Blood pressure, cholesterol and premature death: towards the real relationships

Sarah Lewington

St Peter's College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary Term, 1999

In memory of my dad, with love

Acknowledgements

My thanks go to the many people within the Clinical Trial Service Unit & Epidemiological Studies Unit who have given me intellectual, moral and friendly support since I joined the unit in 1996. In particular I would like to thank the following people (in no particular order):

Richard Peto and Rory Collins for supporting me in my work on the Prospective Studies Collaboration, and for supporting me in writing up this work for a DPhil thesis.

Robert Clarke for his friendship and support, and for allowing me to share my ideas with him.

Christina Davies, for her particular friendship and help with this thesis.

Wei Zhang, Arlene Donald, Paul Sherliker and Alison Palmer for their help with the programming and for their tolerance over the last few weeks.

Colin Baigent and John Danesh for their friendship and useful ideas.

I am very grateful to Cathie Sudlow for her time and valuable suggestions.

Vaughan Evans and Deb Jackson for proof reading the thesis with such attention to detail.

I would also like to thank Stephen Duffy for his advice and support when I needed it most at the beginning of my studies.

On a personal note, my thanks go to my father for encouraging me to register for a higher degree; and to my mother for her support over the years, and for teaching me to "think positive". Finally, my greatest thanks go to Simon who has supported me in all I have undertaken and for his love, support and unlimited tolerance over the last ten years.

Blood pressure, cholesterol and premature death: towards the real relationships

Sarah Lewington, St Peter's College, University of Oxford
A thesis submitted for the degree of Doctor of Philosophy
Hilary Term, 1999

ABSTRACT

This thesis is based on a worldwide overview (meta-analysis) of prospective observational studies of blood pressure and cholesterol, involving a centralised collection of data on over one million individuals from 59 studies, which I have co-ordinated since its inception. Analytically, the aim has been to develop and to use appropriate statistical techniques to assess the age- and sex-specific associations of usual blood pressure and of usual cholesterol with cause-specific mortality. Since the data set is uniquely large, and because appropriate methods of analysis (with full account taken of the time-dependent nature of the regression dilution bias) have been developed and used, these associations have been established more reliably. An integral part of the methodological element of the thesis has been to investigate the systematic underestimation of associations between risk factor and disease that are obtained when only a single baseline measurement is used to assess levels of such risk factors (the regression dilution bias). The extent of this bias has been investigated in each study that had repeat measurements of risk factors during follow-up. One particularly novel aspect has been the emphasis on, and methods developed to account for, the regression dilution bias in several studies simultaneously and in an appropriately time-dependent way.

This thesis illustrates the extent to which random error and inappropriate statistical analysis lead to misleading conclusions concerning the importance of blood pressure and blood cholesterol, particularly in premature death. Only by studying adequate numbers of deaths (136,000 deaths among 1 million adults during 13 million person-years of follow-up) and by using appropriate statistical techniques - taking proper account of (a) the regression dilution bias; (b) the full range of blood pressure and cholesterol; (c) the opposing effects of HDL and the remaining non-HDL cholesterol; and (d) age at death - did it become possible to provide reliable results on the true relationships between blood pressure, cholesterol fractions and vascular and other causes of death.

These analyses have demonstrated reliably that, as causes of IHD death in early middle age, blood pressure and blood lipids are three to five times more important than suggested by inappropriate analyses, with no clinically relevant inverse associations with cancer or other non-vascular mortality (except, surprisingly, COPD).

Abbreviations and notation

χ^2	chi-square
BMI	Body Mass Index
CI	Confidence Interval
DBP	Diastolic Blood Pressure
HDL	High Density Lipoprotein
HR	Hazards Ratio
IHD	Ischaemic Heart Disease
LDL	Low Density Lipoprotein
MI	Myocardial Infarction
mmHg	millimetres of mercury
mmol/l	millimoles per litre
p	probability value for a statistical test
PSC	Prospective Studies Collaboration
ρ	correlation coefficient
RDR	Regression Dilution Ratio
SBP	Systolic blood pressure
SD	Standard deviation
SE	Standard error
Stroke	All cerebrovascular diseases (ICD-9: 430-8)
Haemorrhagic stroke	Primary intra-cerebral haemorrhage (ICD-9: 430-8)
Ischaemic stroke	Cerebral ischaemia (ICD-9: 433-4)

Table of Contents

List of figures and tables

Chapter 1: Background and objectives

Worldwide, half of all adult mortality (and extensive severe disability) is caused by vascular diseases, and most of these deaths involve ischaemic heart disease or stroke. Many of those killed by these diseases are still in middle age (here defined as 35-69) when they die, losing about 20 years of life expectancy. In the year 2000, there will be nearly 3 million such deaths among people under 60 years of age, and over 3 million among people aged 60-69 (Murray and Lopez, 1994).

Much of our knowledge about the relevance of the key risk factors for chronic diseases, such as heart disease and stroke, comes from prospective observational studies in which various characteristics of a cohort are recorded on enrolment to the study (the "baseline" survey). The characteristics of those who had died from (or developed) a particular disease at a particular time are then compared with the characteristics of those who had not done so by that time. The associations of blood pressure and cholesterol with various causes of death have been investigated in many such prospective observational studies over the latter half of this century. Yet, the real importance of these risk factors has previously been seriously underestimated because of limitations in the design and, particularly, the analysis of these studies:

- Most diseases will affect only a small proportion of a study population, especially if follow-up is of limited duration. Thus, even for relatively common diseases, a prospective study (following unaffected individuals until a fair proportion have developed disease) should ideally involve some hundreds of thousands of individuals followed for at least one or two decades. Only then, would sufficient

deaths from the disease of interest ("events") be observed to allow reliable assessment of the true nature of any risk relationships. Although no individual study can provide such data, sufficient events are available from appropriate combination of data from all existing prospective observational studies (that is, an overview or "meta-analysis").

- Some studies have reported "J-shaped" relationships of both blood pressure and cholesterol with many chronic diseases. This has often been claimed as evidence for "thresholds" above or below which the relationship between risk factor and disease is no longer direct. Yet, these observations have been based on relatively small numbers of events in the extremes of the risk factor distributions, and have been distorted by failure to remove those who already had disease, which might itself affect their blood pressure or cholesterol: see below. Furthermore, no single study has been able to provide reliable data throughout the entire range of a risk factor distribution. For example, a study of the relevance of total cholesterol in Finland (where cholesterol levels are generally high) would not be informative about the role of cholesterol in Chinese populations (where cholesterol levels are generally low, and even very low levels are common).

- Many risk factors, such as blood pressure and cholesterol, are proportionally much more relevant in middle age than they are in old age. Therefore, any analyses of disease-risk relationship must be performed within groups of individuals according to the age at which they died.

- Life-threatening diseases can themselves cause low blood pressure and low cholesterol ("reverse causality"). Analyses excluding those who were already

known to be ill at the start of the study, and those who died within a few years of the risk factor measurement, are needed to explore the possibility that any inverse associations can be explained by long-term effects of pre-clinical disease on blood pressure and cholesterol.

- Due to the combined effects of measurement errors, short-term biological variability and longer-term systematic changes within individuals, baseline measurements may not reliably indicate the "usual" level of a risk factor either at around the time of the baseline measurement or during a later period (e.g. the second decade of follow-up). Hence, unless some account is made for this in the analysis, the true relationship between usual levels of a risk factor in a particular period and the subsequent risk of disease during that same (or some later) period will be misrepresented, often by a substantial amount. Generally, the real importance of a risk factor will be systematically under-estimated unless some correction is made for this, so-called, "regression dilution" bias. The regression dilution bias is directly relevant to the analysis of all observational studies, irrespective of their quality or size.

Individual prospective studies have typically involved too few outcome events to address these issues reliably. As a consequence they have been unable, in isolation, to provide reliable estimates of the nature of the age-specific associations of blood pressure and cholesterol with cause-specific mortality. Moreover, all such studies have related risk to risk factor measurements made at a single baseline visit.

To help resolve these issues, I have co-ordinated a worldwide collaboration between investigators of prospective observational studies, who have agreed to pool their data

to establish a centralised collection of data from individuals in these studies: the "Prospective Studies Collaboration". By appropriate analysis of data from all relevant prospective observational studies, simultaneously accounting for each of the above factors, this collaboration will characterise with less bias and with more precision than before the age- and sex-specific relevance of usual blood pressure and usual cholesterol to various causes of death.

The primary objective of this thesis was to establish a unique centralised collection of data on individuals in prospective observational studies world-wide, and to develop and to use appropriate ways of assessing the age- and sex-specific associations of usual blood pressure and of usual cholesterol with cause-specific and all-cause mortality within these data. An integral part of this objective has been to investigate the underestimation of associations with risk that are obtained when only a single baseline measurement is used to assess risk factor levels (the regression dilution bias), and to investigate the differing extent of this underestimation in those studies that have re-measured risk factor levels during follow-up. One particularly novel aspect of this thesis is the emphasis on, and methods developed to account for, the regression dilution bias in several studies simultaneously in an appropriately time-dependent way.

Chapter 2: Introduction to the "regression dilution bias"

Chapter summary

Until the present decade, the importance of established risk factors for vascular disease, such as blood pressure and cholesterol, were being systematically and seriously underestimated (or diluted) by failure to account for variability in biological measurements. Reliable methods are required to correct for this regression dilution bias in epidemiological studies, and such methods need to be simple and robust if they are to be widely used. Otherwise, the strength of the relationships between risk factors and disease will continue to be underestimated. This chapter describes the principles underlying the need for such methods.

