 1.96 \times \frac{35.2}{\sqrt{1000}} \approx 32 \pm 2.2$$

Therefore, the confidence interval for the number of events saved after 16 years of the 3$^{rd}$ intervention is (29.8, 34.2).

## 9.3.2 Blood pressure

The literature review and the analysis of the CaPs data showed that high blood pressure is strongly associated with heart disease. Reducing high blood pressure reduces the risk of heart attack, but selecting the right blood pressure medication can be a challenge. There are many available, but not all of them work equally well in all people. A literature review is needed to find out what other different studies have found.

Calcium channel blockers are a class of drugs that block the entry of calcium into the muscle cells of the heart and the arteries. It is the entry of calcium into these cells that causes the heart to contract and arteries to narrow. By blocking the entry of calcium, CCBs decrease contraction of the heart and widen the arteries. In order to pump blood, the heart needs oxygen. The harder the heart works, the more oxygen it requires. Heart pain occurs when the supply of oxygen to the heart is inadequate for work the heart must do. By dilating the arteries, CCBs reduce the pressure in the arteries. This makes it easier for the heart to pump blood, and, as a result, the heart needs less oxygen. By reducing the heart's need for oxygen, CCBs relieve or prevent heart disease. CCBs are also used for treating high blood pressure because of their blood pressure-lowering effects. CCBs also slow the rate at which the heart beats and are therefore used for treating certain types of abnormal heart rhythms.

The most common side effects with CCBs are peripheral edema, flushing, and headache. High dosage of CCBs results in an increase of 4% on average of total cholesterol [220]. Despite the potential for side effects with CCBs, their potent blood pressure-lowering effect makes them a prerequisite for blood pressure control in many patients [221]. There could be 7% to 9.6% of men having side effects with CCBs, however [222].

A meta analysis paper suggested that the new Calcium Channel Blockers (CCBs) reduce the risk of any cardiovascular event more than the conventional drugs [223]. The mean SBP lowered with these new drugs ranges from 2.1 to 22.1mmHg in 7 trials depending upon the dose of the drug used to lower blood pressure [224]. All of the

interventions suggested anti hypertensive drugs reduce a high percentage of heart attacks. Details of some studies, which had large sample sizes, are shown in Table 9.4.

The effects of these drugs depend on the dosage. Normally a general practitioner decides the dosage of the drug depending on the level of blood pressure. For men with SBP just around 160, GP might give them a low dosage but use a high dosage for those with even higher blood pressure. Suppose that the GP prescribes two different dosages of CCBs depending on the level of SBP. This information is used in the simulation of hypothetical interventions with two levels of treatments. The blood pressure reduction depends on the dosage and this study explains it clearly [225].

There are several ways of intervening to reduce blood pressure, such as drug therapy, meditation, yoga etc. Only drug therapy interventions are discussed here because of their simplicity and effectiveness in reducing blood pressure.

A problem postulated is that higher dosages of Hydrochlorothiazide (mostly CCBs and other major drugs) used in the large trials cause increases in serum lipid levels, insulin resistance and uric acid levels. These adverse metabolic effects counteract the positive cardiovascular benefits of blood pressure reduction. Such effects do not occur when drugs are administered in a low dosage, such as 6.25 or 12.5 mg per day of hydrochlorothiazide [225]. However, the reduction in number events could be lower with lower dosages.

There was no information found with this brief literature review on the percentage of subjects rejecting CCBs or any other anti hypertensive drugs. This is probably due to their proven effect, there would be very few who do not want to take CCBs or any other antihypertensive drug. Therefore, in this simulation, it was assumed that the effect of these other hypertensive drugs was 10 – 20 % lower then the effect of CCBs.

Suppose that a GP prescribes 50mg of these Hydrochlorothiazide for those who have 140mmHg $\leq$ SBP $\leq$ 160mmHg and 100mg for those above 160 mmHg, the mean reduction achieved is 34 (range 30 – 36) & 16 (range 10 – 18)mmHg of SBP and DBP for higher dosage and 23(15 – 33) & 17(5 – 23) mmHg of SBP and DBP for lower dosage [225]. Since there was no information available on their distributions, the

reduction is assumed normal with standard deviation different for each case, depending upon the type of intervention (Table 9.4 and 9.5). The time taken for these blood pressure reductions was assumed instantaneous and hence the effect of reduction on the risk of MI and stroke was assumed either from the start of intervention or after one year. Various parameters that are specified in the simulation model are as follows.

- Column 1 – scenario number
- Column 2 – time take for the intervention to take effect.
- Column 3 – condition used to identify subjects at low risk. Here the low risk group are the men with SBP<140 (no intervention is applied on them)
- Column 4 – separate interventions applied for risk groups of men with 140mmHg ≤ SBP ≤ 160mmHg and SBP>160.
- Column 5 – Intervention side effects.
- Column 6 – Mean and SD of events with no intervention
- Column 7 – Mean and SD of events after 16 years of intervention
- Column 8 – Mean and SD of saved events after 5 years of intervention
- Column 9 – Mean and SD of saved events after 10 years of intervention
- Column 10 – Mean and SD of saved events after 16 years of intervention

**Table 9.4 Hypothetical SBP intervention scenarios for reducing MI events**

| Scenario number | Time for change (Low risk group) | STEP1 groups 1 & 2 | STEP 2 Intervention | Intervention Side effects | Mean(SD) bef intervention | 16Y Inter- Mean(SD) | 5Y-Evns Mean(SD) | 10Y-Evns Mean(SD) | 16Y-Evns Mean(SD) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 year | SBP<140 | SBP by 34 and 23 mmHg (SD 3, 2) | 10% | 1193(33.3) | 1003(30.6) | 64(22.3) | 136(33.7) | 190(42.6) |
| 2 | 0 year | SBP<140 | SBP by 34 and 23 mmHg (SD 3, 2) | 20% | 1191(30.7) | 1007(30.1) | 63(22.1) | 131(33.2) | 184(40.7) |
| 3 | 0 year | SBP<140 | SBP by 30 and 18 mmHg (SD 3, 2) | 10% | 1191(32.5) | 1036(30.9) | 53(22.6) | 110(32.9) | 155(44.2) |
| 4 | 0 year | SBP<140 | SBP by 30 and 18 mmHg (SD 3, 2) | 20% | 1192(32.6) | 1038(29.6) | 53(22.3) | 110(32.2) | 154(42.6) |
| 5 | 1 year | SBP<140 | SBP by 34 and 23 mmHg (SD 3, 2) | 10% | 1192(32.2) | 1024(29.6) | 57(22.1) | 119(33.1) | 168(41.1) |
| 6 | 1 year | SBP<140 | SBP by 34 and 23 mmHg (SD 3, 2) | 20% | 1192(34.0) | 1028(30.7) | 55(22.3) | 113(33.2) | 164(44.8) |
| 7 | 1 year | SBP<140 | SBP by 30 and 18 mmHg (SD 3, 2) | 10% | 1192(32.1) | 1054(30.1) | 47(23.1) | 98(33.4) | 138(43.3) |
| 8 | 1 year | SBP<140 | SBP by 30 and 18 mmHg (SD 3, 2) | 20% | 1192(32.0) | 1056(31.1) | 46(23.3) | 98(33.2) | 136(43.1) |
| 9 | 1 year | SBP<140 | SBP by 30 and 18 mmHg (SD 10, 7) | 10% | 1192(32.1) | 1056(30.1) | 48(23.6) | 99(33.4) | 136(43.1) |

## Table 9.5 Hypothetical SBP intervention scenarios for reducing Stroke events

| Scenario number | Time for change | STEP1 (No risk groups 1 & 2) | STEP 2 Intervention groups 1 & 2 | Intervention Side effects | Mean(SD) bef intervention | 16Y Inter- Mean(SD) | 5Y-Evns Mean(SD) | 10Y-Evns Mean(SD) | 16Y-Evns Mean(SD) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 year | SBP<140 | SBP by 34 and 23 mmHg (SD 3, 2) | 10% | 701(25.3) | 631(24.1) | 13(17.1) | 45(27.4) | 70(34.1) |
| 2 | 0 year | SBP<140 | SBP by 34 and 23 mmHg (SD 3, 2) | 20% | 699(25.6) | 630(23.9) | 13(16.4) | 44(26.5) | 68(35.3) |
| 3 | 0 year | SBP<140 | SBP by 30 and 18 mmHg (SD 3, 2) | 10% | 700(25.2) | 641(24.7) | 11(17.3) | 38(27.1) | 59(34.9) |
| 4 | 0 year | SBP<140 | SBP by 30 and 18 mmHg (SD 3, 2) | 20% | 701(24.5) | 642(24.9) | 11(16.8) | 37(26.9) | 59(33.6) |
| 5 | 1 year | SBP<140 | SBP by 34 and 23 mmHg (SD 3, 2) | 10% | 699(24.4) | 637(24.1) | 12(16.1) | 40(27.0) | 62(33.2) |
| 6 | 1 year | SBP<140 | SBP by 34 and 23 mmHg (SD 3, 2) | 20% | 699(25.5) | 636(25.6) | 11(17.3) | 41(26.2) | 63(34.6) |
| 7 | 1 year | SBP<140 | SBP by 30 and 18 mmHg (SD 3, 2) | 10% | 700(25.8) | 648(24.3) | 10(17.4) | 33(25.9) | 52(35.4) |
| 8 | 1 year | SBP<140 | SBP by 30 and 18 mmHg (SD 3, 2) | 20% | 700(25.6) | 647(24.1) | 9(15.9) | 33(27.8) | 52(35.2) |
| 9 | 1 year | SBP<140 | SBP by 30 and 18 mmHg (SD 10, 7) | 10% | 702(25.4) | 649(24.2) | 10(17.1) | 32(26.6) | 53(35.4) |

For every 10,000 men there would be approximately 1193 events of MI in the simulated population after 16 years of follow-up. In the Caerphilly study, there were 12.01% MI events for 1190 men, whereas in the simulated data, there were 11.9% events.

In the scenarios shown in tables 9.4 and 9.5, it can be seen that the mean number of events saved depends on the effect of the intervention and the time taken for their effects to show the decreased risk. However, it is possible to assume in a hypothetical intervention scenario that the risk of MI comes down immediately (possibly after a few weeks) from the start of the therapy. Instead of assuming one year as the time taken for the risk to go back to normal, a simulation was also performed by assuming that the CCBs show their effect after 10 weeks of the commencement of the therapy. However, there was no huge difference in the events saved (2 to 4 events were added). And also, the literature review shows that the blood pressure reduction is possible as in the scenarios 1 to 4 (Table 9.4 & 9.5). There is not a huge effect on the number of people having side effects with the usage of CCBs because the other drugs are also effective. Therefore, from the table 9.5, it can be concluded that the 3rd or 4th scenarios are the most plausible. This means there would be about 15 – 16% reduction in the MI events.

Also of importance is to look at the number of events saved at 5 years and 10 years intervals. It appears that there are approximately 34% and 71% events saved at 5 and 10 years of the 16 year follow-up respectively. This shows that the intervention on SBP reduces the risk MI at an approximately equal rate.

There were around 700 (7%) stroke events for every 10,000 men and when no intervention was applied, the number of events in the simulation compared favourably to the actual number in the CaPs (6.4%). After the hypothetical interventions, the number of stroke events ranged from 624 to 648 depending on the intervention scenario. This shows that there is no huge difference between the any of the hypothetical interventions shown in the Table 9.6. If the 3rd and 4th scenarios are assumed to be plausible, then approximately 8 – 11% events could be saved.

After 5 years and 10 years, it appears that the reduction of stroke events is around 20% and 64% of the reduction in the number of events at 16 years. The number of stroke events saved appears to be 50% at 10 years compared to the 5 years.

## 9.3.3 Smoking

In a prospective study of 7178 people aged 65 years or older, total mortality among those who smoked was twice that of those who have never smoked [226], whereas subjects who smoked previously had cardiovascular mortality rates similar to those who never smoked. The risk of developing coronary heart disease can be reduced by half after one year of smoking cessation, and after 2 years, the risk equals to that of people who never smoked [227]. However, these men's ages were below 55 years suggesting that the risk reduction is higher in relatively younger men then. Another study found that among men who had quit smoking, mortality from CHD decreased to almost the level of subjects who had never smoked after 5 years or more of smoking abstinence [228]. From both of these studies, it could be assumed that the risk of MI will be as of non-smokers after quitting smoking for about 2-10 years.

There are several pharmacological methods of smoking cessation depending on the level of smoking. A meta analysis of 17 studies involving various nicotine patches reported abstinence rates of 22% at 6 months for the treatment group as compared with 9% for the group which received a placebo patch [229]. The nicotine nasal spray delivers nicotine more rapidly than the gum or patch [230]. The results of a randomised control trial showed that 32% of patients who used nasal spray were abstinent at 6 months, compared to 12% who received placebo [231]. Therefore, it can be concluded that there are good pharmacological methods that would help smokers quit smoking.

Reducing smoking could reduce CHD events. Of the many attempts to reduce smoking, few have been successful (Table 9.6). John R et al [232] showed that very few people have reduced smoking after 4 years of follow-up. Inter99 study [233] proved that most smokers have no plans to quit smoking. The literature review on interventions on smoking cessation suggests that achieving high cessation rates is difficult. Although making people quit smoking is not easy, even if it were to be successful in lowering smoking levels, a study showed that lowering number of cigarettes do not decrease the risk of heart disease [234]. All these studies are not based on pharmacological interventions. From these studies, it appears that quitting smoking is not easy if it is just advice based and a devised intervention should be based on pharmacological methods that would help smokers better.

# Table 9.6

## Intervention studies on the effect of reducing smoking on vascular risk

| Study/type of intervention | Design/sample size. | Age group and follow-up. | RR | Conclusion |
|---|---|---|---|---|
| [232] Ability of smokers to reduce their smoking and its association with future smoking Cessation.(1999) | The 1410 subjects who smoked at baseline. After questionnaire filling, heavy smokers were sent to intervention treatment. | 4 years of follow-up. | After 2 years of follow-up, 60% have increased or remained on same smoking habits. | Very few people have reduced their smoking habits. |
| [233] Smoking cessation intervention in a large randomised population-based Study. The Inter99 study(2005) | 2408 daily smokers. After questionnaire filling, heavy smokers were sent to intervention treatment. | 1 year of follow-up. | 16.3% and 12.7% reduction in smoking for high and low intensity groups. | Most smokers in the population have no plans to quit in the near future. |
| [235] Physicians Taking Action Against Smoking.(2001) | 5-year intervention program for men and women. | 5 years follow-up | All possible methods of reducing smoking in heavy smokers are discussed. | suggests that an integrated, theory-based program to improve physicians' counselling practices could reduce tobacco use. |

Smoking intervention is not as straightforward as a blood pressure intervention; it is difficult to make people quit smoking. Moreover, the risk does not appear to go down immediately when the cessation starts. If the cessation intervention is reasonably a good one, the smoking levels may down by 15-20 percent. Suppose that the intervention on the hypothetical population is a reasonably good one (a pharmacological intervention, say,) so that there is a decrease in smoking by 20-25%. Since the Caerphilly data was recoded into binary data in the multivariate analysis, it will be assumed that the effect of the intervention simply changes the percentage of smokers; no attempt will be made to model changes in the number of cigarettes

smoked among men who continue to smoke. Those who do not continue with an intervention are also taken into account because of the low cessation rate [236]. Therefore, it was assumed that 20 to 25% of the smokers quit smoking. Other possibilities are also simulated. The time taken for the risk of smokers to come down to 'normal' (i.e., equal to those who do not smoke) was assumed to be between 2 to 10 years. In Figure 9.3, the intervention flowchart shown can be ideal for this scenario by excluding steps 6 & 7. Due to the very low cessation rates, step 6 and step 7 are not considered.

- Column 1 – scenario number
- Column 2 – time take for the intervention to take effect.
- Column 3 – intervention scenario for making people quit smoking. A random selection of men (50% – 100%) depending on the scenario quit smoking.
- Column 4 – percentage of people who accept intervention.
- Column 5 – percentage of people who reject intervention.
- Column 6 – Mean and SD of events with no intervention
- Column 7 – Mean and SD of events after 16 years of intervention
- Column 8 – Mean and SD of saved events after 5 years of intervention
- Column 9 – Mean and SD of saved events after 10 years of intervention
- Column 10 – Mean and SD of saved events after 16 years of intervention

The Table 9.7 shows different intervention scenarios for smoking intervention.