1 Initial view of the regression dilution bias

"Regression to the mean" is the phenomenon whereby a variable that is extreme on first measurement will tend to be closer to the mean of its distribution on re-measurement. Sir Francis Galton, when experimenting with peas in 1886, observed the tendency for two tall plants to produce offspring that were also tall, albeit, on average, shorter than their parents. Conversely, small plants were observed to produce offspring that were also small, but taller than the parents (Galton, 1886). Galton called this effect "regression towards mediocrity". It occurs in any measurement that fluctuates randomly within individuals, or is prone to error in measurement. A single measurement may be higher or lower than the individual's long-term average (or "usual" value). Consequently, a group of individuals classified as "high" from a single measurement would tend to include a disproportionate

number whose measurement was higher than their usual value. So, the mean of their measured values would tend to exceed the mean of their usual values. If, however, this same group of individuals were re-measured, then their re-measured values would differ only randomly from their true, or "usual" values. So, the mean of their re-measured values would, in expectation, be the same as the mean of their usual values. Thus, the mean of the re-measured values would tend to be lower than the mean of the initial values for this group of individuals with initially "high" values. Conversely, for those with initially "low" values, the mean of the re-measured values would tend to be higher than the mean of the initial values.

The importance of this phenomenon in epidemiology was not appreciated until recently. MacMahon et al. (1990) reported how fluctuations in blood pressure measurements resulted, not in fluctuations in the strength of the relationship between the risk factor and disease but, in a systematic underestimation of that relationship. This has become known as the "regression dilution bias". Despite efforts to minimise errors in the measurement of blood pressure or cholesterol, some random error will always persist. It was widely (but incorrectly) believed that such random errors would cancel out in any analysis, particularly in a study involving many thousands of participants. The following exemplifies why this will never be so, whatever the size of the study. It illustrates the impact on risk estimates of purely random errors in the exposure measurements, and suggests a simple solution to the problem.

1.1 The epidemiology gremlin

The relevance of increasing blood pressure to stroke risk is assessed in three situations. First, in a hypothetical setting with no random errors; next, in a situation that shows how non-random biases arise when purely random errors (with no bias) are introduced into the measurement of blood pressure; and finally, in a situation that shows how a simple correction can be made that eliminates the bias introduced by these random errors.

Consider an idealised world in which there are only two possible diastolic blood pressure (DBP) values (either 80 mmHg or 100 mmHg, with half the population having a DBP of 80 mmHg, and the other half having a DBP of 100 mmHg), and in which there is no random variation or long-term change in blood pressure. So, those born with a DBP of 80 mmHg (or 100 mmHg) will live their entire lives with a DBP of 80 mmHg (or 100 mmHg) and, irrespective of other circumstances, will have the same DBP throughout each day. Next, consider a large prospective study involving four million individuals designed to assess the relevance of DBP to the risk of stroke. Such a study is large enough not to be much affected by the play of chance, so roughly two million would have a DBP measurement of 80 mmHg and two million a DBP measurement of 100 mmHg. If the cohort is then followed for 10 years to monitor deaths from stroke, the risk ratio for stroke that is associated with this 20 mmHg difference in DBP could be calculated using standard statistical techniques. Because, in this ideal world, no errors are ever made in the measurement of blood pressure, either by the observer or by the sphygmomanometer, this large prospective study would give the *true* relation of "usual" DBP to stroke risk. Those

with a DBP of 80 mmHg would be at relatively "low" risk of stroke, while those with a DBP of 100 mmHg would be at relatively "high" risk of stroke. A graph of stroke risk against DBP would show the true slope or relation to risk (Figure 1).

Next, as in the previous scenario, the clinic staff still take perfect measurements, but the sphygmomanometer now randomly adds or subtracts an error to every measurement. These errors are not systematic and are independent of the true DBP. However, for half the individuals a random 10 mmHg is added to each individual's true DBP, while for the remaining half a random 10 mmHg is subtracted. So now, every measurement yields 70, 90 or 110 mmHg (see Figure 2). These error-prone measurements will subsequently be referred to as the "baseline" measurements.

People with a true DBP of 80 mmHg (but baseline measurements of 70 or 90) are still at the same "low" risk of stroke, and those with a true DBP of 100 mmHg (but baseline measurements of 90 or 110) are still at the same "high" risk of stroke. However, when the population is divided according to "baseline" measurements there are now three groups denoted by **L**, **M** and **H**:

L contains all those whose true DBP is 80 mmHg (and who are, therefore, at low stroke risk), but with a baseline measurement of 70 mmHg because 10 mmHg was randomly **subtracted** from their true DBP;

H contains all those whose true DBP is 100 mmHg (and who are, therefore, at high stroke risk), but with a baseline measurement of 110 mmHg because 10 mmHg was randomly **added** to their true DBP; and,

M contains a mixture of individuals. Firstly, it includes those whose true DBP is 80 mmHg (and who are, therefore, at low risk of stroke), but with a baseline measurement of 90 mmHg because 10 mmHg was randomly **added** to their true DBP. Secondly, it includes those whose true DBP is 100 mmHg (and who are, therefore, at a high risk of stroke), but with a measurement of 90 mmHg because 10 mmHg was randomly **subtracted** from their true DBP. So, half of group **M** are at low risk and half are at high risk, which means collectively, they have an intermediate level of risk.

Now, when stroke risk is plotted against the mean of the baseline measurements (Figure 3) the relation to risk appears much *shallower* than (that is, half a steep as) the true relationship seen in Figure 1. Thus, use of a single baseline reading, measured with *random* error, significantly underestimates the importance of DBP to stroke risks. This is the so-called regression dilution bias.

To relate the risk of stroke to the true (or "usual") DBP, *some correction* is needed to account for the regression dilution bias caused by random errors in the DBP measurements. This can be achieved simply by re-measuring the DBP some years later in a sample of the survivors from each of the three baseline-defined groups (i.e. L, M and H). The clinic staff still take perfect measurements (i.e. exactly 80 mmHg or exactly 100 mmHg), and the sphygmomanometer still randomly adds or subtracts 10 mmHg to each reading. However, these errors are still not systematic, and are totally independent of the true DBP.

The means of the re-measurements in each of the *baseline-defined* groups equal the means of the true values in those groups (Figure 2). This is because group **L** consisted of people whose true DBP is 80 mmHg (although they measured 70 mmHg at baseline), and so on re-measurement half give a reading of 70 mmHg (i.e. 80–10) and half give a reading of 90 mmHg (i.e. 80+10), and the mean of these re-measurements is 80 mmHg. Similarly, group **H** consists of people whose true DBP is 100 mmHg (although they measured 110 mmHg at baseline), and so on re-measurement half give a reading of 90 mmHg (i.e. 100–10) and half give a reading of 110 mmHg (i.e. 100+10), with a mean of 100 mmHg. Group **M** contains people whose true values are either 80 or 100, so on re-measurement one-quarter give readings of 70 mmHg (80-10), one-half give readings of 90 mmHg (80+10 or 100-10), and one-quarter give readings of 110 mmHg (100+10). Hence the mean DBP reading in group **M** is 90 mmHg. When stroke risk is plotted against the mean of the re-measurements in these baseline-defined groups, the true slope emerges (Figure 4).

This is a simple and trivial example, but it serves to illustrate the regression dilution, or systematic underestimation (hence the term "dilution"), in risk associations caused by purely random errors in the measurement of an exposure. Clearly, if appropriate conclusions are to be drawn from observational studies, then acknowledgement of the bias, and correction for it, are fundamental. The importance of this bias has attracted increasing interest over the last decade (MacMahon et al., 1990; Qizilbash, Duffy, and Rohan, 1991; Rosner, Spiegelman, and Willett, 1990; Rosner, Spiegelman, and Willett, 1992; Rosner, Willett, and Spiegelman, 1989; Duffy,

Maximovitch, and Day, 1992; Duffy, Rohan, and Day, 1989; Elton and Duffy, 1983). Bashir and Duffy (1997) and Frost and Thompson ([*in press*]) have reviewed various methods to correct for this bias, along with the assumptions required by each method. Many of the methods proposed to deal with measurement error require some assumptions about the relationship between the initial and the repeat measurements, which may not hold over prolonged follow-up. The main advantage of the methods developed and described in this thesis is that they are simple and do not involve any unjustified assumptions: see Chapter 3 (Data collection and statistical methods).

However, debate over *how* to account for the regression dilution bias should not detract from the importance of *making some such correction*. Epidemiologists need to incorporate ways to correct for this bias into the design of their study. Primarily they must aim to re-survey at least a (representative) sample of the survivors of their original cohort within a few years of the initial screening visit and then again some years later – perhaps one re-survey in every decade of follow-up would be appropriate (see below).

2 The regression dilution bias over prolonged follow-up: a broader definition

Re-measurements taken within just a few years of baseline do not take appropriate account of within-person variability over more prolonged periods. Causes of longer-term variability may include changes in measurement techniques, differences over time in risk factor management, public health initiatives, other environmental influences and the systematic effects of ageing. There are consistent differences

between individuals in the rate of change of certain physiological characteristics, such as blood pressure and cholesterol, during adult life that may be due partly to genetic differences and partly to environmental influences (Miall and Lovell, 1967). These genetic differences and environmental influences will affect both the actual level of the risk factor in middle age and the rate of change (or slope). Such physiological changes within a particular individual define a "track" leading him through many decades of life: the so-called "Horse-Racing Effect" (Peto, 1977).

Consider a group of men whose diastolic blood pressure (DBP) was 90 mmHg at age 50. This group could contain men who had a DBP of 85 mmHg at age 30 and a low rate of change, but also men who had a DBP of 80 mmHg at 30 and a higher rate of change (Figure 5). That is, the systematic physiological change within each individual has placed them on a track. Without taking periodic re-measurements as a way of establishing the rates of change (or tracks) among each of these individuals, it would be difficult to estimate the average DBP of this group some years later.

Both short and longer-term fluctuations leading to "regression to the mean" and systematic changes with time have major implications in the analysis of data from prospective observational studies. If individuals are grouped as "High", "Medium" or "Low" according to a single ("baseline") measurement of a characteristic that both fluctuated randomly and changed systematically over time, then the original means of each of these groups would provide inaccurate and biased estimates of the true means some years later. In some groups, these two phenomena might act in opposite directions and thus cancel each other out. However, provided the re-measurements used to calculate the mean of a group are not those used to define

membership of the group, they will provide an unbiased estimate of the mean of the group at the time these re-measurements are taken. The term "regression dilution bias" will be used to cover not only the effects of measurement error (causing regression to the mean), but also the effects of any causes of longer-term fluctuations as well as systematic physiological changes within individuals.

2.1 Influence of the interval between measurements on regression dilution

We previously examined the importance of regression dilution for blood pressure and cholesterol over prolonged follow-up (Clarke et al., [*in press*]), using biennial measurements of exposures over a 30-year period in 5,209 persons from the Framingham study (Anderson, Castelli, and Levy, 1987), and in a 26-year re-survey of 401 men in the Whitehall study (Lichtenstein, Shipley, and Rose, 1985). The correlation between replicate measurements of systolic blood pressure and baseline values in the Framingham study declined progressively with increasing duration of follow-up: 0.61 for pairs of measurements taken 6 years apart, 0.44 for pairs of measurements taken 16 years apart, and 0.30 for pairs of measurements taken 26 years apart. The corresponding correlations for total cholesterol also declined progressively: 0.68, 0.53 and 0.46, respectively. Thus, for both blood pressure and cholesterol, the agreement between measurements taken some years apart diminished markedly with increasing separation.

In contrast with the modest changes in the overall mean value of blood pressure and cholesterol over time in the Framingham study, substantial artefactual changes were seen in the mean values at subsequent follow-up in the top and bottom fifths defined

only by the baseline measurements (Figure 6). In general, there was a sharp convergence by year 2 of measurements in the top and bottom fifths, probably due to the effects of measurement error and short-term fluctuations in the baseline measurements. This was followed by a slower, but progressive, convergence over the next few decades. For example, the mean systolic blood pressure (in mmHg) for those in the top fifth initially (i.e. those with a value of 146 mmHg or greater on first measurement) declined from 162 at the start, to 156 after 2 years, and was then 155 after 6 years, 153 after 16 years and 147 after 26 years. At the other extreme, the mean systolic blood pressure (in mmHg) for those in the bottom fifth initially (i.e. those with a systolic blood pressure of less than 118 mmHg on first measurement) increased from 111 at the start, to 118 after 2 years, and then 120 after 6 years, 126 after 16 years and 130 after 26 years. Hence, the absolute differences (or ranges) between the mean values in the top and bottom fifths of systolic blood pressure declined throughout the follow-up period: 51 mmHg at the start, 38 mmHg after 2 years, 35 mmHg after 6 years, 27 mmHg after 16 years, and 18 mmHg after 26 years. Similar patterns were seen for diastolic blood pressure and for total cholesterol. For systolic blood pressure and for total cholesterol, the mean of the top fifth declined at a similar rate that the mean of the bottom fifth rose (that is, the convergence was more or less symmetrical). In contrast, for diastolic blood pressure, the mean of the top fifth continued to decline sharply while the mean of the bottom fifth remained fairly constant after an initial rise by year 2 (Figure 6). By comparing analyses of the entire population with the same analyses for those who survived the

entire follow-up period, it was found that little of this convergence was due to differential mortality.

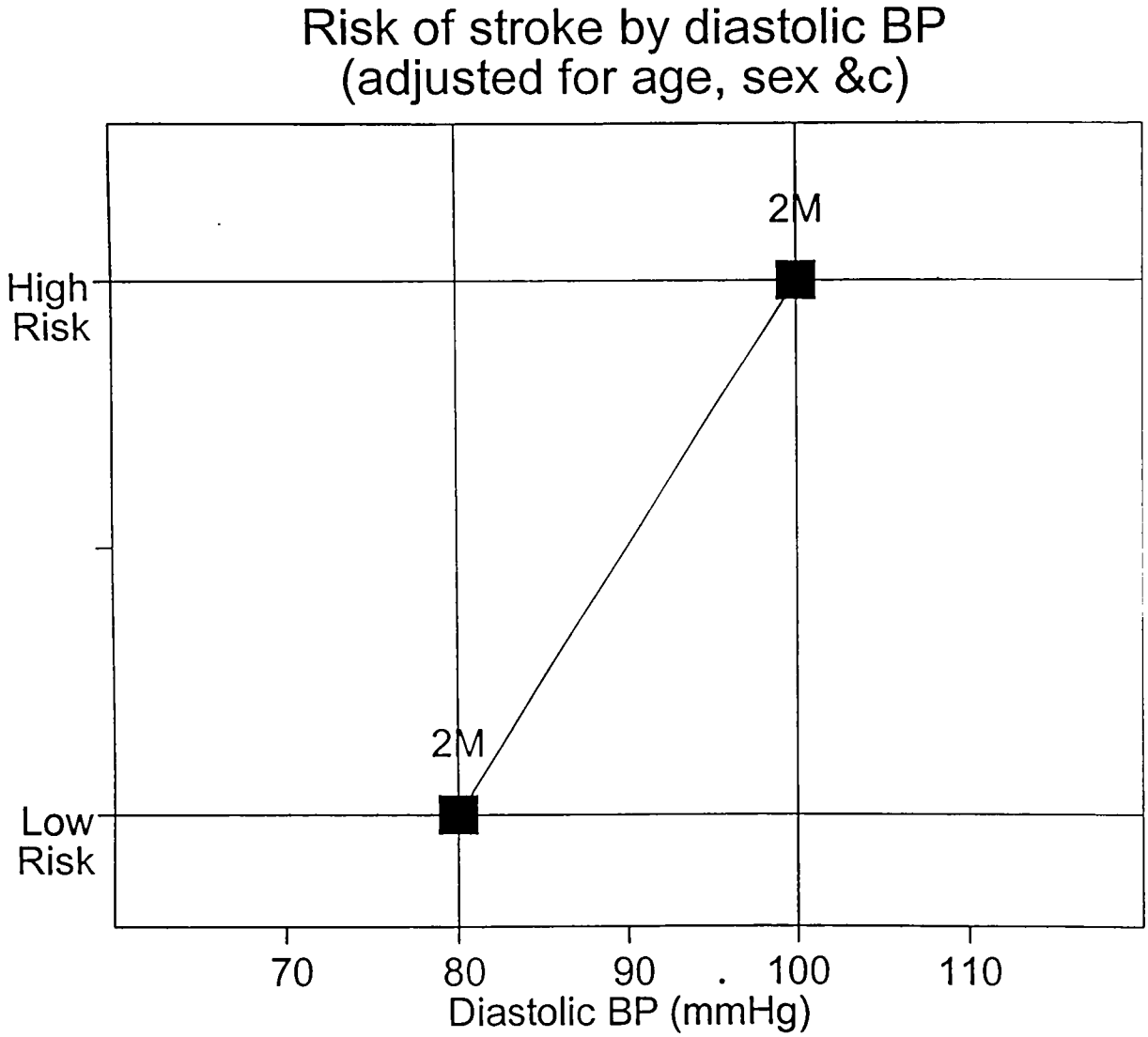
We concluded that the use of uncorrected single measures of blood pressure and cholesterol at baseline in epidemiological studies may underestimate the magnitude of the associated risks in the subsequent follow-up (due to regression dilution) by as much as one-third in the first decade, one-half in the second decade and two-thirds in the third decade of follow-up.

Correction for regression dilution in a prospective study with long-term follow-up should be appropriate to the exposure period of interest if the aim of the analysis is to assess long-term risk with long-term exposure. So if exposure in the second decade of follow-up is to be related to risk in that same, or some later, decade, the relative risk associated with the baseline measurement should be corrected for regression dilution using re-measurements taken about 15 years or more apart. Failure to do so will have significant implications for the interpretation of the study data. In many studies, however, no re-surveys have been performed and, even where re-surveys have been done, they are usually within only a few years of the original baseline survey. This presents a serious limitation on the interpretation and application of the findings from these studies. Yet, using data from other population-based cohorts, such as Framingham, where re-measurements are available over prolonged follow-up, may provide a way to present and interpret the results appropriately.