Table 9.7 Hypothetical smoking intervention scenarios for reducing MI events

| Scenario number ↓ | Time taken for change | Intervention for smoking reduction | Accept inter | Reject inter | Mean(SD) bef intervention | 16Y Inter- Mean(SD) | 5Y-Evns Mean(SD) | 10Y-Evns Mean(SD) | 16Y-Evns Mean(SD) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 years | 50% quit randomly | 70% | 30% | 1216(32.4) | 1136(31.4) | 28(24.6) | 58(36.5) | 80(45.3) |
| 2 | 2 years | 50% quit randomly | 50% | 50% | 1215(33.8) | 1162(31.5) | 19(24.2) | 39(36.6) | 53(46.1) |
| 3 | 2 years | 100% quit smoking | 50% | 50% | 1216(32.4) | 1103(31.8) | 39(23.9) | 82(35.4) | 112(45.9) |
| 4 | 2 years | 100% quit smoking | 100% | 0% | 1214(32.4) | 994(29.5) | 77(23.7) | 161(35.1) | 221(44.6) |
| 5 | 10 years | 50% quit randomly | 70% | 30% | 1215(32.7) | 1162(32.6) | 19(24.7) | 38(37.1) | 53(46.2) |
| 6 | 10 years | 50% quit randomly | 50% | 50% | 1215(33.5) | 1174(32.7) | 14(27.4) | 29(36.7) | 41(47.3) |
| 7 | 10 years | 100% quit smoking | 50% | 50% | 1216(32.6) | 1135(31.8) | 28(27.0) | 59(36.1) | 81(45.2) |
| 8 | 10 years | 100% quit smoking | 100% | 0% | 1216(33.6) | 1057(30.7) | 56(26.7) | 116(35.5) | 159(46.0) |

For the 16 year follow-up, there were 12% MI events recorded in the Caerphilly study. When compared to the average number of MI events in the simulation model, there were 12.2% which is a close estimate.

In the above table, it can be seen that the higher the smoking cessation rate, the higher the number of events saved. However, the most plausible intervention scenario is the one with the 25% reduction rate in all smokers. If this intervention is applied, then there could be a saving of 3.4 to 4.4% MI events (scenarios 2 & 6) at the end of 16 years. This reduction appears to be low due to the low cessation rates. If there could be a slightly better intervention which could possibly increase the cessation rates from 25% to 35%, then 4.4 to 6.6% MI events could be saved. This simulation shows that the increase in cessation rate decreases MI events. If all the smokers were persuaded to quit smoking in 2 to 10 years, which is very unlikely, there could be a reduction of 13.1% to 18.2% of MI events implying that smoking is a major risk factor. This shows that the maximum reduction achievable by the smoking intervention is about 18%.

In chapter 4, it was seen that the risk of stroke is significantly higher in smokers and ex-smokers compared to the never smoked men. When it comes to intervention, one can only change current smokers to ex-smokers. However, ex-smokers also have higher risk of stroke which was found almost equal to that of smokers (Section 4.2.2.2). Therefore, this simulation is not possible.

## 9.3.4 Cholesterol levels

Trials of cholesterol lowering drugs in the primary prevention of coronary heart disease have demonstrated that lowering cholesterol levels in middle aged men with hypercholesterolemia reduces the incidence of myocardial infarction [237-240]. A meta analysis paper with 65 studies including 200,607 subjects in it, showed that statins reduce the risk of coronary disease in patients with or without coronary heart disease [241].

The univariate analysis and the multivariate analysis in the Caerphilly study showed that total cholesterol is a major risk factor and this literature review suggests that it can be modifiable in order to reduce the MI incidence. The statin trials described in the following table add a bit more to the studies discussed in the previous paragraph.

**Table 9.8**

**Intervention studies on the effect of reducing cholesterol levels on vascular risk**

| Study | Design/sample size. | Results | Variables analysed | Outcome/length of follow-up. | Conclusions |
|---|---|---|---|---|---|
| [242] The Distribution of 10-Year Risk for Coronary Heart Disease Among U.S. Adults(2004) | Longitudinal study with 13,769 subjects. Self reported medication on lowering cholesterol. | There is no difference between those who took the medicine and those who did not. | All measured risk factors of CHD | 10 year follow-up. CHD mortality is the outcome. | Self-reported use of cholesterol-lowering medications showed no evidence of a difference for those who used medication and those who did not. |
| [243] Effects of LDL-lowering with simvastatin on stroke and other major vascular events.(2004) | 20,536 people with Cerebrovascular disease or other high-risk conditions. | 22% (95% CI 15-29) relative reduction in the rate of first major vascular event. | LDL-C | Vascular event mortality is the outcome and 5 years of follow-up. | Using 40mg of simvastatin has a definite effect on reducing heart attacks. |
| [244] Coronary Drug Project.(1996) | 8,341 men with high cholesterol. | Using Statin mono therapy for reducing LDL-C by 10%, 8 year non fatal MI was reduced by 27%. Results were statistically significant. | LDL-C | Non fatal MI. follow-up time of 8 years. | patients with pure LDL abnormalities often cannot reach their treatment goals with statins alone. |

250

| Study | Design/sample size. | Results | Variables analysed | Outcome/length of follow-up. | Conclusions |
|---|---|---|---|---|---|
| [245] Scandinavian Simvastatin Survival Study (4S)(2000) | 4,444 men with high cholesterol. 2,223 men in control group and remaining in statin group. Statin dose used is 10-40mg of simvastatin. | 36% risk of coronary events in statin intervention group. | LDL reduction. | All major coronary events were the outcome and 5.4 years was the follow-up time. | Simvastatin has a beneficial effect in reducing CHD and stroke and acts as a primary preventive measure. |
| [138] Can change in high-density lipoprotein cholesterol Levels reduce cardiovascular risk?(1995) | Systematic review on LDL, HDL and Total Cholesterol. 51 trials were identified out of which 19 were statin trials. | 32% of total cholesterol was reduced and 45% of LDL-C was reduced. | HDL, LDL and Total Cholesterol. | Stroke is the outcome. | LDL reduction is given more importance in reducing the risk stroke. |

251

The majority of the studies show a high reduction of MI events with the use of statins. A recent study showed that there is no significant difference between any of the different statins available in the market [246]. As with the SBP reduction using CCBs, the reduction in total cholesterol levels depends on the dosage of statins. 40 mg of any statin (Simvastatin, Pravastatin, Lovastatin) showed similar reductions of total cholesterol levels ranging from 1.1 − 2.5 mmol/L reduction [247]. All these effects were achieved on the subjects with more than 5mmol/L total cholesterol level. Therefore, with the simulation model, the hypothetically simulated subjects with total cholesterol more than 5mmol/L were intervened to reduce their levels, assuming that they were treated with 40 mg of statin. In this brief literature review, no side effects and rejection rates of statins were found. Assuming that there would be some side effects and these get another kind of statin (which has slightly lower effect, i.e., 20% lower effects), intervention scenarios were simulated. The reduction in the high risk group was assumed to be 1.8mmol/L with standard deviation of 0.5 because it seems to be in the middle of the range of reduction suggested by a study [247]. Various possibilities of intervention scenarios were simulated and are shown in the following table. The time taken for risk reduction was also assumed to be either immediate effect or after one year. The lists of columns in Table 9.9 are explained below.

- Column 1 − scenario number
- Column 2 − time take for the intervention to take effect.
- Column 3 − condition used to identify subjects at low risk. Here the low risk group are the men with Total cholesterol less than 5 or 6 depending on the intervention scenario.
- Column 4 − Intervention scenario − standard one for everyone. Reduce total cholesterol in the high risk men by a mean 1.8 and SD of 0.5.
- Column 5 − Intervention side effects.
- Column 6 − Mean and SD of events with no intervention
- Column 7 − Mean and SD of events after 16 years of intervention
- Column 8 − Mean and SD of saved events after 5 years of intervention
- Column 9 − Mean and SD of saved events after 10 years of intervention
- Column 10 − Mean and SD of saved events after 16 years of intervention

The following table gives various hypothetical intervention results.

## Table 9.9 Hypothetical Total cholesterol intervention scenarios for reducing MI events

| Scenario number ↓ | Time for change (low risk group) | STEP1 Intervention on subjects | STEP 2 | Intervention Side effects | Mean(SD) bef intervention | 16Y Inter- Mean(SD) | 5Y-Evns Mean(SD) | 10Y-Evns Mean(SD) | 16Y-Evns Mean(SD) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 year | TCHOL< 5mmol/L | By 1.8 mmol/L (SD 0.5) | 0% | 1175(32.2) | 915(28.6) | 60(22.2) | 174(32.8) | 260(40.8) |
| 2 | 0 year | TCHOL< 5mmol/L | by 1.8 mmol/L (SD 0.5) | 10% | 1177(31.7) ( | 917(28.2) | 58(21.7) | 177(31.7) | 260(41.6) |
| 3 | 0 year | TCHOL< 6mmol/L | by 1.8 mmol/L (SD 0.5) | 0% | 1176(32.1) | 1062(30.9) | 26(21.5) | 76(32.5) | 114(41.8) |
| 4 | 0 year | TCHOL< 6mmol/L | by 1.8 mmol/L (SD 0.5) | 10% | 1175(31.1) | 1062(31.0) | 26(20.9) | 75(32.2) | 113(40.8) |
| 5 | 1 year | TCHOL< 6mmol/L | by 1.8 mmol/L (SD 0.5) | 0% | 1175(32.6) | 940(30.0) | 54(21.9) | 158(31.8) | 236(42.4) |
| 6 | 1 year | TCHOL< 5mmol/L | By 1.8 mmol/L (SD 0.5) | 10% | 1174(33.4) | 941(29.5) | 54(21.4) | 157(33.3) | 234(42.4) |
| 7 | 1 year | TCHOL< 6mmol/L | by 1.8 mmol/L (SD 0.5) | 0% | 1175(30.8) | 1073(30.6) | 24(21.7) | 69(33.5) | 103(39.7) |
| 8 | 1 year | TCHOL< 6mmol/L | by 1.8 mmol/L (SD 0.5) | 10% | 1176(32.8) | 1074(29.6) | 23(23.2) | 68(31.9) | 102(40.4) |
| 9 | 1 year | TCHOL< 5mmol/L | by 1.5 mmol/L (SD 1) | 10% | 1175(32.6) | 968(29.6) | 49(26.2) | 135(34.6) | 207(42.5) |

For every 10,000 men there were approximately 1175 MI hypothetical events without any intervention compared to the 12% in Caerphilly men, it is a close estimate. For an hypothetical intervention to reduce total cholesterol for every subject with more than 5 mmol/L using 40mg statins, there were 233 to 261 events saved (approximately 19.8 to 22.2%) depending upon the time taken for the interventions to show an effect and the percentage of subjects with side effects. If the same treatment was simulated on men with total cholesterol more than 6mmol/L, there were 101 to 113 (approximately 8.6 – 9.6%) events saved depending upon time and percentage of subjects with side effects. This shows that if the total cholesterol is lowered for every subject with total cholesterol levels above 5mmol/L, there could be a higher reduction in MI events. Since this total cholesterol reduction interventions discussed in Table 9.9 are based on the drug therapy, they should be possible to achieve. Reduction of total cholesterol was thought to show effect after 10 weeks of commencement of the intervention and the results showed no huge difference. Even with a large standard deviation, there appears to be very little difference. An intervention with a mean reduction of 1.5 mmol/L and a standard deviation 1 showed similar results.

Total cholesterol was not a significant factor for stroke according to the Caerphilly analysis (section 5.7.2 and section 5.8.1.2) and therefore, intervention scenarios for stroke are not simulated.

## 9.3.5 Exercise and obesity

There is compelling epidemiological evidence from many different populations that leisure time physical activity is associated with a reduced risk of CHD in middle age. Although there is a consensus on an inverse relationship between physical activity and CHD there is still debate on the amount and intensity of physical activity required to achieve the benefit of reduced risk of CHD.

A study conducted an intervention to change the physical fitness of middle aged men for improving their physical fitness from unfit to fit. It showed a 44% reduction in age-adjusted all-cause mortality and a 52% reduction in CVD risk after adjustment [248]. A review of six studies [249] showed that being physically active is associated with about a 40 to 50% reduction in risk of stroke and CHD.

The role of exercise in the treatment of obesity in 11 studies was reviewed and the results show that weight loss was statistically not significant with the combination of exercise and reduced energy intake [250]. However, the aim here is to find whether the physical activity decreases the risk of MI and stroke or not. Another study assessing the role of physical activity in the prevention of cardiovascular disease reviewed six studies and concluded that being physically active was associated with a 40-50% reduction in the risk of stroke and CHD [251]. This review suggests that being physically active reduces the risk of MI and stroke.

In the Caerphilly data, BMI showed significant associations to the risk of MI but showed otherwise for stroke in the uni-variate models (section 5.12). However, BMI was not significant in the multivariate models suggesting that all the effects caused by BMI are explained by other significant risk factors.

Increased leisure activity showed a very significant decreased risk of MI. There was no significant difference between moderately active men and highly active men for having an MI. Moderate/high physical activity can help the heart get stronger. Most people find that exercise improves their symptoms, reduces stress and boosts energy levels. Regular physical activity also may lead to other important health advantages,

including weight control, weight loss, better circulation and blood pressure, and lower cholesterol levels, all of which are especially important in controlling heart disease.

For instance, in intervention trials it was found that providing physical activity advice verbally and through a written prescription was more effective than verbal advice alone, and that supplementing brief advice and a written prescription with self-help booklets was more effective (33% increased physical activity) than advice and a prescription alone (23% increased physical activity) after 7-8 months of follow-up [252]. In a study testing four levels of intervention [253] it was found that only the most intensive intervention, which included six counselling sessions and vouchers for attending a leisure centre, achieved significant increases in physical activity (55% increased physical activity).

There is a lot of information available in the literature about the effect of physical activity reducing other biological risk factors such as weight, blood pressure and cholesterol levels [254, 255]. Since this type of intervention is a multivariate problem, this is discussed in the multivariate interventions follows the individual risk factor interventions. With this brief literature review, there was no clear information available on the possible increase in physical activity levels with an intervention and success in achieving these levels.

There is a need to make some assumptions in order to simulate interventions to increase physical activity and hence estimate the effects of an intervention. It was thought that the majority of the people nowadays understand the advantages of being physically active. However, even after accepting the importance of physical activity, there could be a high chance that a high proportion of people would neglect it. Therefore, it is assumed that 90% subjects accept the idea of increasing their physical activity but only 50% of them actually workout. Since the effects are not going to be immediate with physical activity, it was assumed that the effect would be shown after one year of the commencement of intervention. The list of columns in the intervention results Table 9.10 are explained below.

- Column 1 – scenario number
- Column 2 – time take for the intervention to take effect.
- Column 3 – Percentage of people who accept intervention given that they are were not physically active before.
- Column 4 – Percentage of people who reject intervention given that they are were not physically active before.
- Column 5 – Percentage of people who actually workout after accepting the intervention
- Column 6 – Mean and SD of events with no intervention
- Column 7 – Mean and SD of events after 16 years of intervention
- Column 8 – Mean and SD of saved events after 5 years of intervention
- Column 9 – Mean and SD of saved events after 10 years of intervention
- Column 10 – Mean and SD of saved events after 16 years of intervention

Various other possible intervention scenarios were simulated and are shown in the Table 9.10.

**Table 9.10 Hypothetical exercise intervention scenarios for reducing MI events**

| Scenario number | Time for change | Accept intervention | Reject intervention | Intervention | Mean(SD) bef intervention | 16Y Inter- Mean(SD) | 5Y-Evns Mean(SD) | 10Y-Evns Mean(SD) | 16Y-Evns Mean(SD) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 years | 90% | 10% | 50% men workout | 1189(32.0) | 1135(32.1) | 19(25.4) | 38(36.6) | 54(46.0) |
| 2 | 1 years | 90% | 10% | 60% men workout | 1187(33.2) | 1126(31.6) | 21(24.9) | 43(36.2) | 61(46.0) |
| 3 | 1 years | 70% | 30% | 70% men workout | 1187(31.7) | 1132(31.7) | 19(25.3) | 39(36.9) | 55(45.5) |
| 4 | 1 years | 100% | 0% | 100% men workout | 1186(32.8) | 1074(30.2) | 39(25.1) | 79(35.4) | 112(46.3) |
| 5 | 2 years | 90% | 10% | 50% men workout | 1187(30.6) | 1139(31.8) | 17(25.7) | 35(36.0) | 48(44.0) |
| 6 | 2 years | 90% | 10% | 60% men workout | 1187(31.5) | 1127(32.0) | 21(24.4) | 43(36.1) | 60(44.3) |
| 7 | 2 years | 70% | 30% | 70% men workout | 1187(32.3) | 1133(32.0) | 19(26.3) | 38(37.1) | 54(49.5) |
| 8 | 2 years | 100% | 0% | 100% men workout | 1188(31.8) | 1078(31.2) | 39(26.1) | 80(35.8) | 111(44.8) |

For every 10,000 simulated subjects to whom the intervention was applied, there were approximately 11.9% MI events. In the Caerphilly data, there were 12% events observed suggesting that this simulation model has estimated MI events reasonably well.