Ideally studies would take repeat measurements every few years from a representative sample of their original cohort to allow a reasonably accurate assessment of the size of the bias over time, and to account for it in the analysis. Many epidemiological studies have re-survey data from a sample of the original cohort in which the exposures (that is, the risk factors under study) were later re-measured, and hence, the potential to account for the bias is available. However, despite the obvious need to do so, and the availability of statistical methods, correction for regression dilution bias is still seldom undertaken in published analyses of prospective studies. This might be because investigators are unaware of the extent to which the use of a single measurement underestimates the true risk, or because they are unaware of the methods available to correct for this bias. Whatever the reasons of it, however, the widespread failure to make proper correction for the regression dilution bias means that the relevance of many risk factors for disease has been seriously underestimated.

Figures

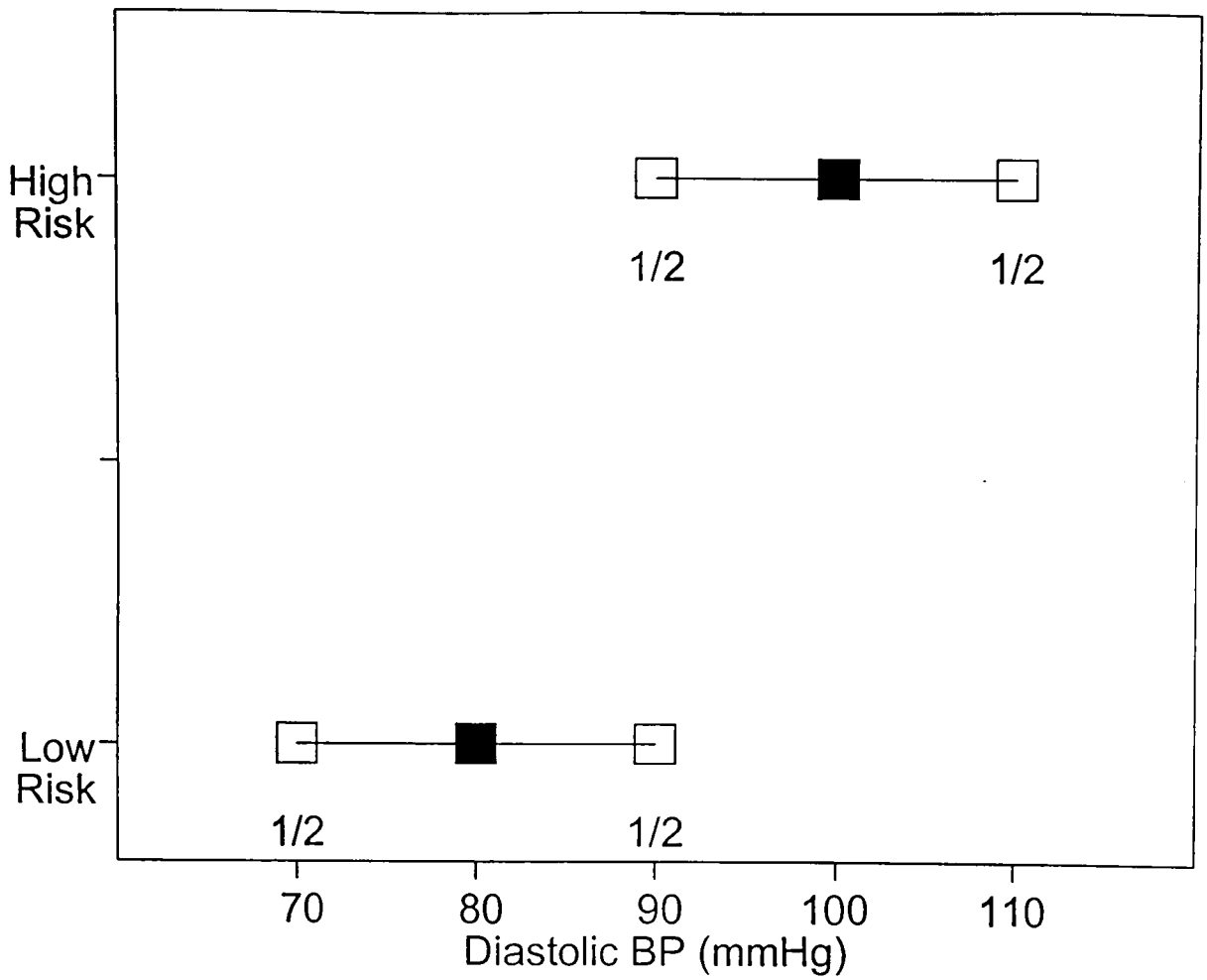
Figure 1: Large prospective study in ideal world



Measurements give just TWO groups:

exactly 80 / exactly 100

Figure 2: Real and measured values of DBP (mmHg)



Mean of "baseline" and "re-survey" in groups defined by "baseline" measurements

	Group		
	L	M	H
True	80	80 or 100	100
Baseline	70	90	110
Re-measurement	80±10	80±10 , 100±10	100±10
	√	√	√

Figure 3: Ideal world (80/100) + random errors (±10)

→ the WRONG answer

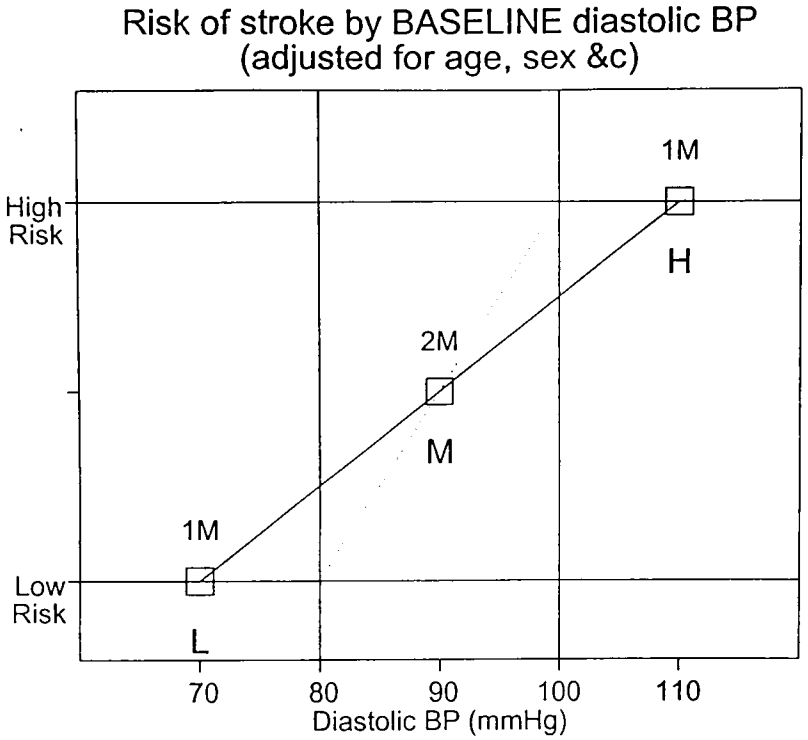


Figure 4: Ideal world (80/100) + random errors (±10) + correction (using re-survey)