The fourth and the eighth intervention scenarios in the above table assumed all men with low physical activity change their activity level. Though this cannot be practically possible, it can be observed that the maximum number of events that could be saved using an intervention on exercise would be around 110 (9.3%). All the remaining intervention scenarios seem to show very similar effects. It can be seen that there is no huge difference in the reduction of events even when the effect of reduction was assumed to be two years rather then one year. Therefore, if a plausible intervention could be chosen, (1 – 3 and 5 – 7 intervention scenarios were thought to be practically possible, according to the literature review), there could be a reduction of 4% – 5% MI events.

Physical activity was not a significant factor for stroke in the Caerphilly data. Therefore, intervention scenarios were not simulated.

## 9.3.6 Psychological factors

The psychological factors that were measured in the Caerphilly cohort showed no significance in measuring MI or stroke other than anger out for an increased risk of MI. Hence, an intervention scenario to reduce anger out could reduce the risk of MI and therefore reduce the number of events. An intervention scenario for reducing anger out can be simulated theoretically, but backing up these interventions as plausible interventions could only be done with a literature review. Anger out being a specific anger scale, it is unlikely to find an intervention to reduce it. Therefore, literature available on general anger reduction for cardiovascular intervention was searched.

There are no clinical trials of anger management interventions with VAD events as the outcome. However, several clinical trials using stress management techniques in combination with other modalities do show reduced hostility and reduced risk of new VAD events in coronary patients who survived a cardiac event [256, 257]. A recurrence rate of MI was found to be 13% in the treated group compared to the 29% in the no treatment group (Friedman et al study [256]). Blumenthal et al [257] concluded that behavioural interventions offer additional benefit over and above usual medical care in cardiac patients with evidence of heart disease. Their study was based on a relatively small number of subjects (total 107 subjects).

Hostility is closely related to anger. The association of hostility with unhealthy behaviours occurs early in the natural history of heart disease [258, 259]. Siegler et al [258] suggested that the hostility showed an association with various other cardiovascular risk factors and therefore it is important to try reducing anger. Similarly Räikkönen et al [259] also showed that the hostility is associated with various cardiovascular risk factors after a 3 year follow-up study in adolescent and young men. This suggests that reducing hostility, which is similar to anger, on reduction may reduce the risk of MI.

Concerted efforts to promote exercise, prevent weight gain, and reduce the likelihood of smoking may be especially beneficial in hostile young adults. Mental health services frequently concern anger management issues, with treatment often being

successful [260]. Thus, referring people for anger treatment is a good option for reducing risk of MI.

A meta analysis study with 23 studies showed that 18 (713 subjects) of those studies which used cognitive behavioural therapy showed a significantly decreased anger (OR=0.68 95% CI: 0.52 – 0.83) compared to the other therapies (other therapies are cognitive, relaxation, other) [261]. Therefore, it shows that anger could be reduced using various therapies and significant results can be obtained.

There was very little information available of effect sizes of anger reduction; therefore, various possibilities of anger reduction were simulated. Some of them might not be practically possible, but they are simulated to check the maximum reductions that can be obtained by the anger management therapy.

It could be argued that most people may not accept when they are asked to take part in anger management intervention. Therefore, it was assumed that only a small percentage of subjects accept anger management and of those, only a small percentage is successful. Since the Anger-out scale was recoded into binary data, it will be assumed that the effect of the intervention simply changes the percentage of high anger-out subjects. Various other possibilities are simulated and listed in Table 9.11. The columns of the following table are explained below.

- Column 1 – scenario number
- Column 2 – time take for the intervention to take effect.
- Column 3 – Percentage of people who accept intervention given that they have high anger out score.
- Column 4 – Percentage of people who reject intervention given that they have high anger out score.
- Column 5 – Percentage of people who succeed in reducing their anger.
- Column 6 – Mean and SD of events with no intervention
- Column 7 – Mean and SD of events after 16 years of intervention
- Column 8 – Mean and SD of saved events after 5 years of intervention
- Column 9 – Mean and SD of saved events after 10 years of intervention
- Column 10 – Mean and SD of saved events after 16 years of intervention

**Table 9.11 Hypothetical Anger intervention scenarios for reducing MI events**

| Scenario number ↓ | Time taken for change | Accept intervention | Reject intervention | Intervention for Anger | Mean(SD) bef intervention | 16Y Inter- Mean(SD) | 5Y-Evns Mean(SD) | 10Y-Evns Mean(SD) | 16Y-Evns Mean(SD) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 years | 50% | 50% | 40% men succeed | 1179(33.0) | 1105(31.9) | 26(24.5) | 52(36.2) | 74(45.7) |
| 2 | 1 years | 50% | 50% | 60% men succeed | 1179(31.7) | 1067(30.2) | 39(24.2) | 78(36.1) | 111(44.1) |
| 3 | 1 years | 70% | 30% | 70% men succeed | 1176(31.4) | 1000(29.8) | 62(24.9) | 123(36.7) | 176(42.7) |
| 4 | 1 years | 100% | 0% | 100% men succeed | 1178(32.1) | 816(27.6) | 127(24.0) | 253(35.3) | 362(41.8) |
| 5 | 2 years | 50% | 50% | 40% men succeed | 1178(33.9) | 1106(31.5) | 25(23.7) | 50(35.7) | 72(46.8) |
| 6 | 2 years | 50% | 50% | 60% men succeed | 1177(32.1) | 1069(32.1) | 38(24.7) | 76(36.5) | 108 (44.5) |
| 7 | 2 years | 70% | 30% | 70% men succeed | 1178(32.5) | 1004(31.3) | 61(25.3) | 122(36.1) | 174(44.3) |
| 8 | 2 years | 100% | 0% | 100% men succeed | 1177(32.7) | 823(27.6) | 124(25.9) | 248(36.9) | 354(43.4) |

For every 10,000 men there would be approximately 1178 MI events before intervention. This is very close to what was observed in the Caerphilly study.

If suppose that all the men with high anger out score change their anger out score to low anger out, there appears to be a reduction of 30 – 31% MI events (scenarios 4 & 8 in the Table 9.11). Though reducing anger in every case is practically unlikely, this example illustrates that there could be a very high number of MI events saved with reduction of anger. Depending on the intensity of the intervention that could be used to reduce anger out, there is a possibility of saving 6 – 31% MI events. Even if only 30% take up the intervention and only 25% of them succeed, there will be a reduction of 2.4% MI events.

## 9.4 Multiple intervention scenarios

Hypothetical intervention scenarios discussed in the previous section were only performed to find what could happen if a single risk factor were changed. However, more often then not, in a realistic situation, a change in any risk factor modifies other risk factors. For example, an intervention on leisure activity showed reduction in blood pressure and cholesterol levels [254, 255]. There is a Meta analysis study which showed a reduction in blood pressure with walking [262]. There is an evidence of exercise decreasing anxiety, depression and stress [263].

It has been known for many years that high-dose antihypertensive drugs (say CCBs ≥50 mg) negatively affect the lipid levels. In 1976 Ames and Hill [220] reported an average rise in total cholesterol levels of 12 mg/dL in patients treated with high doses of antihypertensive drugs. Since then multiple reports have documented the adverse effects of various antihypertensive drugs on lipid levels. Traditional high-dose antihypertensive therapy has been reported to increase total cholesterol levels by approximately 4%, and serum LDL cholesterol levels by 10%, with lesser effects on VLDL cholesterol and HDL cholesterol [220, 264-267]. One study showed that increasing exercise decreases DBP by 4mmHg and total cholesterol by 0.12 mmol/L [254]. Smoking is a risk factor all by itself and the literature did not show any evidence of smoking affecting other risk factors. Though the literature did not show any direct evidence, it should be noted that someone giving up smoking may be deciding on a healthy lifestyle and change other aspects of his life.

Multiple intervention scenarios are implemented with standard recommendations from the literature review. A task force report [268] recommended that the primary prevention could be achieved by keeping the blood pressure 140/90, total cholesterol below 5 mmol/L, BMI ≤ 25 and quit smoking. These are the recommendations that are supposed to be achieved for everyone but are always not possible.

There could be various possible multivariate scenarios for hypothetical interventions. All the plausible and probably achievable interventions discussed previously for each individual risk factor could be performed together to check the number of events that could be saved. Alcohol consumption and SBP were the only two risk factors that were used for the simulation modelling for stroke. Though the remaining factors discussed for MI were not significant for stroke, those were included in the stroke multivariate model to see their additional effect. Interventions on alcohol consumption were also considered for MI. Alcohol consumption was included in the multivariate model for MI because it enable the model to be more realistic and also, alcohol consumption would not make huge impact anyway.

The following two tables (9.12 & 9.13) show the different simulations for reducing the risk of MI and stroke. Age was also adjusted for each multivariate simulation model.

In each individual risk factor discussed previously, the most plausible scenarios were selected and were used for these multivariate intervention scenarios. Impractical scenarios are also discussed to assess the maximum impact of such multivariate hypothetical intervention.

**Table 9.12 Hypothetical multivariate intervention scenarios for reducing MI events**

| Scenario number ↓ | Risk factor (scenario number) | Mean(SD) before intervention (95% CI) | Mean(SD) after intervention (95% CI) | Mean(SD) of events saved (95% CI) |
|---|---|---|---|---|
| 1 | ALC (4) SBP(4) Smoking (6) TCHOL (6) Exerc (3) Anger (5) | 16Ys- 1284(31.9) (1281.9, 1285.8) | 16 Ys-956(29.6) (954.5, 958.2) | 5Ys-101(21.5) (99.4, 103.1) 10Ys-213(34.2) (211.0, 215.2) 16 Ys-327(43.1) (324.8, 330.1) |
| 2 | ALC (1) SBP(4) Smoking (5) TCHOL (5) Exerc (2) Anger (2) | 16 Ys-1284(33.0) (1282.2, 1286.3) | 16 Ys-928(30.8) (926, 929.8) | 5Ys-110(22.1) (108.5, 113.4) 10Ys-231(32.9) (229.7, 233.8) 16 Ys-356(45.4) (353.5, 359.1) |
| 3 | ALC (3) SBP(1) Smoking (2) TCHOL (1) Exerc (4) Anger (1) | 16 Ys-1285(33.7) (1282.6, 1286.8) | 16 Ys-660(23.8) (658.1, 661.1) | 5Ys- 205(22.6) (204, 206.8) 10Ys- 429(33.8) (427, 431.3) 16 Ys-665(40.4) (622.6, 627.6) |

The first scenario shown in the above table discusses the most plausible multivariate intervention scenario and shows a reduction of 25.5% of MI events. The second intervention scenario is designed from the second most plausible scenarios of each individual risk factor, which shows a reduction of 28.4% Mi events. The final scenario was designed with the theoretically possible but practically impossible scenarios for each individual risk factor. This intervention shows a possible reduction of 51.8% MI events. Though this is not possible practically, it appears that higher number of events could be saved if these interventions were to be somehow achieved.

**Table 9.13 Hypothetical multivariate intervention scenarios for reducing Stroke events**

| Scenario number ↓ | Risk factor (scenario number) | Mean(SD) before intervention (95% CI) | Mean(SD) after intervention (95% CI) | Mean(SD) of events saved (95% CI) |
|---|---|---|---|---|
| 1 | ALC (4) SBP(4) Smoking (6) TCHOL (6) Exerc (3) Anger (5) | 16Ys-771(26.4) | 16Ys-283(16.5) | 5Ys-141(16.4) 10Ys-317(23.8) 16Ys-488(31.3) |
| 2 | ALC (1) SBP(4) Smoking (5) TCHOL (5) Exerc (2) Anger (2) | 16Ys-773(25.5) | 16Ys-275(16.5) | 5Ys-143(16.20) (142.4, 144.4) 10Ys-331(24.9) (329.2, 332.2) 16Ys-499(31.3) (496.8, 500.6) |
| 3 | ALC (3) SBP(1) Smoking (2) TCHOL (1) Exerc (4) Anger (1) | 16Ys-772(27.5) | 16Ys-228(15.0) | 5Ys-158(17.1) 10Ys-354(24.1) 16Ys-544(31.5) |

In the above table, it appears that the multiple intervention scenarios for MI, when applied for strokes, showed a high percentage of reduction in stroke events compared to the hypothetical MI interventions. Similar to the MI multiple interventions, there is an increased number of events predicted for stroke when additional risk factors were added in the model though they were not significant independently. With the first intervention scenario, there appears to be a reduction of 63% stroke events. The last intervention scenario may not be practically possible, but if it were to be achieved, there could be a further reduction of stroke event of the order of 70%.

Additional effects of exercise on blood pressure and cholesterol reduction (6 mmHg SBP instead of 4 mmHg DBP and 0.12 mmol/L of total cholesterol [254]) and increased total cholesterol (4%) for those who use CCBs were taken into account in the next step. All these additional effects were thought to follow normal distribution with SD 1 for SBP and 0.2 for total cholesterol. An average increase of 4% cholesterol in the subjects who are

intervened with CCBs was included in the simulation model. These side effects did not change the results hugely.

All these simulation models are data driven and it is only possible to test any other recommendations if the Caerphilly dataset has similar risk factors measured. One of the studies has shown a strategy to reduce 88% of the IHD events and 80% of the stroke events [269]. The risk factors that were used for the intervention study included LDL, serum homocysteine and platelet function. LDL was not measured in $2^{nd}$ phase of the Caerphilly cohort. Therefore, it is not possible to see the effect of this strategy on the Caerphilly subjects. Serum homocysteine and platelet function were measured in the Caerphilly cohort. In the Caerphilly data, serum homocysteine was weakly predictive of IHD events, but when the data was adjusted for other factors, the significance was lost [270]. Platelet function in the Caerphilly data also showed no predictive power [271] for either MI or stroke. Hence, the comparison of this 'polypill' strategy could not be made [272].

## 9.5 Discussion

At the start of the chapter, the assumptions that were needed for a simulation program were discussed in detail with the statistical methods used for simulating the data discussed in detail with examples. The limitations of these simulations were of not including the interactions between the risk factors as the Caerphilly data did not show high correlations between the data. Most of the possible scenarios were discussed and were implemented in the hypothetical interventions using a flowchart wherever suited.

A hypothetical intervention to reduce alcohol consumption appears to show mixed results. At the five years follow-up there were more events then what was expected and as the follow-up continues, there were a few events saved depending upon the type of intervention. The maximum that could be saved with alcohol consumption appears to be lower than 5% events in 16 years. This shows that reducing alcohol consumption may not be a high priority for allocating funds for cardiovascular interventions. However, intervening to reduce alcohol consumption appears to be cheap from the studies in the literature review and there might be other possible risks that could be reduced.

Systolic blood pressure is a proven risk factor of MI and stroke. There were several intervention therapies available in the literature review but only drug therapy was discussed in this scenario because of the proven ability of the anti-hypertensive drugs that are on the market today. There could be 13-16% and 8.4-10% MI and stroke events saved if these interventions were to be real.

Smoking is a major risk factor for MI and other diseases like lung cancer. However, there are not many interventions that were proven to be highly successful. Depending upon the quality of the intervention, the reduction in the MI events range from 3.4-13% at the end of 16 years. Achieving higher rates of abstaining smoking showed fewer MI events. Therefore, carefully designed interventions to make people reduce or quit smoking would lead to considerable reductions in MI events.

Smoking is a major risk factor for stroke too. The Caerphilly analysis showed that those who smoke and those who were ex-smokers have similar risks of having stroke compared to those who never smoke. Therefore, there should be initiatives taken to deliver these results to the public to not only give up smoking to reduce the risk of MI, but also not to smoke for youngsters as they increase the risk of stroke.

Total cholesterol is another risk factor for MI which is well established and has very good treatments. When everyone with a total cholesterol level more than 5 mmol/L can be targeted to reduce their cholesterol levels, there could be a reduction of 20-22% of MI events. These interventions are drug based and highly proven.

Exercise is a risk factor which might have an influence on BMI and also other risk factors of MI. It appears that a plausible intervention scenario to increase leisure activity would decrease MI events by 4-6%. It might look small but the effects of leisure activity on other risk factors will make it an important factor to be intervened.

The emphasis in this thesis is placed on the inclusion of the psychological risk factors for reduction of the risk of MI and stroke. In addition, it tries to explore the possible reduction of events with the intervention of these factors. Hypothetical intervention on anger out seems to lead to a decent reduction in MI events. The literature review for interventions on anger to reduce cardiovascular risk is very brief but one of the major Meta analysis suggested that there could be significant reductions in anger with various therapies. Using this information, various possible and improbable scenarios were simulated to reduce anger. These simulations resulted in a decrease of MI events by 6-31% depending upon the type of intervention. This shows latent anger is a major risk factor and a possible intervention could reduce huge events.

Multivariate analysis showed that the plausible interventions of all the above discussed risk factors at a time might reduce 25-30% MI events by the end of 16 years whereas, the same intervention scenarios that of stroke showed a reduction above 50%. Though most of the risk factors were not significantly associated with stroke, it appears that the overall

effect of these risk factors could have a huge impact on stroke. Not including the correlation between the risk factors for the simulations is a limitation for this multivariate model. If the set of improbable intervention scenarios were applied on the simulated populations, there appears to be a reduction of about 50% MI and 70% stroke events.

In conclusion, there need to be more emphasis given to the design of various intervention scenarios for reducing and stopping smoking, anger out and to increase leisure activity in order to reduce more MI events in addition to the drug therapies for the reduction of blood pressure and cholesterol levels.

But there are many limitations in the simulation models. Primarily, models that are used for the simulation should be relevant. These models depend on many assumptions which might not be valid all times. For example, assumptions were made for the SBP reduction to be normally distributed in the intervened populations. This may not be true and it could affect the results.

The time taken for an intervention to show an effect was assumed to vary between 0-10 years depending upon the risk factor and the type of intervention. This is also on assumption as there is no clear data from the literature review.

Assumptions were made for simulation data for smoking, leisure activity and anger-out as binary variables. However, in a real world, this is not the care. These intervention scenarios assumed that men will switch from, say, current smokers to current non-smokers after they successfully participate in the intervention. It is very hard to believe that men suddenly become ex-smokers as more often then not, a reduction in smoking might be achieved. It was not possible to model the effects of such changes because of the limitations of the CaPs analysis.

Advantages of simulations are many. Instead of spending time and money to find the effect sizes of an intervention, using the available literature, simulation models can estimate the effect sizes. But there are relative difficulties associated with different

interventions. For example taking an anti-hypertensive or statin is very easy, whereas an intervention on anger would be much harder as it would take a lot of effort to perform an intervention on high anger people. These things should be accounted for the design process of any intervention.

Simulations are only as good as the model used and the small size of the Caerphilly study means there is much doubt over that model. In an ideal world, one would have done sensitivity analysis to investigate this.

Changing one risk factor leads to a change in many other risk factors as the risk factors are interrelated. No attempt has been made to do this as the Caerphilly data did not show plausible associations between the risk factors presumably due to its small size.

Another important issue is the reliability of Weibull model used to simulate the number of events. Since a different Weibull model was used for each risk factor to simulate the survival time, the number of events generated by the simulation model varied depending upon the association of that risk factor to MI or stroke. If the Caerphilly study were to be big, then these variations between the events generated by the simulation models might have been pretty close.

# 10. Overall discussion

## 10.1 What has been done so far?

The purpose of the project is to simulate the effect of early detection and intervention of vascular disease in the Caerphilly cohort that has been followed for about 17 years. Vascular risk factor data included haemostatic and thrombosis factors, lipids, hormones, lifestyle (diet, smoking exercise) and psycho-social (anger, anxiety, and depression) factors. Vascular related outcomes analysed were stroke and MI. To simulate the effect of early detection and intervention of vascular disease, there were several intermediate goals set at the start of the project to ensure all the analyses was carried out systematically. These goals were

1. To construct risk models linking individual risk factors with the risk of MI and stroke and to compare these with those found from literature reviews.
2. To construct multivariate models linking the risk of vascular disease with multiple risk factors.
3. To compare risk models constructed from the CaPS data with others in the literature.
4. To review the literature on interventions aimed at modifying risk factors
5. To simulate interventions to assess their public health impact on vascular disease.

**Goal 1:**
The first goal of the project was to construct risk models linking individual risk factors with the risk of MI and stroke and to compare these with those found from literature reviews. Various modelling strategies such as logistic, fractional polynomial, Cox's proportional hazards models were used to achieve the best possible fit for each of the risk factor.

Various biological and behavioural risk factors were analysed individually. Psychological factors were measured in the Caerphilly study and they were the most significant part of

273

the thesis. These psychological factors were analysed in many ways to estimate the association with the risk of MI and stroke. Anger out was the only risk factor that was highly significant in measuring the risk of MI. None of the psychological factors were associated with stroke.

In the Caerphilly heart study, attitudes towards seven coronary related behaviours were measured using the Health Attitude Inventory (HAI). The development of the HAI was based on Fishbein and Ajzen's Theory of Reasoned Action (TRA) discussed in the Chapter 6 (Section 6.5). This dataset is unique to the Caerphilly study. Analysis with these seven scales of attitudes and social norms revealed that there is very little that these attitudes and social norms add to the present knowledge of cardiovascular risk factors.

Measurement error and the random variation was estimated for various risk factors where there were repeated measurements available. It appeared that the regression dilution bias adjustments doubled the association between MI and SBP whereas; it was 21% between MI and psychological distress indicator. Regression adjustment for psychological distress indicator has never been reported before.

The goal was to model the Caerphilly data using various statistical techniques in order to make the best use of the data and to investigate possible associations between various risk factors and vascular disease. Therefore, it could be said that first aim was achieved. However, interpretation of non-significant associations must be cautious. The Caerphilly study is not a large study and some of the non-significant results could be consistent with moderate effects.

**Goal 2:**
The second goal was to use multivariate analysis in order to link various risk factors with the risk of MI and stroke. Logistic regression and survival analysis using Cox's proportional hazard models were used here. In-depth analysis was carried out giving more emphasis on psychological risk factors. In the multivariate analysis, a comparison between men with low negative emotions (low anger out, low anxiety and low

274

depression) and men with high anger out expression only showed a higher risk of MI throughout the follow-up period. However, the addition of anxiety and depression to high anger out increases the risk (Figure 7.1).

The risk of stroke showed no significant association with the psychological factors. It could mean that either the stroke events were too few to detect any of the psychological factors' effects or there is no actual improvement of the model using the psychological factors for stroke.

**Goal 3:**

The next aim was to construct models similar to the Framingham risk models and to compare them with the classic Framingham models. This was also done successfully. The results suggested that the Framingham models over estimate the risk of MI and stroke at the prevalence cutoff points but the ROC curve analysis suggested that these models were not very dissimilar to the Caerphilly 1 models. These Caerphilly 1 models were based on the same variables as the Framingham models but constructed using the Caerphilly data. These results mean that a different risk cutoff should be administered when using the Framingham model for identifying men with high risk of MI and stroke in Caerphilly men or perhaps in Welsh and English men.

The newly constructed Caerphilly 2 models were significantly better predictors of MI and stroke. These were improved models based on the addition of extra risk factors (Anger-out and leisure activity for MI, alcohol consumption and re-defined smoking category for stroke). Using these newly constructed Caerphilly 2 models, the probability of having an MI for an individual can be estimated significantly better then the Framingham model; this could help in the early detection of VAD. However, these results should be used cautiously as these were constructed using the Caerphilly data and tested on the same data. In other words, these new models are tailored to the Caerphilly data. Therefore, these models should be applied to different populations in order to measure accurately their improved significance.

**Goal 4:**

The 4th goal was to carry out a literature review in order to identify plausible intervention scenarios. A literature review for each individual risk factor was performed to identify plausible intervention scenarios for reducing them. This aim was also achieved.

**Goal 5:**

The ultimate aim was to build a simulation model for each outcome to examine the effect of an intervention. Information from the intervention scenarios identified as part of the 4th goal was used in these simulation models. Various statistical techniques were used to simulate the risk factor datasets. Weibull regression models were used to simulate the survival times. Hypothetical interventions were first carried out on each individual risk factor in order to assess their impact on VAD. Several possible and some improbable scenarios were simulated to check the maximum possible event savings achieved by that particular risk factor.

This was followed by hypothetically intervening on every risk factor simultaneously in a multivariate scenario. These multivariate hypothetical intervention scenarios suggested that there could be a maximum of 50-55% MI saved if all the interventions were applied. If plausible interventions were applied to the hypothetical populations, there could be a saving of 25-30% MI events. For stroke, there appears to be reduction of more than 50% events if the same set of plausible hypothetical interventions for MI were applied. Though only alcohol consumption and SBP emerged as significant risk factors for stroke in the multivariate analysis for stroke (Section 7.4) excluding age and diabetes, there appears to be a higher reduction in events.

The important part of this simulation model is the hypothetical intervention to reduce anger-out in men in order to reduce the risk of MI. Depending on the intensity of the intervention that could be used to reduce anger-out; there is a possibility of reducing the number of MI events by between 6 and 31%. Assuming that there would be only 30% of the subjects accepts intervention and only 25% of them succeed, 2.4% MI events can be saved. Considering the fact that these intervention scenarios may not be entirely possible

and assuming such a low success level also fetches 2.4% reduction of MI events, there should me some plausible interventions scenarios needs to be identified. There are quite a few intervention scenarios for the standard biological and behavioural risk factors available in the literature. But there is little or no evidence to reduce psychological risk factors in order to reduce the risk of VAD. This is the grey area of the present interventions to reduce VAD therefore; more emphasis should be placed on these.

## 10.2 Limitations of the thesis

One of the major limitations of the Caerphilly study is that it consists of only men. The sample size for the study was 2398 at the start of the baseline. With missing values, and excluding men with prevalence IHD at the baseline, the effect size was even smaller. This is sample is small compared to other studies.

Considering the fact that the repeated measurements are needed for regression dilution bias and usually the follow-up reduces over the period of time, this further reduces the power of the analysis. However, it is important to note that the repeated measurements were apart from each other by 4-5 years. Methods used assume that the mean of the risk factor is constant over the period of time which is a big assumption as there is a big time gap between the repeated measurements and therefore, there is a high possibility to other variations to effect along with the measurement error.

Other issue is the lack of standard outcomes. There are various cardiovascular outcomes available in the literature and this makes comparisons with other studies rather difficult. For example, most of them had an outcome as CVD death as an outcome whereas this thesis deals with MI and stroke.

Advantages of simulations are many. Instead of spending time and money to find the effect sizes of an intervention, with the available literature, simulation models would measure the effect sizes. But there are many limitations. Primarily, models that are used

for the simulation should be relevant. These models depend on many assumptions which might not be valid all times.

## 10.3 Future work

Another dataset with Framingham anger-out measurements should be used to measure the risk of MI by the Caerphilly 2 models in order to estimate its true improvement of the model over the Framingham model. This would enhance the findings. The Caerphilly 2 Stroke model should be used on other Welsh or English populations because the role of smoking appears to be different in the Caerphilly and the Framingham models.

A larger dataset should be used to do a similar kind of analysis which would improve the level of accuracy. Other possibility is to impute various missing data that could be used to reduce the effects of low sample sizes encountered in the analysis. Much care needs to be taken while doing this because some of the risk factors are thought to be highly correlated (SBP & DBP). This analysis was limited to simulating hypothetical interventions on uncorrelated data as the Caerphilly data did not show high correlations between the risk factors.

These models could be extended to incorporate dynamic random effects of various repeated measurements and different time points. This would enhance the findings and also address the regression dilution bias in the multivariate scenarios.

These simulation models are incomplete to a certain extent if the cost effectiveness of these interventions is not brought into the model. Therefore, a literature review on the cost effectiveness of plausible intervention scenarios should be done. By doing that, the cost effectiveness of each intervention could be analysed and proper allocation of funds to reduce the risk of VAD could be done more effectively.

# References

1. Press Release, in *British Heart foundation*. 2000.
2. WHO, *World health report*. 2002(World Helath organisation).
3. Kannel WB, Dawber TR, Kagan A, *Factors of risk in the development of coronary heart disease: 6-year follow-up experience. The Framingham Study*. Intern Med, 1961. **55**: p. 33-50.
4. Grundy SM, Balady GJ, Criqui MH,, *Guide to primary prevention of cardiovascular diseases*. Circulation, 1997. **95**: p. 2329-31.
5. Castelli WP, Garrison RJ, Wilson PW, , *Incidence of coronary heart disease and lipoprotein cholesterol levels. The Framingham Study*. JAMA, 1986. **256**: p. 2835-8.
6. Stamler J, Wentworth D, Neaton JD. , *Is 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 Trial (MRFIT)*. JAMA, 1986. **256**: p. 2823-8.
7. Castelli WP, A.K., Wilson PW, Levy D, *Lipids and risk of coronary heart disease. The Framingham Study*. Ann Epidemiol, 1992. **2**: p. 23-28.
8. Iso H, Jacobs DR, Jr., Wentworth D, Neaton JD, Cohen JD., *Serum cholesterol levels and six-year mortality from stroke in 350,977 men screened for the multiple risk factor intervention trial*. N Engl J Med, 1989. **320**: p. 904-910.
9. Collaborative Research Group, *Blood pressure, cholesterol, and stroke in eastern Asia. Eastern Stroke and Coronary Heart Disease* Lancet, 1998. **352**: p. 1801-1807.
10. Rockville, M., *Reducing the health consequences of smoking: 25 years of progress*. DHHS publication, 1989. **CDC**: p. 89-8411.
11. Pooling Project Research Group, *Relationship of blood pressure, serum cholesterol, smoking habit, relative weight and ECG abnormalities to incidence of major coronary events: final report of the Pooling Project*. J Chronic Dis 1978. **31**: p. 2002-306.
12. Mulcahy. R, *Influence of cigarette smoking on morbidity and mortality after myocardial infarction*. Br Heart J 1983. **49**: p. 410-5.
13. Steenland K, *Passive smoking and the risk of heart disease*. JAMA, 1992. **267**: p. 94-99.
14. Kawachi I, *A prospective study of passive smoking and coronary heart disease*. Circulation, 1997. **95**: p. 2374-2379.
15. Poikolainen. K, *It can be bad for the heart, too-drinking patterns and coronary heart disease*. Addiction, 1993. **93**: p. 1757-1759.
16. Preddy, V.R., , Richardson, P. J.,, *Ethanol induced cardiovascular disease*. British Medical Bulletin, 1994. **50**: p. 152-163.
17. Ahlawat S.K, Siwach. S.B, *Alcohol and coronary artery disease*. International Journal of Cardiology, 1994. **44**: p. 157-162.
18. Marmot, M., , & Brunner, E,, *Alcohol and cardiovascular disease: the status of the U-shaped curve*. British Medical Journal, 1991. **303**: p. 565-568.

19. Foster WR, Burton BT,, *National Institutes of Health Consensus Development Panel on the health implications of obesity.* . Ann Intern Med, 1985. **103**(suppl 6): p. 1073-7.

20. Hubert HB, *The importance of obesity in the development of coronary risk factors and disease: the epidemiological evidence.* Annu Rev Public Health 1986. **7**: p. 493-502.

21. Linda C. Gallo, Shiva G. Ghaed, *Emotions and Cognitions in Coronary Heart Disease: Risk, Resilience, and Social Context.* . Cognitive Therapy and Research, 2004. **28**(5): p. 669-694.

22. Jenkins, C.D., Rosenman, *prediction of clinical CHD by a test of coronary prone behaviour pattren.* New England J Med, 1974. **290**: p. 1271-5.

23. Shapiro P, *Psychiatric aspects of cardiovascular disease.* Psychiatric Clinics of North America, 1996. **19**: p. 613-628.