→ the RIGHT answer

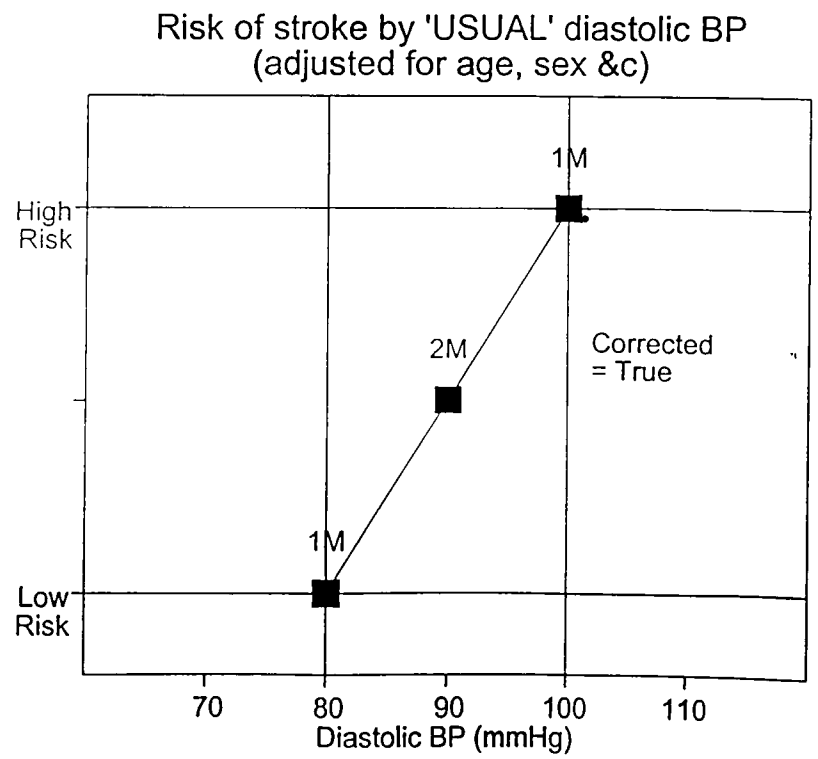


Figure 5: The Horse-Racing Effect

Possible diastolic BP "tracks" for two individuals with DBP = 90 mmHg at age 50

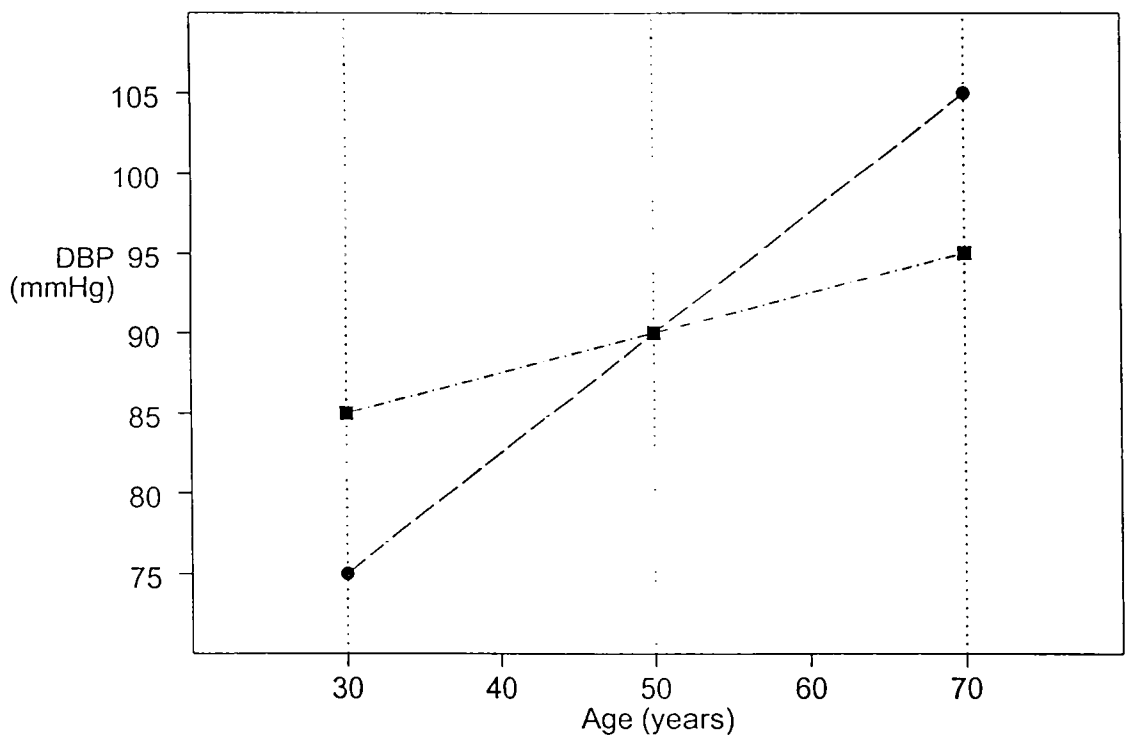
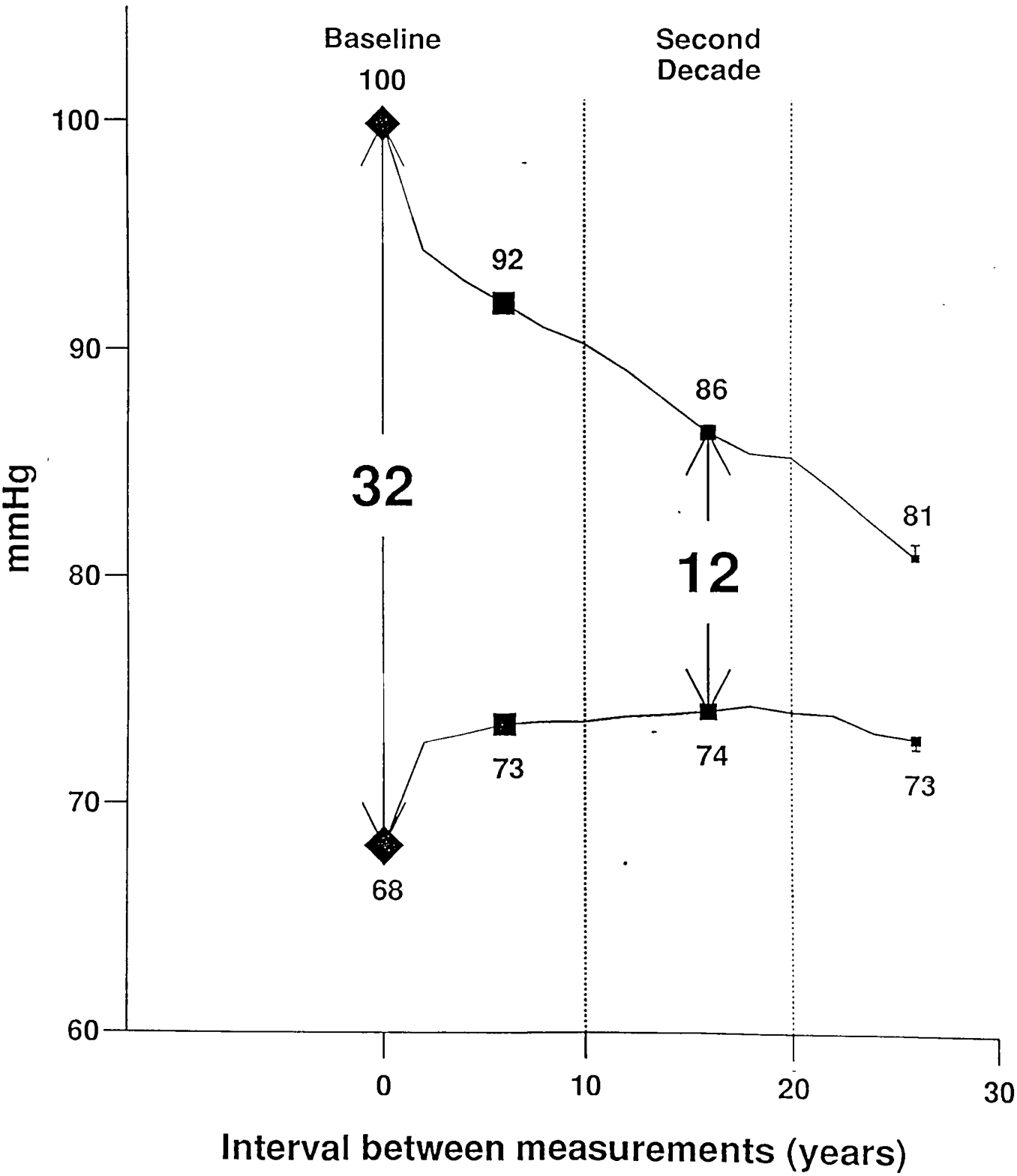


Figure 6: Mean values of diastolic BP at biennial follow-up in the Framingham study, for groups defined by first measurement



Chapter 3: A worldwide collaboration of prospective observational studies

Data collection and statistical methods

Chapter summary

The associations of blood pressure and cholesterol with the risks of stroke and CHD have been investigated in many prospective observational studies, and the Prospective Studies Collaboration (PSC) has become a worldwide collaboration between the principal investigators of such studies. This chapter defines the study selection criteria, data collection process and statistical methods used for an appropriate analysis of data from individuals in all relevant studies.

1 Study selection criteria and identification of studies

To enable this research, collaboration was sought from the investigators of all prospective observational studies in which data on blood pressure, cholesterol, date of birth (or age) and sex were recorded at a baseline screening visit, and in which cause and date of death (or age at death) were routinely sought thereafter. Relevant studies were identified through computer searches of MEDLINE and EMBASE, by hand-searches of meeting abstracts and by discussion with investigators and experts in the field. The aim was to include all studies with more than 5,000 person-years of follow-up provided participants were not selected on the basis of a positive disease history or diagnosis.

2 Risk factors being studied

The primary risk factors of interest are blood pressure (systolic and diastolic) and blood total cholesterol (subsequently referred to as total cholesterol). Secondary risk factors are lipid fractions (HDL cholesterol, and the remaining "non-HDL" cholesterol as a surrogate for LDL-cholesterol (see Chapter 4: Study Characteristics)), relative weight (that is, weight relative to the square of height – the so-called "body mass index"), alcohol consumption and smoking, but these secondary risk factors will not be considered for the purposes of this thesis. Data for each of these factors were sought from the baseline examination, and from all subsequent examinations conducted during follow-up. Data on age at baseline and at death (or last follow-up) and sex were sought to allow age- and sex-specific analyses. Data on possible effect-modifiers (such as ethnicity, and prior heart disease, stroke or diabetes) were also sought for consideration (Table 1). Many studies recorded smoking and alcohol consumption in categories (e.g. never, ex, light, medium, heavy) that could not be quantified reliably, but such data could nevertheless be used to adjust for confounding. A few studies have provided only self-reported measurements of blood pressure or total cholesterol, and a sensitivity analysis has been performed to determine the relevance of such self-reported results. For the remaining studies, information on the methodology used to assess risk factors (for example, measurement techniques, conditions under which measurements were made, etc.) was sought from the collaborators.

2.1 Re-survey data for the primary risk factors

Only about half of the studies involved in the PSC could provide information on individuals with more than one measurement of blood pressure or cholesterol taken some years apart. Hence, when investigating regression dilution only those studies that could provide data from at least 500 individuals with blood pressure or cholesterol measured at both an initial screening visit and at least one re-survey have been considered. To ensure that this subset was representative of the entire population, a comparison was made between the baseline characteristics of this subset of studies, and those of the one million individuals contributing to the outcome analyses.