24. Bryne, D.G., , Rosenman, R.H, *Anxiety and the Heart.* 1990, New York: Hemisphere Publishing Corporation.

25. Gallacher JEJ, Yarnell JWG, *Anger and Incident Heart Disease in the Caerphilly Study.* Psychosomatic medicine, 1999. **61**: p. 446–453.

26. Lazarus, R.S., , & Folkman, S, *Stress, Appraisal and Coping,* ed. Guilford. 1984, New York.

27. Friedman M, Powell LH, Thoresen CE. , *Effect of discontinuance of type-A behavioural counselling on type-A behaviour and cardiac recurrence rate of post myocardial infarction patients.* Am Heart J, 1987. **114**: p. 483-90.

28. Beswick A, Brindle P,, *Risk scoring in the assessment of cardiovascular risk.* Curr Opin Lipidol, 2006. **17**(4): p. 375-386.

29. The Caerphilly and Speedwell Collaborative Group, *Caerphilly and Speedwell collaborative heart disease studies.* . J Epidemiol Community Health, 1984. **38**: p. 259-262.

30. Group, T.C.a.S.C., *Caerphilly and Speedwell collaborative heart disease studies.* . J Epidemiol Community Health, 1984. **38**: p. 259-262.

31. Caerphilly and Speedwell collaborative heart disease studies, *The Caerphilly and Speedwell Collaborative Group* J Epidemiol Community Health, 1984. **38**: p. 259-262.

32. Fehily Ann M, M.J.E., Yarnell JWG, Hayes TM, Kubiki AJ, Eastham RD, , *2) Dietary determinants of lipoproteins, total cholesterol, viscosity, fibrinogen, and blood pressure.* Am J Clin Nutr, 1982. **36**: p. 890-896.

33. Stansfeld SA, S.D., Gallacher JEJ, Yarnell JWG, *A population survey of ischaemic heart disease and minor psychiatric disorder in men.* Psychol Med, 1992. **22**: p. 939–49.

34. Goldberg DP, *The detection of psychiatric illness by questionnaire.* 1972, London: Oxford University Press.

35. Marmot MG, *Stress, social and cultural variations in heart disease.* J Psychosom Res 1983. **27**: p. 377–84.

36. Taylor HL, J.D., Schacker B, Knudsen J, Leon AS, de Backer G,, *A questionnaire for the assessment of leisure time physical activities.* J Chron Dis, 1978. **31**: p. 741–55.

37.     Yarnell JWG, F.A., Millbank JE, Sweetnam PM, Walker CL, , *A short dietary questionnaire for use in an epidemiological survey: comparison with weighed dietary records*. Hum Nutr Appl Nutr, 1983. **37**: p. 103–12.

38.     Rose GA, *The diagnosis of ischaemic heart pain and intermittent claudication in field surveys*. Bull World Health Organ, 1962. **27**: p. 645–58.

39.     Speilberger, C.D., *Manual for the State-Trait Anxiety Inventory (Form Y)*, ed. P. Alto. 1983, CA: Consulting Psychologists Press.

40.     Haynes SG, L.S., Scotch N, Feinleib M, Kannel WB, *The relationship of psychosocial factors to coronary heart disease in the Framingham study. I. Methods and risk factors*. Am J Epidemiol, 1978. **107**: p. 362–82.

41.     Anastasi A, *Psychological testing*. 1963, New York.

42.     Goldberg D.P, B.B., *psychiatric illness in a suburban general practice. A detailed study using a new method of case identification*. BMJ, 1982. **439.**

43.     Stansfeld SA, Gallacher JEJ, Sharp DS, Yarnell JWG, *Social factors and minor psychiatric disorder in middle-aged men: a validation study and a population survey*. Psychol Med, 1991. **21**: p. 157-167.

44.     Jenkins CD, Z.S., Roseman RH,, *Jenkins Activity Survey Manual*. The Psychological Corporation. 1979, New York: Harcort Brace Jovanovich.

45.     Bortner RW, *A short rating scale as a potential measure of pattern*. A. J Chron Dis 1969. **22**: p. 87-91.

46.     Thomas K, *Attitudes and behaviour*. 1971, Middx, UK: Penguin.

47.     Gallacher JEJ, Yarnell JWG, *Attitudes and ischaemic heart diesease in the Caerphilly study: The health attitude invernory*. International Journal of Epidemiology, 1995. **24**(2): p. 321-330.

48.     Fishbein M, Ajzen I, *Belief, attitude, Intension and Behaviour. A introduction to theory and research*. MA. 1975, Reading: Addision-Wesley.

49.     Patrick Royston, Douglas G. Altman, *Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling*. Applied Statistics, 1994. **43**(3): p. 429-467.

50.     Peterson AV Jr, *Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions*. Journal of the American Statistical Association, 1977. **72**: p. 854-858.

51.     Kaplan EL, M.P., *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association, 1958. **53**: p. 457-481.

52.     Anderson, K.M., *A Nonproportional Hazards Weibull Accelerated Failure Time Regression Model*. Biometrics, 1991. **47**: p. 281-288.

53.     Keaven M. Anderson, PhD, Patricia M. Odell, PhD, Peter W. F. Wilson, MD, and and M. William B. Kannel, MPH, *Cardiovascular disease risk profiles*. Am Heart J, 1990. **121**: p. 293-8.

54.     Usha Govindarajulu, Lisa Sullivan, Ralph B. D' Agostino, *SAS Macro for confidence interval estimation for CHD risk prediction by different survival models*. Computational Statistics and Data Analysis, 2004. **46**: p. 571 - 592.

55.     Snedecor, G.W.a.C., W. G,, *Statistical Methods*. 6 ed. Ames. 1967: Iowa State University Press.

56. Haitao Chu, Zhaojie Wang, Stephen R. Cole, Sander Greenland, *Sensitivity analysis of misclassification: A graphical and Bayesian approach.* Ann Epidemiol, 2006. **16**: p. 834-841.

57. Patrick DL, Chaedle A, Thompson DC, *The validity of self repoted smoking: A review and meta-analysis.* Am J Public Health, 1994. **84**: p. 182-195.

58. Walker AM, Lanes SF, *Missclassification of covariates.* Statistics in Medicine, 1991. **10**: p. 1181-1196.

59. T Laatikainen, L Manninen, K Poikolainen, E Vartiainen, *Increased mortality related to heavy alcohol intake pattern.* J Epidemiol Community Health, 2003. **57**: p. 379-384.

60. Annie Britton, M.M., *Different measures of alcohol consumption and risk of coronary heart disease and all-cause mortality.* Addiction, 2004. **99**: p. 109-116.

61. T Laatikainen, L.M., K Poikolainen, E Vartiainen, *Increased mortality related to heavy alcohol intake pattern.* J Epidemiol Community Health, 2003. **57**: p. 379-384.

62. Djousse L, *Long-term alcohol consumption and the risk of atrial fibrillation in the Framingham Study.* Am J Cardiol, 2004. **93(6)**: p. 710-3.

63. Ajani UA, G.J., Lotufo PA, Liu S, Hennekens CH, Buring JE, Manson JE, *Alcohol consumption and risk of coronary heart disease by diabetes status.* Circulation, 2000. **102**: p. 500-505.

64. Jamrozik K, Broadhurst RJ, Forbes S, Hankey GJ, Anderson CS, *Predictors of death and vascular events in the elderly: the Perth Community Stroke Study.* Stroke, 2000. **31**: p. 863-8.

65. Shaper AG, Wannamethee G, Walker M., *Alcohol and mortality in British men: explaining the U-shaped curve.* Lancet, 1988. **2**: p. 1267-1273.

66. Giovonni Corrao, Luca Rubbiati, Vincenzo Baganardi, Antonella Zambon & Kari Poipaplinen, *Alcohol and coronary heart disease: a meta-analysis.* Addiction, 2000. **95(10)**: p. 1505-1523.

67. Giovanni Corrao, Ph.D., Vincenzo Bagnardi, Sc.D., Antonella Zambon, Sc.D., and Carlo La Vecchia, M.D, *A meta-analysis of alcohol consumption and the risk of 15 diseases.* Preventive Medicine, 2004. **38(2004)**: p. 613-619.

68. Gorelick PB, *The status of alcohol as a risk factor for stroke.* Stroke, 1989. **20**: p. 1607-1610.

69. Mazzaglia G, Britton AR, Altmann DR, Chenet L, *Exploring the relationship between alcohol consumption and non-fatal or fatal stroke: a systematic review.* Addiction, 2001. **96**: p. 1743-1756.

70. Gill JS, Zezulka AV,, *stroke and alcohol consumption.* N Engl J Med, 1986. **315**: p. 1041-1046.

71. Gill JS, Shipley MJ,, *Alcohol consuption - a risk factor for hemorrhagic and non-hemorrhagic stroke.* Am J Med, 1991. **90**: p. 489-497.

72. Sacco RL, Lkind M, Kargman DE,, *The protective effect of moderate alcohol consumption on ischemic stroke.* JAMA, 1999. **281**: p. 53-60.

73. Gaziano JM, Gaziano TA, Glynn RJ, Sesso HD, Ajani UA, Stampfer MJ, Manson JE, Hennekens CH, Buring JE, *Light-tomoderate alcohol consumption and mortality in the Physicians' Health Study enrollment cohort.* J Am Coll Cardiol, 2000. **35**: p. 96-105.

74. Reynolds K, Lwesi B, Nolen JD,, *Alcohol consumption and risk of stroke. A meta analysis.* JAMA, 2003. **289**: p. 579-588.

75. Elwood PC, *Smoking, drinking, and other life style factors and cognitive function in men in the Caerphilly cohort.* J Epidemiol Community Health, 1999. **53**(1): p. 9-14.

76. Fehily AM, *Diet and incident ischaemic heart disease: the Caerphilly Study.* Br J Nutr, 1993. **69**(2): p. 303-314.

77. Renaud SC, *Alcohol and platelet aggregation: the Caerphilly Prospective Heart Disease Study.* Am J Clin Nutr, 1992. **55**(5): p. 1012-1017.

78. Annie Britton, Michael Marmot, *Different measures of alcohol consumption and risk of coronary heart disease and all-cause mortality.* Addiction, 2004. **99**: p. 109–116.

79. Arthur L. Klatsky, Mary Anne Armstrong, Gary D. Friedman,, *Relations of Alcoholic Beverage Use to Subsequent Coronary Artery Disease Hospitalization.* Am J Cardiol, 1986. **58**: p. 710-714.

80. Carole L Hart, George Davey Smith, David J Hole, Victor M Hawthorne, *Alcohol consumption and mortality from all causes, coronary heart disease, and stroke: results from a prospective cohort study of Scottish men with 21 years of follow up.* BMJ, 1999. **318**: p. 1725-1729.

81. S. Goya Wannamethee, PhD; A. Gerald Shaper, FRCP *Patterns of Alcohol Intake and Risk of Stroke in Middle-aged British Men* Stroke, 1996. **27**: p. 1033-1039.

82. Klatsky AL, Armstrong MA, Friedman GD, *Relations of alcoholic beverage use to subsequent coronary artery disease hospitalization.* Am J Cardiol, 1986. **58**: p. 710-4.

83. www.nida.gov. *Nicotine.* 2007 [cited 2007 26/7]; Available from: http://nida.gov/researchreports/nicotine/.

84. Doll R, P.R., *mortality in relation to smoking: 50 years' observations on male british doctors.* BMJ, 2004. **328**: p. 1519-1528.

85. Edwards R, *The problem of tobacco smoking. ABC of smoking cessation.* BMJ, 2004. **328**: p. 217-219.

86. Hammond EC, Garfinkel L, Seidman H, Lew EA, *"Tar" and nicotine content of cigarette smoke in relation to death rates. Environmental Research.* 1976. **12**(3): p. 263-274.

87. Borland C, *carbon monoxide yield of cigarettes and its realtion to CAD.* BMJ, 1983. **287**(6405): p. 1583-1586.

88. Higenbottam, *Cigarettes, Lung Cancer, and CHD: the effects of inhalation and Tar yields.* Jour of epidem and Commun Health, 1982. **36**(2): p. 113-7.

89. Carole L. Hart, David J. Hole and George Davey Smith, *Risk Factors and 20-Year Stroke Mortality in Men and Women in the Renfrew/Paisley Study in Scotland.* Stroke, 1999. **30**: p. 1999-2007.

90. Tobias Kurth, Carlos S. Kase, Klaus Berger, Elke S. Schaeffner, Julie E. Buring, *Smoking and the Risk of Hemorrhagic Stroke in Men.* stroke, 2003. **34**: p. 1151-1155.

91. McCarron P, *The incidence and aetiology of stroke in the Caerphilly and Speedwell Collaborative Studies II: risk factors for ischaemic stroke.* Public Health, 2001. **115**(1): p. 12-20.

92. Doll R, Peto R, *mortality in relation to smoking: 50 years' observations on male british doctors*. BMJ, 2004. **328**: p. 1519-1528.

93. Chave SPW, Morris JN, Moss S, Semmence AM,, *Vigorous exercise in leisure time and the death rate: a study of male civil servants*. J Epidemiol Community Health, 1978. **32**: p. 239–43.

94. Chave SPW, M.J., Moss S, Semmence AM,, *Vigorous exercise in leisure time and the death rate: a study of male civil servants*. J Epidemiol Community Health, 1978. **32**: p. 239–43.

95. Leon AS, C.J., Jacobs DR Jr, Raurama R, *Leisure-time physical activity levels and risk of coronary heart disease and death: the Multiple Risk Factor Intervention trial*. JAMA, 1987. **258**: p. 2388–95.

96. Morris JN, C.D., Everitt MG, Semmence AM, Burgess EH. , *Exercise in leisure time: coronary attack and death rates*. Br Heart J, 1990. **63**: p. 325–34.

97. Paffenbarger RS Jr, H.R., Wing AL, Hsieh CC,, *Physical activity, all cause mortality and longevity of college alumni*. N Engl J Med, 1986. **304**: p. 605–13.

98. Shaper AG, W.G., Weatherall R, *Physical activity and ischaemic heart disease in middle-aged men*. Br Heart J, 1991. **66**: p. 384–94.

99. Slattery ML, J.D.J., Nichaman MZ, *Leisure time physical activity and coronary heart disease death: the US Railroad Study*. Circulation, 1989. **79**: p. 304-11.

100. Wannamethee G, S.A., *Physical activity and risk of stroke in British middle-aged men*. BMJ, 1992. **304**: p. 597–601.

101. Kaplan GA, S.T., Cohen RD, Knudsen LP, Guralnik J, *Mortality among the elderly in the Alameda Count Study: behavioural and demographic risk factors*. Am J Public Health, 1987. **77**: p. 307–12.

102. Sherman SE, D.A.R., Cobb JL, Kannel WB., *Does exercise reduce mortality rates in the elderly? Experience from the Framingham Heart study.* . Am Heart J, 1994. **128**: p. 965–72.

103. Shaper AG, W.G., Walker M, *Physical activity, hypertension and risk of heart attack in men without evidence of ischaemic heart disease*. J Hum Hypertens, 1994. **8**: p. 3-10.

104. Wannamethee SG, Shaper AG,Walker M., *Changes in physical activity, mortality and incidence of coronary heart disease in older men*. Lancet 1998, 1998. **351**: p. 1603-8.

105. James C. Puffer, *Exercise and Heart Disease*. SPORTS MEDICINE, 2002. **3**(5).

106. Fang J, Wylie-Rosett J, Cohen HW, *Exercise, body mass index, caloric intake, and cardiovascular mortality*. Am J Prev Med, 2003. **25**: p. 283–9.

107. Gang Hua, *Joint effects of physical activity, body mass index, waist circumference and waist-to-hip ratio with the risk of cardiovascular disease among middle-aged Finnish men and women*. European Heart Journal, 2004. **25**: p. 2212–2219.

108. GCW Wendel-Vos, AJ Schuit, EJM Feskens, HC Boshuizen, WMM Verschuren, WHM Saris, D Kromhout, *Physical activity and stroke. A meta-analysis of observational data*. International Journal of Epidemiology, 2004. **33**: p. 787–798.

109. Taylor HL, Jacobs DR, Schacker B, Knudsen J, Leon AS, de Backer G,, *A questionnaire for the assessment of leisure time physical activities*. J Chron Dis, 1978. **31**: p. 741–55.

110. S Yu, J W G Yarnell, P M Sweetnam, L Murray, *What level of physical activity protects against premature cardiovascular death? The Caerphilly study* Heart 2003. **89**: p. 502-506.