3 Outcomes being studied

Cause-specific mortality data were obtained in the greatest detail available in the data files of the collaborators. Deaths were considered in various specific groups of causes (Table 2), each contributing at least 500 deaths. Other causes were analysed if they were responsible, in the aggregated data, for over 250 deaths (which provided 95% power to detect a two-fold relative risk between the top and bottom fifths of a risk factor distribution at the 5% significance level (Breslow and Day, 1987)).

4 Data transfer and checking

Data from individual studies were transferred using any machine-readable medium that was locally available (e.g. tape; disc; e-mail; anonymous ftp). Mortality and all of the risk factor data were accepted in whatever format they were originally coded and

stored by the study investigators. The data obtained from each participating study were checked for internal consistency and with published results. Any queries were referred back, in confidence, to the study collaborator. The data were then converted to a standard format for incorporation into a central database to be used for all collaborative analyses. The content of the data remained unchanged by this process, and detailed computer-generated summary tabulations based on the converted data were returned to each collaborator for review and confirmation.

The data provided from each study remained entirely the property of the principal investigators of that study, and were held in confidence. Anonymous data on individual participants in each of the studies were stored securely on the central computer database.

5 Statistical Methods

The fundamental method of analysis used to assess the separate associations of the risk factors with mortality was Cox's "proportional hazards" model (Cox, 1972) stratified by study, sex and age at risk. Only individuals with data available for each of systolic blood pressure, diastolic blood pressure, total cholesterol, sex, age, survival duration and cause of death (when appropriate) were included in the analyses. This allowed direct comparison between analyses of the importance of blood pressure and total cholesterol in the same individuals. In the subset for whom the necessary data were available, some of the analyses of blood pressure and cholesterol were, when appropriate, adjusted for (relative) weight, alcohol, smoking, ethnicity and disease history. Analyses were conducted of the relationships of the

primary risk factors to death for men and women separately, for sub-groups of age at risk, and for different periods of follow-up. (N.B. Such period-specific analyses remain appropriate whether or not the "proportional hazards" assumption is valid. Indeed, comparisons between them provide a direct test of its validity.) To explore the possibility that associations during the early period may reflect the effects of pre-existing disease on the risk factor, some analyses were repeated after exclusion of patients who died within five years of baseline.

Cox's model was used in two ways to describe the dose-response relationships. The first involved the estimation of regression coefficients per unit increase in exposure (i.e. the hazard ratio per one mmol/litre increase in usual cholesterol or per 10 mmHg increase in usual blood pressure). Such estimates describe only the *steepness* of the straight lines that best fit the data when risk is plotted on a logarithmic scale. A fuller description of the *shape* of the relationships was obtained by plotting the relative hazard of death in groups defined by baseline measurements of the risk factor of interest against the mean "usual" levels of the risk factor during the period of interest. To avoid issues of different measurement techniques between studies, individuals were ranked *within* their study according to baseline measurements of the risk factor of interest, and then divided into similar sized groups. Confidence intervals for each risk factor level were estimated using "floating" variances (Easton, Peto, and Babiker, 1991). These floating variances do not alter the estimates of the relative hazard, but do reduce the variance attributed to those that are not exactly unity (and also greatly reduce the unwanted covariances between them). All analyses were adjusted for the

effects of the "regression dilution" bias, and therefore relate risk to the usual levels of the relevant risk factors.

5.1 Variances of floating "absolute" risks

Suppose that, apart from the baseline group, there are *n* other groups. The logarithm of the relative risk and the variance of this log relative risk is calculated for each group in the usual way. For group *i*, let c_i denote the average of the (*n-1*) covariances between the log relative risk in that group and the (*n-1*) other log relative risks, and let the average of these *n* values (c_1 to c_n) be called *c*. To calculate "floating" variances, the variance attributed to the baseline log relative risk becomes *c*, and the variance of the i^{th} log relative risk has ($2c_i$ - *c*) subtracted from it. Letting the subscript zero denote the baseline group (in which the log relative risk has no covariance with anything else, so $c_0 = 0$), the covariance between the i^{th} and the j^{th} log relative risk has ($c_i + c_j$ *c*) subtracted from it. Although the procedure does not change their values at all, the relative risks are renamed as "floating absolute risks". The covariances between these log floating absolute risks are, in general, negligibly small.

5.2 Adjustments for regression dilution

MacMahon et al. (MacMahon et al., 1990) proposed two methods (one parametric, one non-parametric) to correct analyses of prospective observational data using two imprecise measurements taken some years apart, and it is these methods that will be considered in more detail throughout this thesis. However, the principles

developed apply to any correction method preferred by the analyst. The non-parametric methods proposed by MacMahon et al. do not assume either that the "measurement error" is due to purely random fluctuation or that the true level of the exposure of interest remains constant over time. The parametric method provides a good approximation to the non-parametric method when these assumptions hold.

Terminology used throughout this thesis:

Baseline levels, usual levels, exposure period, risk period and regression dilution ratio

The value of a risk factor recorded from a measurement made at the initial screening visit (year 0 by definition) of a prospective study will be referred to as the "*baseline*" measurement (or value) and will be denoted by Z_0. The true underlying level (or average) of the risk factor during the period of interest (the "*exposure period*") will be referred to as the "*usual*" level (or value) and will be denoted by X_t (so X_0 would be the *usual* level at the time of the initial screening visit). The *baseline* measurement will not accurately reflect the *usual* level during any *exposure period* due to the combined effects of measurement errors, longer-term physiological fluctuations or systematic changes over time (see earlier). Yet, using data from prospective studies, it is often desirable to relate exposure during a particular period (the "*exposure period*") to the risk of disease during a particular period (the "*risk period*"). The *risk period* may well be the same as the *exposure period*, but it does not have to be. For example, one could choose to relate stroke rates during the second decade of follow-up to the usual blood pressure in that decade, or to the usual blood pressure during the previous decade.

Let S_u (where *u* denotes *usual*) denote the "*steepness*" of the relationship between the outcome rate (e.g. stroke rate) during the *risk period* and the *usual* level of the risk factor during the *exposure period*. Let S_b (where *b* indicates *baseline*) denote the *steepness* of the relationship between the same outcome rate (during the same *risk period*) and the *baseline* measurement of the risk factor. Finally, let *RDR*, the regression dilution ratio, be the ratio of S_b to S_u (i.e. $RDR = S_b / S_u$). To avoid any unjustifiable assumptions about straight-line relationships between outcome rates and risk factors, "*steepness*" (in this context) is defined as follows:

First, use the *baseline* measurements to divide people into a few groups, from L (those with the lowest values) to H (those with the highest values). Next, define the "range of *baseline* values" (r_b) to be the difference between the mean of the *baseline* values in group L and the mean of the *baseline* values in group H; and the "range of *usual* values" (r_u) to be the difference between the mean of the *usual* values in group L and the mean of the *usual* values in group H (NOTE: the groups L and H are defined only by the *baseline* measurements and do not otherwise depend on the *usual* values). Then, the *steepness* S_b is the difference in the outcome rates between groups L and H divided by r_b, while the *steepness* S_u is the difference in the outcome rates between groups L and H divided by r_u, so $RDR = S_b/S_u = r_u/r_b$. It is important to note that *RDR* depends only on the relationship between the *baseline* measurements and the *usual* values of the risk factor during the exposure period and is, therefore, independent of both the disease being studied and the *risk period* of interest.

The MacMahon-Peto methods of correction for the regression dilution bias

Non-parametric corrections

In the non-parametric method proposed by MacMahon et al. for estimating, and correcting for, the regression dilution bias associated with a particular *exposure period*, individuals with pairs of measurements separated by an appropriate interval (e.g. from the time of the *baseline* survey to the mid-point of the *exposure period*) are sub-divided into a few similar-sized groups according to the *baseline* measurement, Z_0 (with the lowest values in group *L* and the highest values in group *H*). The means of these *baseline* measurements in each of the groups are then calculated. Next, although the re-measurements, Z_t, do not determine which group that individual is in, the means of the re-measurements in each of the groups are calculated. This second set of group means provides an unbiased estimate of the *usual* levels of the risk factor in each group during the *exposure period*.

More formally this can be described as follows, in the special case where the usual value is constant, and where the baseline measurement and the re-measurement are far enough apart for the random errors in them to be uncorrelated:

Suppose $Z_0 = X + \varepsilon_0$ and $Z_t = X + \varepsilon_t$ where X is the usual underlying exposure (assumed to be constant over time in this case). ε_0 and ε_t are the measurement errors, and are uncorrelated with each other and with X. By definition, ε_0 and ε_t both have zero mean, so

$$E(Z_t \mid Z_0 \in (a,b]) = E(X + \varepsilon_t \mid X + \varepsilon_0 \in (a,b])$$

$$= E(X \mid X + \varepsilon_0 \in (a,b]) + E(\varepsilon_t \mid X + \varepsilon_0 \in (a,b])$$

$$= E(X \mid Z_0 \in (a,b])$$

since ε_t has zero mean and is uncorrelated with both X and ε_0.

Often, the range of the second set of means will be substantially narrower than the range of the baseline set of means, so the ratio of these two ranges (i.e. the estimate of the regression dilution ratio) will be substantially less than 1. This implies that, if the relative risks for each group are plotted, not against the mean of the *baseline* measurements for each group, but rather against the **mean of the re-measurements** the slope associating risk with exposure will be somewhat steeper (see Introduction: Epidemiology Gremlin). This method can readily be extended to prospective studies with prolonged follow-up, where X may not necessarily remain constant over time (Clarke et al. [*in press*]).

This "ratio of ranges", r, provides an assumption-free estimate of the regression dilution ratio, *RDR*, during the *exposure period* that allows appropriately for the effects of all other sources of variation (including selective survival or tracking). The regression coefficient relating risk (during the *risk period*) to *usual* levels of a risk factor during the *exposure period* can be estimated as $1/r$ times the "uncorrected" regression coefficient relating risk to the *baseline* measurements. MacMahon et al. do not account for imprecision in their correction method when calculating confidence intervals for the "corrected" plot.

Parametric correction

When certain assumptions are met, for example when the values recorded at baseline and re-measurement have similar variances, parametric methods can be used to assess the magnitude of the regression dilution bias. The parametric method proposed by MacMahon et al. involves calculating the correlation between the baseline values and the re-survey values (the "self-correlation" denoted by ρ_{self}).

$$\rho_{self} = \frac{\mathrm{cov}(Z_0, Z_t)}{\sqrt{\mathrm{var}(Z_0)\mathrm{var}(Z_t)}}$$

$$\approx \frac{\sigma^2_{true}}{\sigma^2_{true} + \sigma^2_{error}} \qquad (6)$$

(when the errors, ε_0 and ε_t, are uncorrelated with each other and with X).

The regression coefficient relating risk (during the *risk period*) to *usual* levels of a risk factor during the *exposure period* can be estimated as $1/\rho_{self}$ times the "uncorrected" regression coefficient relating risk to the *baseline* measurements.

Standard deviations and confidence intervals for these "corrected" log odds ratios are obtained by dividing the standard errors of the "uncorrected" regression coefficients by r or ρ_{self}. This, in effect, assumes that r and ρ_{self} are known constants, and does not account for the fact that they have also been estimated from the data, and so have a standard error associated with them. Bashir and Duffy (Bashir and

Duffy, 1997) and Frost and Thompson (Frost and Thompson, [*in press*]) have derived variance estimates which account for the uncertainty in the correction factor.

Comparison between non-parametric and parametric methods

The parametric approach should provide similar estimates of the *RDR* to those obtained by the non-parametric method when the baseline measurements and re-measurements have similar variances. Estimates based on a self-correlation (using all the data) will have less random variation than those based on a ratio of ranges (using only data from the top and bottom groups), which may make them preferable if there are only a limited amount of data available. Nevertheless, since the non-parametric method requires no assumptions, it may be more reliable for estimating the regression dilution ratio over longer periods of follow-up. Neither of the methods proposed by MacMahon et al. require a validation study or the existence of a "gold standard" and are easily incorporated into a conventional analysis. Furthermore, they make no assumptions about the distributions of Z_0 and Z_t, or about the shape of the relation between Z_0 and Z_t.

Assessing the magnitude of regression dilution in each study of the PSC by the interval of follow-up

The magnitude of the *RDR* at each re-survey in each study was estimated using first the non-parametric and then the parametric methods of MacMahon et al.

Non-parametric methods

The following was performed for each cohort:

1. The baseline measurements were divided into fifths, and individuals allocated a group number according to which fifth their baseline measurement fell (with the lowest values in group I and the highest values in group V).

2. The means of the re-measurements within each of these groups were calculated separately for each survey, and plotted against the average time to the survey. [NOTE: the mean of the entire cohort will subsequently be referred to as the "overall mean" to distinguish from these group means.]

3. The absolute difference between the top and bottom fifths was calculated at each visit, and this difference was called the "interquintile" range for that visit.

4. The ratio of the interquintile range at baseline to that at the re-survey, i.e. the ratio of ranges (r), was calculated for each re-survey within each study and then plotted together against the average time to re-survey.

Parametric method

Correlation coefficients were computed between the baseline and re-survey measurements, ρ_{self}, within each study (separately for each re-survey) and plotted against the average time to re-survey.

A regression line of the estimated *RDR*s against time has been fitted separately to both the ratios of ranges and self-correlations. These regression lines were weighted by the number of people contributing to each point estimate of *RDR*. Only data from surveys where at least 500 individuals were re-surveyed have been included

because any estimates of the *RDR* based on smaller numbers would be too dominated by the play of chance.

Comparisons of regression dilution ratios from studies in different regions of the world

Based on the differing distributions of risk factors and differing mortality patterns observed in different countries (WHO MONICA Project Principal Investigators, 1988; Stegmayr et al., 1997; Murray and Lopez, 1994), studies in the PSC were divided into different regions. By far the most data in the PSC came from Europe (700,000 individuals) followed by North America, Japan and China, with the least data from Australia. There were no ideal groupings of these countries, but due to the differing patterns of disease between Europe, North America and East Asia there was some justification for choosing these three broad "regions". There were only two studies from Australia – Busselton and Perth – of which only the Busselton study had conducted re-surveys of a sample of the cohort. Thus, for practical, rather than any more defensible, reasons, these two studies have been included in a "North America & Australia" region.

Framingham "multiple pairs" and a comparison with other studies

The Framingham Heart Study involved 5209 men and women who were aged 30 to 62 when first examined between 1948 and 1952 (Anderson, Castelli, and Levy, 1987). Participants were invited every two years for 16 consecutive surveys to have their blood pressure measured and a blood sample collected. Because measurements were made every two years it was possible to classify measurements

for each participant into "multiple pairs" separated by the same time interval. For example, to estimate the *RDR* from measurements separated by a 2-year interval, up to 15 pairs of measurements per person were constructed using data from the first (i.e. "baseline" at year 0) and second (i.e. year 2) examinations, the second and third, the third and fourth, and so on, up to the fifteenth and sixteenth examinations. In this way around 5,000 individuals contributed 53,000 pairs with a 2-year interval between them. Pairs separated by 4 years were constructed using measurements from the first (i.e. "baseline" at year 0) and third (i.e. year 4) examinations, the second and fourth, the third and fifth, and so on, up to the fourteenth and sixteenth examinations. Pairs separated by 6 years, 8 years, etc. up to 30 years were constructed in a similar way. (A missing value in either one of a pair of measurements resulted in a missing value for that pair). The *RDR* was then estimated in these "multiple pairs" using the methods described above.

Results from the studies contributing to the regression dilution analyses in the PSC were compared with those obtained from the Framingham "multiple pairs".

Studies excluded from the analyses of regression dilution

The Framingham Heart Study

Our analyses of data from the Framingham Heart Study (Clarke et al. [*in press*]) (summarised in Chapter 2: "Influence of interval between measurements on regression dilution") generated the hypothesis of a need for time-dependent correction procedures, and are therefore excluded from any combined analyses. The

results obtained from Framingham will, however, be presented for comparison with those from the remaining studies in the PSC.

Studies in which risk factors were self-reported

In two North American studies - the US Physicians Health Study (Buring et al., 1995) and the US Health Professionals Follow-Up Study (Rimm et al., 1993) - individuals were asked to measure their own blood pressure and cholesterol and to send a completed questionnaire, including these measurements and details of other cardiovascular risk factors, to the co-ordinating centres for these studies. It was anticipated that the issues surrounding the regression dilution bias may have been somewhat different in these studies than from conventional observational studies, in which blood pressure was measured by an attendant nurse or physician and blood was taken from the participant for analysis at the study laboratory. It was anticipated that the potential for measurement error was far greater from these studies, and that issues of standardisation for methodology were more complex. Therefore, although the extent of the regression dilution bias will need to be assessed for these studies prior to correction of results from outcome analyses, they have been excluded for the purposes of this thesis.

Tables

Table 1: Data sought from collaborating studies

Table 2: Outcomes of interest

Table 1: Data sought from collaborating studies*

From Baseline Examination

- Some unique (but anonymous) identifier
- Date of birth (if available)
- Date of the baseline visit (or, if not available, age at baseline)
- Blood cholesterol (total and, if available, absolute levels HDL and LDL sub-fractions)
- Blood pressure (systolic and diastolic, if both are available)
- Gender and ethnicity
- History of heart disease, cerebrovascular disease and diabetes
- Height and weight
- Tobacco and alcohol use (current and past, if available)

From Follow-up

- Date last known to be alive (if not recorded as dead)
- Date of death (or, if not available, age at death)
- Underlying cause of death, preferably coded according to some specified version of the 3-digit International Classification of Diseases (but if a 3-digit ICD code is not available then whatever code the study already uses)

From Repeat Examinations

- The unique (but anonymous) identifier used for the baseline visit
- Date of the visit (or, if not available, age at visit)
- Blood cholesterol (total and, if available, absolute levels of HDL and LDL sub-fractions)
- Blood pressure (systolic and diastolic, if both are available)
- Height and weight
- Tobacco and alcohol use (current)

* Details of the coding conventions and the methodology used to assess risk factors were sought from all studies, together with a copy of the questionnaire or other data entry forms used.