111. Shaper AG, Wannamethee G, Weatherall R, *Physical activity and ischaemic heart disease in middle-aged men.* Br Heart J, 1991. **66**: p. 384–94.

112. Fuller W.A, *Measurement error models.* Wiley. 1987, New York.

113. William B. Kannel, *Risk Stratification in Hypertension: New Insights From the Framingham Study.* AJH, 2000. **13**: p. 3S–10S.

114. Wilhelmsen, *Heart failure in the general population of men morbidity, risk factors and prognosis.* J Intern Med, 2001. **249**: p. 253-261.

115. Howard D. Sesso, M.J.S., Bernard Rosner, Charles H. Hennekens, J. Michael Gaziano, JoAnn E. Manson, Robert J. Glynn, *Systolic and Diastolic Blood Pressure, Pulse Pressure, and Mean Arterial Pressure as Predictors of Cardiovascular Disease Risk in Men.* Hypertension, 2000. **36**: p. 801-807.

116. Benetos A, Thomas F, BeanK,Gautier S,, *Systolic blood pressure is a better predictor of cardiovascular disease risk than diastolic blood pressure in hypertensive men.* Evidence-based CardiovascularMedicine, 2002. **162**: p. 577-581.

117. A Menotti, M Lanti, P E Puddu, D Kromhout, *A systematic reanalysis of 10 year coronary heart disease incidence data from the northern and the southern European cohorts of the seven countries.* Heart, 2000. **84**: p. 238-244.

118. Sarah Lewington, Robert Clarke, Nawab Qizilbash, Richard Peto, Rory Collins, *Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies.* LANCET, 2002. **360**: p. 1903-1913.

119. Robert Clarke, Martin Shipley, Richard Peto, *Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies.* Am J Epidemiol, 1999. **150**(4): p. 341-353.

120. Viola Vaccarino, MD, PhD, Alan K. Berger, MD, Jerome Abramson, PhD, Henry R. Black, MD, John F. Setaro, MD, Janice A, , *Pulse Pressure and Risk of Cardiovascular Events in the Systolic Hypertension in the Elderly Program.* Am J Cardiol, 2001. **88**: p. 980–986.

121. Franklin SS, K.S., Wong ND, Larson MG, Levy D,, *Is pulse pressure useful in predicting risk for coronary heart Disease? The Framingham heart study.* Circulation, 1999. **100**(4): p. 354-360.

122. Howard D. Sesso, Meir J. Stampfer, Bernard Rosner, Charles H. Hennekens, J. Michael Gaziano, JoAnn E. Manson, Robert J. Glynn, *Systolic and Diastolic Blood Pressure, Pulse Pressure, and Mean Arterial Pressure as Predictors of Cardiovascular Disease Risk in Men.* Hypertension, 2000. **36**: p. 801-807.

123. Fabio Angeli, Paolo Verdecchia, Gian Paolo Reboldi,, *Calcium Channel Blockade to Prevent Stroke in Hypertension.* AJH, 2004. **17**: p. 817–822.

124. Millar JA, Lever AF, Burke V. , *Pulse pressure as a risk factor for cardiovascular events in the MRC Mild Hypertension Trial.* J Hypertens Suppl, 1999. **17**: p. 1065–1072.

125. Corti M, G.J., Salive ME, Harris T, Ferrucci L, Glynn RJ, , *Clarifying the direct relation between total cholesterol levels and death from coronary heart disease in older persons*. Ann Intern Med, 1997. **126**: p. 753-760.

126. Tervahauta M, P.J., Nissinen A, *Risk factors of coronary heart disease and total mortality among elderly men with and without pre-existing coronary heart disease. Finnish cohorts of the Seven Countries Study*. J Am Coll Cardiol, 1995. **26**: p. 1623–1629.

127. Reed. D, Benfante. R,, *Lipid and lipoprotein predictors of coronary heart disease in elderly men in the Honolulu Heart Program*. Ann Epidemiol, 1992. **2**: p. 29–34.

128. Weijenberg MP, F.J., Kromhout D, *Total and high density lipoprotein cholesterol as risk factors for coronary heart disease in elderly men during 5 years of follow-up. The Zutphen Elder Study*. Am J Epidemiol, 1996. **143**: p. 151–158.

129. Houterman S, M.V.W., Giampaoli S, Nissinen A, Feskens EJM, Menotti A,, et al. *Total but not high-density lipoprotein cholesterol is consistently associated with coronary heart disease mortality in elderly men in Finland, Italy, and the Netherlands*. Epidemiology, 2000. **11**: p. 327–332.

130. EMMANUEL A. ANUM, T.A., *Hypercholesterolemia and Coronary Heart Disease in the Elderly: A Meta-analysis*. Ann Epidemiol, 2004. **14**: p. 705–721.

131. Tervahauta M, Pekkanen J, Nissinen A, *Risk factors of coronary heart disease and total mortality among elderly men with and without pre-existing coronary heart disease. Finnish cohorts of the Seven Countries Study*. J Am Coll Cardiol, 1995. **26**: p. 1623–1629.

132. Houterman S, Monique Verschuren WM, Giampaoli S, Nissinen A, Feskens EJM, Menotti A,, et al. *Total but not high-density lipoprotein cholesterol is consistently associated with coronary heart disease mortality in elderly men in Finland, Italy, and the Netherlands*. Epidemiology, 2000. **11**: p. 327–332.

133. Weijenberg MP, Feskens JM, Kromhout D, *Total and high density lipoprotein cholesterol as risk factors for coronary heart disease in elderly men during 5 years of follow-up. The Zutphen Elder Study*. Am J Epidemiol, 1996. **143**: p. 151–158.

134. Iso H, J.D., Jr., Wentworth D, Neaton JD, Cohen JD., *Serum cholesterol levels and six-year mortality from stroke in 350,977 men screened for the multiple risk factor intervention trial*. N Engl J Med, 1989. **320**: p. 904–910.

135. Sacco RL, B.R., Kargman DE,, *High-density lipoprotein cholesterol and ischemic stroke in the elderly: the Northern Manhattan troke Study*. JAMA, 2001. **285**: p. 2729–2735.

136. Bots ML, E.P., Nikitin Y,, *Total and HDL cholesterol and risk of stroke. EUROSTROKE: a collaborative study among research centres in Europe*. J Epidemiol Community Health, 2002. **56**(1): p. 119–24.

137. Shahar E, C.L., Rosamond WD, , *Plasma lipid profile and incident ischemic stroke: the Atherosclerosis Risk in Communities (ARIC) study*. Stroke, 2003. **34**: p. 623–631.

138. Prospective Studies Collaboration, *Cholesterol, diastolic blood pressure, and stroke: 13 000 strokes in 450 000 people in 45 prospective cohorts*. Lancet, 1995. **346**: p. 1647-1653.

139.    Corti M, Guralnik JM, Salive ME, Harris T, Ferrucci L, Glynn RJ, , *Clarifying the direct relation between total cholesterol levels and death from coronary heart disease in older persons.* Ann Intern Med, 1997. **126**: p. 753-760.

140.    Hubert HB, F.M., McNamara PM,, *Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study.* Circulation, 1983. **67**: p. 968–77.

141.    Daniell L Mcgee, *Body Mass Index and Mortality: A Meta-analysis Based on Person-level Data from Twenty-six Observational Studies.* Ann Epidemiol, 2005. **15**: p. 87–97.

142.    Kurth T, Gaziano JM, Berger K, Kase CS, Rexrode KM, Cook NR, Buring JE, Manson JE, *Body mass index and the risk of stroke in men.* arch Intern Med, 2003. **163**(8): p. 980.

143.    Katarina Jood, MD; Christina Jern, MD, PhD; Lars Wilhelmsen, MD, PhD; Annika Rosengren, MD, PhD, *Body Mass Index in Mid-Life Is Associated With a First Stroke in Men: A Prospective Population Study Over 28 Years.* Stroke, 2004. **35**: p. 2764-2769.

144.    Ariyo AA, H.M., Tangen CM,, *Depressive symptoms and risks of coronary heart disease and mortality in elderly Americans.* Circulation, 2000. **102**: p. 1773-9.

145.    Ferketich AK, S.J., Frid DJ, *Depression as an antecedent to heart disease among women and men in the NHANES I study.* Arch Intern Med, 2000. **160**: p. 1261-8.

146.    Gaziano JM, G.T., Glynn RJ, Sesso HD, Ajani UA, Stampfer MJ, Manson JE, Hennekens CH, Buring JE, *Light-tomoderate alcohol consumption and mortality in the Physicians' Health Study enrollment cohort.* J Am Coll Cardiol, 2000. **35**: p. 96-105.

147.    Vogt, T., , Mullooly, J, *Mental health status as a predictor of morbidity and mortality: a15 years follow-up of members of health maintenance organization.* American Journal of Public Health, 1994. **84**: p. 227-31.

148.    Cole, S.R., , Kawachi, I., Sesso, H. D.,, *Sense of exhaustion and coronary heart disease among college alumni.* American journal of cardiology, 1999. **84**: p. 1401-5.

149.    Chang M, Hahn RA, Teutsch SM, *Multiple risk factors and population attributable risk for ischeamic heart disease mortality in the United States, 1971 - 1992.* Journal of Clinical epidemiology, 2001. **54**: p. 634-44.

150.    Ford DE, M.L., *Depression is a risk factor for coronary heart disease in men: the precursors study.* Arch Intern Med, 1998. **158**: p. 1422-6.

151.    Wade DT, Legh-Smith J, Hewer RA, *Depressed mood after stroke: a community study of its frequency.* Br J Psychiatry, 1987. **151**: p. 200–205.

152.    Colantonio A, Kasi SV, Ostfeld AM, *Depressive symptoms and other psychosocial factors as predictors of stroke in the elderly.* Am J Epidemiol, 1992. **136**: p. 884-894.

153.    Everson SA, Roberts RE, Goldberg DE, Kaplan GA, *Depressive symptoms and increased risk of stroke mortality over a 29-year period.* Arch Intern Med, 1998. **158**: p. 1133–1138.

154.    Simonsick EM, Wallace RB, Blazer DG, Berkman LF, *Depressive symptomatology and hypertension-associated morbidity and mortality in older adults.* Psychosomatic medicine, 1995. **57**: p. 427–435.

155. Julia Hippisley-Cox, C.C., Yana Vinogradova, John Robson, Margaret May, Peter Brindle, *Derivation of Validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study.* BMJ, 2007.

156. Bashir. S. A, Duffy. S. W, *The Correction of Risk Estimates for Measurement Error.* Ann Epidemiol, 1997. **7**: p. 154-1 64.

157. Spielberger, C.D., & Rickman, R.L, *Assessment of State and Trait Anxiety in Cardiovascular Disorders.* 1990, New York: Hemisphere Publishing Corporation.

158. Todaro JF, *Effect of negative emotions on frequency of coronary heart disease (The Normative Aging Study).* Am J Cardiol, 2003. **92**(8): p. 901-6.

159. MARIANNE HAGMAN, L.W., HANS WEDEL, KJELL PENNERT, *RISK FACTORS FOR ANGINA PECTORIS IN A POPULATION STUDY OF SWEDISH MEN.* J Chron Dis, 1987. **40**(3): p. 265-275.

160. A P HAINES, J.D.I., T W MEADE, *Phobic anxiety and ischaemic heart disease.* BMJ, 1987. **295**: p. 297-299.

161. I Kawachi, G.C., A Ascherio, EB Rimm, E Giovannucci, MJ Stampfer, WC Willett, *Prospective study of phobic anxiety and risk of coronary heart disease in men.* circulation, 1994. **89**: p. 1992-1997.

162. Jonas, B.S., Frank, P. Ingram, D.D., *Are symptoms of anxiety and depression risk factors for hypertension? Longitudinal evidence from the National Health and Nutrition Examination Survey I: Epidemiologic Follow-up Study.* Arch. Fam. Med. , 1997. **5**: p. 44–48.

163. Matthews KA, Glass DC, Rosenmen RH, *Competitive drive, Pattern A, and coronary heart disease: a further analysis of some data from the western colloborative study.* Journal of Chronic disease, 1977. **30**: p. 489-98.

164. Shekelle RB, Gale M, Ostfeld AM, *Hostility, risk of coronary heart disease and mortality.* psychosomatic medicine, 1983. **45**: p. 59-63.

165. Barefoot JC, Dahlstrom WC, Williams RB, *CHD incidence and total mortality: a 25 years follow-up study of 255 physicians.* Psychosomatic medicine, 1983. **45**: p. 59-63.

166. Leon GR, Finn SE, Murray DM, *inability ot predict cardiovascular disease from hostility score or MMPI items related ot Type A behaviour.* J Consult Clin Psychol, 1988. **56**: p. 597-600.

167. Hearn MD, Murray DM, Luepker RV, *Hostility, coronary heart disease and total mortality: a 33 year follow-up study of university students.* Journal of behavoural medicine, 1989. **12**: p. 105-21.

168. Everson SA, Kauhanen J, Kaplan GA, Goldberg DE, Julkunen J, Tuomilehto and S.J. J, *Hostility and increased risk of mortality and acute myocardial infarction: the mediating role of behavioral risk factors.* Am J Epidemiol, 1997. **146**: p. 142–152.

169. Cook WW, Medley DM, *Proposed hostility and pharisaic-virtue scales for the MMPI.* J Appl Psychol, 1954. **38**: p. 414–8.

170. Ulla Christensen, M., PHD, Rikke Lund, MD, PHD, Mogens Trab DAMSGAARD, MAGSCIENTSOC, *Cynical Hostility, Socioeconomic Position, Health Behaviors, and Symptom Load: A Cross-Sectional Analysis in a Danish Population-Based Study.* psychosomatic medicine, 2004. **66**: p. 572-577.

171. Millar TQ, Smith TW, Turner CW, *A meta-analytic review of research on hostility and physical health.* psychological Bulliten, 1996. **142**: p. 477-84.

172. Rosenmen, R.H., *multivariate prediction of CHD during 8.5 year follow-up in western colloborative group study.* American journal of cardiology, 1976. **37**: p. 903-9.

173. Haynes SG, Levine S, Scotch N, Feinleib M, Kannel WB, *The relationship of psychosocial factors to coronary heart disease in the Framingham study. I. Methods and risk factors.* Am J Epidemiol, 1978. **107**: p. 362–82.

174. Hollis, J.F., *stressful life events, type A behaviour, and the prediction of cardiovascular and total mortality over 6 years.* journal of behavoural medicine, 1990. **13**: p. 263-80.

175. Bosma, H., *psychological factors in the eatiology of coronary heart disease: follow-up to the Kaunas-Rotterdam intervention study (KRIS).* cardiology, 1995. **2**: p. 54-59.

176. DR Ragland, R.B., *Type A behavior and mortality from coronary heart disease.* N Engl J Med, 1988. **318**: p. 65-69.

177. Gallacher, J., *RE. "CORONARY HEART DISEASE MORTALITY IN THE WESTERN COLLABORATIVE GROUP STUDY: FOLLOW-UP EXPERIENCE OF 22 YEARS".* Am J Epidemiol, 1989. **130**(2): p. 429-30.

178. David Melzer, Brenda McWilliams, Carol Brayne, Tony Johnson, John Bond, *Socioeconomic status and the expectation of disability in old age: estimates for England The Medical Research Council Cognitive Function and Ageing Study (MRC CFAS) writing committee.* J Epidemiol Community Health, 2000. **54**: p. 286-292.

179. WB Kannel, DL McGee, *Diabetes and cardiovascular risk factors: the Framingham study.* circulation, 1979. **59**: p. 8-13.

180. Mc Carron P, Greenwood R, Elwood P, Ben Shlomo Y, Bayer A, Baker I, Frankel S, Ebrahim S, Murray L, Davey Smith G, *The incidence and aetiology of stroke in the Caerphilly and Speedwell Collaborative studies. II Risk factors for ischaemic stroke.* . Public Health, 2001. **115**: p. 12-20.

181. EMMANUEL A. ANUM, TILAHUN ADERA,, *Hypercholesterolemia and Coronary Heart Disease in the Elderly: A Meta-analysis.* Ann Epidemiol, 2004. **14**: p. 705–721.

182. Stansfeld S, Rasul F, *Mental illness and coronary heart disease. In: Marmot M, Elliott P, editors. Coronary heart disease epidmiology: from aetiology to public health.* Oxford, 2005. **Oxford University Press**: p. 414-431.

183. Rasul F, Stansfeld SA, Hart CL, Davey SG, *Psychological distress, physical illness, and risk of coronary heart disease.* J Epidemiol Community Health, 2005. **59**(2): p. 140-145.

184. Anderson DE, Metter EJ, Hougaku H, Najjar SS, *Suppressed anger is associated with increased carotid arterial stiffness in older adults.* Am J Hypertension, 2006. **19**(11): p. 1129-1134.

185. Eng PM, Fitzmaurice G, Kubzansky LD, Rimm EB, Kawachi I, *Anger expression and risk of stroke and coronary heart disease among male health professionals.* Psychosomatic medicine, 2003. **65**(1): p. 100-110.

186. Julius M, Harburg E, Cottington EM, Johnson EH, *Anger-coping types, blood pressure, and all-cause mortality: a follow-up in Tecumseh, Michigan (1971-1983).* Am J Epidemiol, 1986. **124**(2): p. 220-233.

187. Painuly N, Sharan P, Mattoo SK, *Relationship of anger and anger attacks with depression: a brief review.* Eur Arch Psychiatry Clin Neurosci, 2005. **255**(4): p. 215-222.

188. Wattanakit K, Williams JE, Schreiner PJ, Hirsch AT, Folsom AR, *Association of anger proneness, depression and low social support with peripheral arterial disease: the Atherosclerosis Risk in Communities Study.* Vasc Med, 2005. **10**(3): p. 199-206.

189. Preisig M, Merikangas KR, Angst J, *Clinical significance and comorbidity of subthreshold depression and anxiety in the community.* Acta Psychiatr Scand, 2001. **104**(2): p. 96-103.

190. Jackie A. Cooper, George J. Miller, Steve E. Humphries, *A comparison of the PROCAM and Framingham point-scoring systems for estimation of individual risk of coronary heart disease in the Second Northwick Park Heart Study.* Atherosclerosis, 2005. **181**: p. 93–100.

191. Bhargava, A., *A longitudinal analysis of the risk factors for diabetes and coronary heart disease in the Framingham Offspring Study.* Population Helath Metrics, 2003. 1(3): p. 1-10.

192. Peter Brindle, Jonathan Emberson, Fiona Lampe, Mary Walker, *Predective accuracy of the Framingham coronary risk socre in British men: prospective cohort study.* BMJ, 2003. **327**: p. 1267-1273.

193. Hans-Werner Hense, Helmut Schulte, Hannelore Lo¨wel, Gerd Assmann, Ulrich Keil, *Framingham risk function overestimates risk of coronary heart disease in men and women from Germany—results from the MONICA Augsburg and the PROCAM cohorts.* European Heart Journal, 2003. **24**: p. 937–945.

194. R.M. Conroy, K., , Pyorala, A.P. Fitzgerald, S. Sans, A. Menotti, G. De Backer, D. De Bacquer, P. Ducimetiere, P. Jousilahti, Keil, I. Njølstad, R.G. Oganov, T. Thomsen, H. Tunstall-Pedoe, A. Tverdal, H. Wedel, P. Whincup, L. Wilhelmsen, I.M. Graham,, *Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project.* European Heart Journal, 2003. **24**: p. 987–1003.

195. Joan Cristobal, Fernando Lago, Javier de la Fuente, Jose R. Gonzalez-Juanatey, *Comparison of Cornary Risk estimates Derived Using the Framingham and REGICOR Equations.* Rev Esp Cardiol, 2005. **58**(8): p. 910-5.

196. Jackson R, Barham P,Bills J,, *Management of raised blood pressure in New Zealand: a discussion document.* BMJ, 1993. **307**: p. 107-10.

197. Jackson R, *Updated New Zealand cardiovascular disease risk-benefit prediction guide.* BMJ, 2000. **320**(7236): p. 709-10.

198. Pyorala, K., De Backer, G., Graham, I., Poole-Wilson, P.A., Wood, D.,, *Prevention of CHD in clinical practice: recommendations of the task force of the European society of Cardiology, European Atherosclerotic Society and European Society of Hypertension.* European Heart Journal, 1994. **15**: p. 1300-31.

199. Wood, D., De Backer, G., Faergeman, O., Graham, I., Mancia, G., Pyorala, K, *Prevention of CHD is clinical practice: recommendations of the second joint task*

*force and other societies on coronary prevention.* European Heart Journal, 1998a. **19**: p. 1434-503.

200. Wood, D., De Backer, G., Faergeman, O., Graham, I., Mancia, G., Pyorala, K, *Prevention of CHD in clinical practice: recommendations of the second joint task force and other societies on coronary prevention.* Atherosclerosis, 1998b. **140**(2): p. 199-270.

201. Haq, I.U., Jackson, P.R., Yeo, W.W., Ramsay, L.E.,, *Sheffield risk and treatment table for cholesterol lowering for primary prevention of CHD.* Lancet, 1995. **346**: p. 1467-71.

202. *Joint British recommendations on prevention on CHD in clinical practice.* 2000.

203. Wallis, E.J., , Ramsay, L. E., Haq, I. U.,, *Coronary and cardiovascular risk estimation for primary prevention: validation of a new Sheffield table in the 1995 scottish health survey population.* BMJ, 2000. **320**: p. 671-6.

204. Wood, D., Durrington, P., Poulter, N., McInnes, G., Rees, A., Wray, R. on behalf of the British Cardiac society, British Hyperlipidaemia Association, British hypertension society and British diabetic association, *Joint British recommendations on prevention of coronary heart disease in clinical prectice.* Heart, 1998c. **80**(Suppl. 2): p. S1-S29.

205. Julia Hippisley-Cox, Carol Coupland, Yana Vinogradova, John Robson, Margaret May, Peter Brindle, *Derivation of Validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study.* BMJ, 2007.

206. Miller GJ, Bauer KA, Barzegar S, Cooper JA, Rosenberg RD, *Increase dactivation of the haemostatic system in men at high risk for fatal coronary heart disease.* Thromb Haemost, 1996. **75**: p. 767–71.

207. Haq, I.U., Jackson, P.R., Yeo, W.W., Ramsay, L.E.,, *Is FRamingham risk function valid for northern European populations? A comparision of methods for estimating absolute coronary risk in high risk men.* Heart, 1999. **81**: p. 40-6.

208. Pyorala, K., *Assesment of coronary heart disease in the populations with different levels of risk.* european Heart Journal, 2000. **21**(5): p. 348-50.

209. Menotti, A., Lanti, M., Puddu, P. E., Kromhout, D.,, *CHD incidence in northern and southern European populations: a reanalysis of the Seven Countries study for a European coronary risk chart.* Heart, 2000. **84**: p. 238-44.

210. De Backer, G., Ambrosioni, E., Borch-Johnsen, K.,, *European guidelines on CVD prevention in clinical practice. Executive summary. Third joint task force of European and other societies on CVD prevention in clinical practice.* European Journal of Cardiovascular prevention and Rehabilitation, 2003. **10**(Suppl. 1): p. S2-S78.

211. Cheng, R.C.H., , and G. M. Feast, *Some simple gamma variate generators.* Applied Statistics, 1979. **28**: p. 290–295.

212. Ahrens, J.H., Dieter, U, *Computational Methods for sampling from gamma, beta, Poisson, and binomial distributions.* Computing, 1974. **12**: p. 223-246.

213. Best, D.J., *A note on gamma variate generators with shape parameter less than unity.* Computing, 1983. **30**: p. 185-188.

214. Ralf Bender, Thomas Augustin, Maria Blettner, *Generating survival times to simulate Cox's Proportional Hazard models*. Statistics in Medicine, 2005. **24**: p. 1713-1723.

215. Richard Saitz, M., MPH; Nicholas J. Horton, ScD; Lisa M. Sullivan, PhD; Mark A. Moskowitz, MD, *Alcohol Problems in Primary Care: A Cluster Randomized, Controlled Trial of a Systems Intervention*. Ann Intern Med, 2003. **138**: p. 372-382.

216. Michael F. Fleming, M.P.M., *Brief Physician Advice for Problem Drinkers: Long- Term Efficacy and Benefit-Cost Analysis*. Alcohol Clin Exp Res, 2002. **26**(No 1): p. 36-43.

217. Anders Hammarberg, *A Comparision of two intensities of psychological intervention for alcohol dependent patients*. Alcohol & Alcoholism, 2004. **39**(No 3): p. 251-255.

218. S. Linke, A.B.a.P.W., *Down your drinking: A web based intervention for people with excessive alcohol consumption*. Alcohol & Alcoholism, 2004. **39**(No 1): p. 29-32.

219. Home Health UK. 20/11/2006 [cited; Available from: http://www.homehealth-uk.com/medical/alcohol.htm.

220. Ames RP, Hill P, *Increase in serum lipids during treatment of hypertension with chlorthalidone*. Lancet, 1976. **1**: p. 721-723.

221. Domenic A. Sica, M., *Calcium Channel Blocker Class Heterogeneity: Select Aspects of Pharmacokinetics and Pharmacodynamics* hypertension, 2005. **7**(1): p. 21-26

222. Kristina Chen, Daniel Glass, Chiun-Fang Chiou, Craig Plauschinat, Feride Frech, Ann Harper and Robert Dubois, *Adverse effects and reasons for switch reported by patients on antihypertensive medications*. Am J Hypertension, 2004. **17**(5): p. 102A.

223. Fabio Angeli, P.V., Gian Paolo Reboldi,, *Calcium Channel Blockade to Prevent Stroke in Hypertension*. AJH, 2004. **17**: p. 817–822.

224. Jan A. Staessen, J.-G.W., Lutgarde Thijs. , *Cardiovascular prevention and blood pressure reduction:a quantitative overview updated until 1 March 2003*. Journal of Hypertension 2003, 2003. **21**(6).

225. Neutel JM, *Metabolic manifestations of low-dose diuretics*. Am J Med, 1996. **101**: p. 71S-82S.

226. LaCroix AZ, L.J., Scherr P *Smoking and mortality among older men and women in three communities*. N Engl J Med, 1991. **324**: p. 1619–1625.

227. Rosenberg L, K.D., Helmrich SP, Shapiro S, *The risk of myocardial infarction after quitting smoking in men under 55 years of age*. N Engl J Med, 1985. **313**: p. 1511–1514.

228. Tverdal A, T.D., Stensvold I,, *Mortality in relation to smoking history: 13 years' follow-up of 68,000 Norwegian men and women 35–49 years*. J Clin Epidemiol, 1993. **46**: p. 475–487.

229. Fiore MC, S.S., Jorenby AE, Baker TB, *The effectiveness of the nicotine patch for smoking cessation.Ameta-analysis*. JAMA, 1994. **271**: p. 1940-1947.

230. Schneider NG, L.E., Olmstead RE, Fagerstrom KO, *Clinical pharmacokinetics of nasal nicotine delivery. A review and comparison to other nicotine systems.* Clin Pharmacokinet, 1996. **31**: p. 65-80.

231. Sutherland G, S.J., Russell MA, *Randomised controlled trial of nasal nicotine spray in smoking cessation.* Lancet, 1992. **340**: p. 324-329.

232. John R, *Ability of smokers to reduce their smoking and its association with future smoking cessation.* Addiction, 1999. **94**(1): p. 109-114.

233. Charlotta Pisinger, J.V., Knut Borch-Johnsen, Torben Jørgensen, , *Smoking cessation intervention in a large randomised population-based study. The Inter99 study.* Preventive Medicine, 2005. **40**: p. 285-292.

234. Aage Tverdal, K.B., *Health consequences of reduced daily cigarette consumption.* Tob Control, 2006. **15**: p. 472–480.

235. Michèle Tremblay, A.G., Chantal Lacroix, Jennifer O'Loughlin, Hela Makni, Gilles Paradis,, *Physicians Taking Action Against Smoking: an intervention program to optimize smoking cessation counselling by Montreal general practitioners.* CMAJ, 2001. **165**(5): p. 601-607.

236. Fiore MC, Kenford SL, Jorenby DE, Wetter DW, Smith SS, Baker TB. , *Two studies of the clinical effectiveness of the nicotine patch with different counseling treatments.* Chest, 1994. **105**(2): p. 524-533.

237. investigators, r.f.t.c.o.P., *A co-operative trial in the primary prevention of IHD using clofibrate.* Br Heart J, 1978. **40**: p. 1069-118.

238. 1, T.l.r.c.c.p.p.t.r., *reduction in incidence of coronary heart disease.* JAMA, 1984. **251**: p. 351-364.

239. The lipids research clinics coronary primary prevention trial results 2, *reduction in incidence of coronary heart disease.* JAMA, 1984. **251**: p. 351-364.

240. Frick MH, E.O., *primary prevention trial with gemfibrozil in middle-aged men with dyslipidemia.* N Engl J Med, 1987. **317**: p. 1237-45.

241. Matthias Briel, M.S., Tracy R. Glass, Heiner C. Bucher,, *Effects of Statins on Stroke Prevention in Patients with and without Coronary Heart Disease: A Meta-analysis of Randomized Controlled Trials.* Am J Med, 2004. **117**: p. 596 – 606.

242. Earl S. Ford, M., MPH, Wayne H. Giles. , *The Distribution of 10-Year Risk for Coronary Heart Disease Among U.S. Adults.* JACC, 2004. **43**(10): p. 1791-6.

243. Heart Protection Study Collaborative Group, *Effects of cholesterol-lowering with simvastatin on stroke andother major vascular events in 20 536 people with cerebrovascular disease or other high-risk conditions. .* Lancet 2004; 363: 757-67., 2004. **363**: p. 757-767.

244. Sacks FM, P.M., Moye LA,, *The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels.* N Engl J Med, 1996. **336**: p. 1001-09.

245. Pedersen TR, W.L., Faergeman O, , *Follow-up study of patients randomized in the Scandinavian Simvastatin Survival Study (4S) of cholesterol lowering.* Am J Cardiol, 2000. **86**: p. 257–62.

246. Zhou Z, Rahme E, Pilote L, *Are statins created equal? Evidence from randomized trials of pravastatin, simvastatin, and atorvastatin for cardiovascular disease prevention.* Am Heart J, 2006. **151**: p. 273-281.

247. Jayne E Edwards, R Andrew Moore, *Statins in hypercholesterolaemia: A dose-specific meta-analysis of lipid changes in randomised, double blind trials.* BMC Family Practice, 2003. **4**: p. 18.

248. Pate RR, P.M., Blair SN, , *Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine.* JAMA, 1995. **273**: p. 402–7.

249. Wannamethee SG, S.A., *Physical activity in the prevention of cardiovascular disease: an epidemiological perspective* Sports Med 2001. **31**: p. 101–14.

250. Wannamethee SG, Shaper AG. , *Physical activity in the prevention of cardiovascular disease: an epidemiological perspective* Sports Med 2001. **31**: p. 101–14.

251. Susanne B. Votruba, B., Micah A. Horvitz, BA, *The Role of Exercise in the Treatment of Obesity.* Nutrition, 2000. **16**: p. 179-188.

252. Smith, B., Bauman, A., Bull, F., Booth, M. and Harris, M, *Promoting physical activity in general practice: a controlled trial of written advice and information materials.* British Journal of Sports Medicine, 2000. **34**: p. 262-267.

253. Harland, J., White, M., Drinkwater, C., Chin, D., Farr, L. and Howel, D, *The Newcastle exercise project: A randomised controlled trial of methods to promote physical activity in primary care.* bmj, 1999. **319**: p. 828-832.

254. Peter Lindgren, Per Fahlstadius, M.D., Ph.D.,Mai-Lis Hellenius, M.D., Ph.D., and P.D. Bengt Jo"nsson, Ulf de Faire, M.D., Ph.D, *Cost-effectiveness of primary prevention of coronary heart disease through risk factor intervention in 60-year-old men from the county of Stockholm—a stochastic model of exercise and dietary advice.* Preventive Medicine, 2003. **36**: p. 403-409.

255. Paul D. Thompson, MD; David Buchner, MD; Ileana L. Piña, MD; Gary J. Balady, MD; Mark A. Williams, PhD; Bess H. Marcus, PhD; Kathy Berra, MSN, ANP;, *Exercise and Physical Activity in the Prevention and Treatment of Atherosclerotic Cardiovascular Disease A Statement From the Council on Clinical Cardiology (Subcommittee on Exercise, Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity).* circulation, 2003. **107**: p. 3109-3116.

256. Friedman M, Thoresen CE, Gill J, *Alteration of type A behavior and its effect on cardiac recurrences in post myocardial infarction patients: summary results of the recurrent coronary prevention project.* Am Heart J, 1986. **112**: p. 653-665.

257. Blumenthal JA, Jiang W, Babyak MA, *Stress management and exercise training in cardiac patients with myocardial ischemia.* Arch Intern Med, 1997. **157**: p. 2213-2223.

258. Siegler IC, Peterson BL, Barefoot JC, *Hostility during late adolescence predicts coronary risk factors at mid-life.* Am J Epidemiol, 1992. **136**: p. 146–154.

259. Räikkönen K, Keltikangas-Jarvinen L., *Hostility and its association with behavioral induced and somatic coronary risk indicators in Finnish adolescents and young adults.* Social Medicine, 1991. **33**: p. 1171–1178.

260. DiGiuseppe R, Tafrate RC, *Anger treatment for adults: a meta-analytic review.* Clin Psych Sci Pract, 2003. **10**: p. 70-84.

261. Tamara Del Vecchio, K. Daniel O'Leary, *Effectiveness of anger treatments for specific anger problems: A meta-analytic review.* Clinical Psychology Review, 2004. **24**: p. 15-34.

262. George A., Kelley D.A., Kristi S. Kelley M.Ed, Zung Vu Tran Ph.D *Walking and resting blood pressure in adults: A Meta-analysis.* Preventive Medicine, 2003. **33**(2): p. 120-127.

263. Salmon, P., *Effects of physical exercise on anxiety, depression, and sensitivity to stress.* Clinical Psychology Review. **21**(1): p. 33-61.

264. Glueck Z, Weidmann P, Mordasini R,, *Increased serum lowdensity cholesterol in men treated short-term with the diuretic chlorthalidone.* Metabol Clan Exper, 1980. **29**: p. 240-245.

265. Chrysant SG, Neller GK, Dillard B, Frohlich ED, *Effects of dlurettcs on lipid metabolism in patients with essential hypertension.* Angiology, 1976. **27**: p. 707-711.

266. Lehtonen A, Gordin A, *Metabolic parameters after changtng from hydrochlorothiazide to verapamil treatment in hypertension.* Eur J Clin Pharmacol, 1984. **27**: p. 153-157.

267. van Brummelen P, Gevers Leuven JA, van Gent CM, *Influence of hydrochlorothiazide on the plasma levels of triglycerides, total cholesterol and HDL cholesterol in patients with essential hypertension.* Curr Med Res Opin, 1979. **6**: p. 24-29.

268. David Wood, Guy De Backer, Ole Faergeman, Ian Graham, Giuseppe Mancia, Kalevi Pyo"ra"la", *Task Force Report. Prevention of coronary heart disease in clinical practice: Recommendations of the Second Joint Task Force of European and other Societies on Coronary Prevention.* Atherosclerosis, 1998. **140**: p. 199–270.

269. N J Wald, M.R.L., *A strategy to reduce cardiovascular disease by more than 80%.* BMJ, 2003. **326**: p. 1419-1425.

270. Ubbink JB, Fehily AM, Pickering J, Elwood PC, Hayward Vermaak WJ., *Homocysteine and ischaemic heart disease in the Caerphilly cohort* Atherosclerosis, 1998. **140**: p. 349-356.

271. Elwood PC, Beswick A, Pickering J, McCarron P, O'Brien JR, Renaud SR, Flower RJ, *Platelet tests in the prediction of myocardial infarction and stroke: evidence from the Caerphilly study.* Brit J Haematol, 2001. **113**: p. 514-520.

272. N J Wald, M R Law, *A strategy to reduce cardiovascular disease by more than 80%.* BMJ, 2003. **326**: p. 1419-1425.

# Appendix 1
## Data collected in the Caerphilly cohort study

| Risk factor measured | | Phase 1 | Phase 2 | Phase 3 | Phase 4 |
|---|---|---|---|---|---|
| IHD | LSHTM | ✓ | ✓ | ✓ | ✓ |
| | 12 lead ECG | ✓ | ✓ | ✓ | ✓ |
| | Family history of IHD | ✓ | ✓ | × | × |
| | IHD events | ✓ | ✓ | ✓ | ✓ |
| Life style data | Marital status | ✓ | ✓ | ✓ | ✓ |
| | Social class | ✓ | ✓ | ✓ | ✓ |
| | Father's social class | ✓ | ✓ | × | × |
| | Employment status | ✓ | ✓ | ✓ | ✓ |
| | Diet | ✓ | ✓ | ✓ | ✓ |
| | Smoking habit | ✓ | ✓ | ✓ | ✓ |
| | Alcohol consumption | ✓ | ✓ | ✓ | ✓ |
| | Sleep pattern | × | ✓ | ✓ | × |
| | Sleeping apnoea | × | ✓ | ✓ | × |
| | Leisure activity(exercise) | ✓ | ✓ | ✓ | ✓ |
| Clinical data | Medical history | ✓ | ✓ | × | × |
| | Weight | ✓ | ✓ | ✓ | ✓ |
| | Height, birth weight | ✓ | ✓ | × | × |
| | Skin fold thickness | ✓ | ✓ | ✓ | ✓ |
| | Lung function | ✓ | ✓ | ✓ | ✓ |
| | BP | ✓ | ✓ | ✓ | ✓ |
| | LHSTM | ✓ | ✓ | ✓ | ✓ |
| | Hearing | ✓ | ✓ | ✓ | ✓ |
| | TIA Questionnaire | ✓ | ✓ | ✓ | ✓ |
| | Medication | ✓ | ✓ | ✓ | ✓ |
| Psychological factors | Social support | × | ✓ | × | × |
| | Job satisfaction | × | ✓ | × | × |
| | Type A Behaviour | × | ✓ | × | × |
| | HAI | × | ✓ | ✓ | × |
| | GHQ | × | ✓ | ✓ | ✓ |

| Blood tests | WCC | ✓ | ✓ | ✓ | ✓ |
|---|---|---|---|---|---|
| | Fibrinogen | ✓ | ✓ | ✓ | ✓ |
| | Plasma viscosity | ✓ | ✓ | ✓ | ✓ |
| | Lipids | ✓ | ✓ | ✓ | ✓ |
| | HDL, LDL, VLDL, $HDL_2$, $HDL_3$ | ✓ | ✓ | ✓ | ✓ |
| | Triglycerides | ✓ | ✓ | ✓ | ✓ |
| | Testosterone | ✓ | ✓ | ✓ | ✓ |
| | Oestradiol | ✓ | ✓ | ✓ | ✓ |
| | Cortisol | ✓ | ✓ | ✓ | ✓ |
| | Insulin | ✓ | ✓ | ✓ | ✓ |
| | Blood glucose | ✓ | ✓ | ✓ | ✓ |
| | Liver and kidney function tests | ✓ | ✓ | ✓ | ✓ |

# APPENDIX 2

## HEALTH ATTITUDE INVENTORY

This questionnaire concerns your views on how various things affect you personally. Read each question carefully and *show your answer by circling the number* which best reflects your view. Please answer *all* the questions as best you can and work quickly by giving the first answer that comes to you. For example, in the example given below if you knew that you didn't like playing golf you would circle the number T as shown.

|  | Definitely<br>NO | Probably<br>NO | Probably<br>YES | Definitely<br>YES |
|---|---|---|---|---|
| EXAMPLE: Do you like playing golf? | (1) | 2 | 3 | 4 |
| Will exercise improve your own health and fitness? | 1 | 2 | 3 | 4 |
| Do you generally enjoy exercise? | 1 | 2 | 3 | 4 |
| Is it difficult to find time to exercise? | 1 | 2 | 3 | 4 |
| Will exercise help you lose weight? | 1 | 2 | 3 | 4 |
| Will exercise reduce your risk of a heart attack? | 1 | 2 | 3 | 4 |
| Do you generally enjoy the taste of dairy produce? | 1 | 2 | 3 | 4 |
| Will eating dairy produce improve your health? | 1 | 2 | 3 | 4 |
| Is dairy produce generally convenient to eat? | 1 | 2 | 3 | 4 |

Does dairy produce contain a lot of fat? 1 2 3 4

Will eating dairy produce increase your risk of a heart attack? 1 2 3 4

Do you generally enjoy the taste of most vegetables? 1 2 3 4

Will eating vegetables improve your health? 1 2 3 4

Will eating vegetables give you a better balanced diet? 1 2 3 4

Will eating vegetables reduce your risk of a heart attack? 1 2 3 4

Do you generally enjoy the taste of fried food? 1 2 3 4

Will eating fried food harm your health? 1 2 3 4

Is frying generally a convenient way to cook? 1 2 3 4

Does fried food contain a lot of fat? 1 2 3 4

Will eating fried food increase your risk of a heart attack? 1 2 3 4

Do you generally enjoy the taste of wholemeal bread? 1 2 3 4

Will eating wholemeal bread improve your health? 1 2 3 4

Will eating wholemeal bread improve your digestion? 1 2 3 4

Will eating wholemeal bread reduce your risk of a heart attack? 1 2 3 4

Will smoking harm your health? 1 2 3 4

Is smoking generally unpleasant to others? 1 2 3 4

Do you generally enjoy smoking? 1 2 3 4

Is smoking expensive? 1 2 3 4

Will smoking increase your risk of a heart attack? 1 2 3 4

Will setting deadlines help you get things done? 1 2 3 4

Will getting angry help you get things done? | 1 | 2 | 3 | 4

Will being competitive help you get things done? | 1 | 2 | 3 | 4

Will working hard to meet deadlines increase your risk of a heart attack? | 1 | 2 | 3 | 4

Will getting worked up and angry increase your risk of a heart attack? | 1 | 2 | 3 | 4

Will being very competitive increase your risk of a heart attack? | 1 | 2 | 3 | 4

Does your wife or closest friend like you to take exercise? | 1 | 2 | 3 | 4

Does your wife or closest friend like you to smoke? | 1 | 2 | 3 | 4

Does your wife or closest friend like you to eat wholemeal bread? | 1 | 2 | 3 | 4

Does your wife or closest friend like you to eat vegetables? | 1 | 2 | 3 | 4

Does your wife or closest friend like you to eat dairy produce? | 1 | 2 | 3 | 4

Does your wife or closest friend like you to eat fried food? | 1 | 2 | 3 | 4

Does your wife or closest friend like you to get things done quickly? | 1 | 2 | 3 | 4

### How much do you like or dislike each of the following?

| | Dislike Very much | Dislike | Not Sure | Like | Like Very much |
|---|---|---|---|---|---|
| Getting your enjoyment through smoking | 1 | 2 | 3 | 4 | 5 |
| Spending money on smoking | 1 | 2 | 3 | 4 | 5 |
| Unpleasantness to others caused by smoking | 1 | 2 | 3 | 4 | 5 |
| The taste of dairy produce in general | 1 | 2 | 3 | 4 | 5 |

300

*How much do you value each of the following?*

Improving your health and Fitness through exercise

| | Not At all | A little | Moderately | very much | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| The convenience of eating dairy produce | 1 | 2 | 3 | 4 | 5 |
| The amount of fat in dairy produce | 1 | 2 | 3 | 4 | 5 |
| The taste of vegetables in general | 1 | 2 | 3 | 4 | 5 |
| Eating vegetables to improve the balance of your diet | 1 | 2 | 3 | 4 | 5 |
| The taste of fried food in general | 1 | 2 | 3 | 4 | 5 |
| The convenience of fried food | 1 | 2 | 3 | 4 | 5 |
| The amount of fat in fried food | 1 | 2 | 3 | 4 | 5 |
| The taste of wholemeal bread | 1 | 2 | 3 | 4 | 5 |
| Eating wholemeal bread to improve your digestion | 1 | 2 | 3 | 4 | 5 |
| Getting your enjoyment through exercise | 1 | 2 | 3 | 4 | 5 |
| Putting time aside for exercise | 1 | 2 | 3 | 4 | 5 |
| Exercising to lose weight | 1 | 2 | 3 | 4 | 5 |
| Setting deadlines to get things done | 1 | 2 | 3 | 4 | 5 |
| Getting angry to get things done | 1 | 2 | 3 | 4 | 5 |
| Being competitive to get things done | 1 | 2 | 3 | 4 | 5 |

Improving your health by eating vegetables   1   2   3   4

Improving your health by eating dairy produce   1   2   3   4

Improving your health by eating wholemeal bread   1   2   3   4

Reducing your risk of a heart attack by eating vegetables   1   2   3   4

Reducing your risk of a heart attack through exercise   1   2   3   4

Reducing your risk of a heart attack by eating wholemeal bread   1   2   3   4

| | Not At all | A little | Moderately | very much |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |

*How much do you fear each of the following?*

| | At all | A little | Moderately | very much |
|---|---|---|---|---|
| Harming your health through smoking | 1 | 2 | 3 | 4 |
| Harming your health by eating fried food | 1 | 2 | 3 | 4 |
| Increasing your risk of a heart attack through smoking | 1 | 2 | 3 | 4 |
| Increasing your risk of a heart attack by eating fried food | 1 | 2 | 3 | 4 |
| Increasing your risk of a heart attack by eating dairy produce | 1 | 2 | 3 | 4 |
| Increasing your risk of a heart attack by working hard to meet deadlines | 1 | 2 | 3 | 4 |
| Increasing your risk of a heart attack by getting worked up and angry | 1 | 2 | 3 | 4 |
| Increasing your risk of a heart attack by being very competitive | 1 | 2 | 3 | 4 |

How much do you generally like to do
what your wife or closest friend says?

| Couldn't Care less | Not much | Not sure | Like to | Like to very much |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 1 | 2 | 3 | 4 | 5 |

## Anger Questionnaire Items

| Item | Question | Scale |
|------|----------|-------|
| | When really angry or annoyed do you | |
| 1 | Get tense or worried? | **Anger symptoms** |
| 2 | Get a headache? | |
| 3 | Feel weak? | |
| 4 | Get nervous or shaky? | |
| 5 | Try to act as though nothing had happened? | **Anger in** |
| 6 | Keep it to yourself? | |
| 7 | Apologize even though you are right? | |
| 8 | Take it out on others? | **Anger out** |
| 9 | Blame someone else? | |
| 10 | Get it off your chest? | **Anger discuss** |
| 11 | Talk to a friend or relative? | |

To each question, the responses and score are very likely (2); somewhat likely (1); or not too likely (0).

# Appendix 4

Weibull models used for simulating the survival times for the interventions in chapter 9.

1) Model used to Intervention scenarios in Table 9.3

$$T = \left( \frac{-\log(U)}{15.41 \times \exp(0.001 \times ALC)} \right)^{0.67}$$

2) Model used to Intervention scenarios in Table 9.4

$$T = \left( \frac{-\log(U)}{15.10 \times \exp(0.015 \times SBP)} \right)^{0.78}$$

3) Model used to Intervention scenarios in Table 9.5

$$T = \left( \frac{-\log(U)}{17.24 \times \exp(0.013 \times SBP)} \right)^{0.67}$$

4) Model used to Intervention scenarios in Table 9.7

$$T = \left( \frac{-\log(U)}{13.22 \times \exp(0.75 \times SMOKING)} \right)^{0.78}$$

5) Model used to Intervention scenarios in Table 9.9

$$T = \left( \frac{-\log(U)}{14.29 \times \exp(0.265 \times TCHOL)} \right)^{0.79}$$

6) Model used to Intervention scenarios in Table 9.10

$$T = \left( \frac{-\log(U)}{12.44 \times \exp(0.505 \times LEISURE)} \right)^{0.79}$$

7) Model used to Intervention scenarios in Table 9.11

$$T = \left( \frac{-\log(U)}{13.37 \times \exp(0.782 \times ANGER - OUT)} \right)^{0.79}$$

8) Model used to Intervention scenarios in Table 9.12

$$T = \left( \frac{-\log(U)}{17.03 \times \exp(-0.0007 \times ALC + 0.014 \times SBP + 0.76 \times SMOKING + 0.231 \times TOT\_CHOL - 0.423 \times LEI\_ACT + 0.67 \times ANG\_OUT)} \right)^{0.77}$$

9) Model used to Intervention scenarios in Table 9.13

$$T = \left( \frac{-\log(U)}{17.5 \times \exp(0.0008 \times ALC + 0.012 \times SBP + 0.18 \times SMOKING + 0.20 \times TOT\_CHOL - 0.09 \times LEI\_ACT + 0.075 \times ANG\_OUT)} \right)^{0.67}$$

306