Table 2: Outcomes of interest

In approximate order of ICD-9 code numbers:

- All neoplasms (140-239)
- Certain specific types of cancer, namely stomach (151), large intestine (153-154), liver (155), lung (162) and breast (174) will be considered separately
- All cardiovascular disease (390-459, 795)
- Ischaemic heart disease (410-414)
- All cerebrovascular disease (430-438)
- Subarachnoid haemorrhage (430)
- Haemorrhagic stroke (431-432)
- Ischaemic stroke (433-434)
- Other cardiovascular disease (390-459 excluding 430-438, 410-414)
- Chronic obstructive lung disease and related conditions (490-496)
- All liver disease (070, 570-573)
- All renal causes (403-404, 580-593)
- Violence, suicide and other trauma (800-999)
- Each other separate cause that is responsible, in the aggregated data, for over 250 deaths
- All cause mortality

Chapter 4: Study characteristics

136,000 deaths among 1 million adults during 13 million person-years of follow-up

Chapter summary

By June 1998, investigators of 59 studies in Europe, North America, Australia and Eastern Asia had contributed individual participant data on baseline blood pressure and cholesterol, together with mortality follow-up, for just over one million individuals. In addition, collaboration has been established with the MRFIT screening study of 343,000 middle-aged men, with 37,000 deaths during 5.5 million person-years of follow-up (Shaten et al., 1991) and with the secretariat of the Asia Pacific Cohort Studies Collaboration involving over 600,000 individuals (Eastern Stroke and Coronary Heart Disease Collaborative Research Group, 1998). Collaborative analyses with both these groups have been planned. Data were then known to be unavailable for only 75,000 participants from four studies known to be eligible (Daviglus et al., 1997; Stamler et al., 1993; Kodama, Sasaki, and Shimizu, 1990; Gordon and Doyle, 1987), although it is likely that some smaller studies have not been included. Cause-specific mortality data were available on 136,000 deaths during 13 million years of participant follow-up. About 45% of all deaths were due to vascular causes and 20% were due to cancer. This chapter describes the baseline characteristics and mortality follow-up of the study populations, and discusses some data issues.

1 Details of participating studies

The principal investigators of the 59 studies contributing data to the Prospective Studies Collaboration (PSC) are given in Appendix I. Relevant references for each study (as indicated by the principal investigators) and any acronyms used by the study investigators, and hence throughout the thesis, are given in Appendix II.

2 Baseline risk factor data

Some basic characteristics of each cohort are given in Table 1. 70% of participants came from Europe, 20% from North America, 9% from Eastern Asia and the remaining 1% from Australia (the two Australian cohorts have been presented with those from North America.) Cohorts varied in size from 828 (Seven Countries: Netherlands) to 221,738 (IPC-Paris). 40% of the participants were female: 39% in Europe, 52% in North America & Australia and 20% in East Asia. The mean age at screening in each cohort ranged from 34 to 73 years, and the mean for the overview population was 46 years. Participants from Europe were, on average, 8 years younger than participants from North America & Australia, and 3 years younger than participants from East Asia. Baseline risk factor levels were measured at various times throughout the last half-century. The earliest study – Framingham - began in 1949, and the most recent study - Ohasama - in 1990. The mean age at death in the overview population was 68, ranging from 50 years in the Netherlands Consultation Bureau Project on Cardiovascular Diseases (CB project) to 79 years in the Cardiovascular Health Study (CHS).

2.1 Blood pressure and cholesterol

Data on blood pressure and total cholesterol from a baseline visit together with cause-specific follow-up for mortality were available for about 90% of participants. Blood pressure data were available for a greater proportion of participants than were cholesterol. The majority of participants for whom cholesterol data were unavailable were from studies in which blood pressure and cholesterol were self-reported (see later).

The mean values of systolic and diastolic blood pressure (SBP and DBP) are shown in Table 2. The mean(SD) SBP and DBP were 135(21) mmHg and 82(12) mmHg, respectively. There was considerable variation between cohorts in the mean blood pressures at baseline that, after investigation, could not be explained by regional differences, age at which the blood pressure was measured or the year in which it was measured. However, both SBP and DBP were, on average, highest in the European cohorts and lowest in the North American & Australian cohorts (Table 2). Figure 1(a) and 1(b) show the distributions of the recorded values of SBP and DBP in the overview population. 99% of the SBP measurements lay between 96 and 219 mmHg, and 99% of the DBP measurements lay between 53 and 119 mmHg. These histograms illustrate clear digit preference for measurements as multiples of 10.

Figure 2(a) shows the distributions of the recorded values of total cholesterol in the overview population. 99% of the baseline total cholesterol measurements lay between 3.0 and 8.9 mmol/l. The mean(SD) value in the entire population was 5.7(1.3) mmol/l. The mean cholesterol levels in the East Asian cohorts, however, were all below 5 mmol/l, which contrasts with mean values ranging from 5.0 mmol/l in

Croatia up to 6.9 mmol/l in Oslo from the remaining cohorts. The mean values for each study are shown in Table 3.

Although some of the cholesterol and blood pressure values were considered too extreme to be real, unless refuted by the investigators (when reviewing the summary statistics prepared for them), these values were retained in the database. However, non-parametric correction for regression dilution will prevent these values, which did not provide a true reflection of the person's "usual" cholesterol or blood pressure, from biasing the results. The substantial differences between cohorts in the mean values of all the variables were considered when assessing heterogeneity.

2.2 Cholesterol fractions

Data on cholesterol fractions (high density lipoprotein (HDL), low density lipoprotein (LDL) and triglycerides) were reported in some studies, although data on LDL-cholesterol were far less complete than on HDL-cholesterol, and data on triglycerides were available from only three studies. HDL-cholesterol data were available for 178,000 individuals, of whom only 46,000 also provided data on LDL-cholesterol. 90% of the HDL-cholesterol measurements fell between 0.6 and 2.9 mmol/l (Figure 2 (b)). The mean(SD) HDL-cholesterol was 1.4(0.4) overall, and the mean values in individual studies ranged from 0.9 (Israel) to 2.0 (Saitama) (Table 3). The mean(SD) non-HDL-cholesterol was 4.4(1.3) mmol/litre, with 90% of the values between 2.5 and 6.5 mmol/litre. The distribution of total cholesterol values among this subset with HDL-measurements was similar to the whole population (mean(SD)=5.7(1.2) mmol/l). There was only 10% correlation between total and HDL cholesterol values.

In the 46,000 individuals with an LDL cholesterol measurement, the mean(SD) total cholesterol was 5.6(1.1) mmol/l and the mean HDL-cholesterol was 1.35(0.44) mmol/l. The mean non-HDL-cholesterol was 4.2(1.2) mmol/l and the mean LDL-cholesterol was 2.7(1.8) mmol/l. There was surprisingly little correlation between non-HDL cholesterol levels and the reported LDL cholesterol levels (only 30%), perhaps because the LDL measurement techniques were not reliable. Partly because there were far fewer data on LDL cholesterol and partly because of these uncertainties in the reliability of the available measurements, the LDL data are not considered further in this thesis. However, although beyond the scope of this thesis, a detailed investigation of the relationship between non-HDL and LDL cholesterol will be performed before any analyses of non-HDL cholesterol are published. There was 94% correlation between non-HDL cholesterol and total cholesterol.

2.3 Smoking and alcohol

Information on smoking at baseline, in differing degrees of detail, was available from 90% of participants. Of those, approximately 40% were classified as never having smoked cigarettes (29% male and 56% female), 15% as ex-smokers and 45% as current smokers. However, since in some populations many individuals may have quit smoking after the baseline survey, whilst in other populations the tobacco epidemic is still evolving, these baseline proportions will not represent the true proportions during prolonged follow-up. An indication of this is can be obtained by comparing the proportions of never, ex- and current smokers at baseline in the different regions. In East Asia, where people have only been smoking for relatively short periods, there was a much smaller proportion of ex-smokers (data not shown).

Two studies could provide information only on current smoking habits, making it impossible to distinguish between never- and ex-smokers, and two further studies could provide information only on ever smoking, making it impossible to distinguish between ex- and current-smokers. Only one study could not provide any information on smoking habits at baseline.

Although inadequate to assess fully the role of smoking as a risk factor for disease, these data should allow some preliminary investigations of the effects (or lack of effects) of adjusting blood pressure and cholesterol analyses for smoking.

Information on drinking habits was available from fewer than one-third of the participants, and suffers from the usual problems of inaccurate recall or reporting. Furthermore, in many studies, information was recorded only on current drinking habits, making it impossible to distinguish between never drinkers and ex-drinkers. The data from the 300,000 participants in these studies suggested that 40% (32% of males and 53% of females) were not current drinkers. However, some studies simply recorded answers to questions such as "What did you drink yesterday?". If the response was "Nothing", then that person would currently be classed as a non-drinker in the central database. Thus, most of the information on alcohol is currently of little value. Nevertheless, where alcohol is considered to be an important confounder, the effects of adjustment can be investigated in those few studies able to provide at least moderately reliable alcohol data.

The proportions of "never" smokers and "not current" drinkers for males and females in each study are given in Table 4.

2.4 Prior history of disease

One of the inclusion criteria for studies was that participants must not have been selected on the basis of a prior history of disease, although those with a medical history were not specifically excluded. Most studies recorded at least brief details of prior diseases. From these data, participants were re-classified as having no history of disease (based on the available data from each study), having a history of disease (that is, a definite or probable past history based on the available data within each study), or not known (including those for whom the information provided was ambiguous). As one would expect in studies of primarily healthy populations, the vast majority of participants were disease-free on entry to the study. Of those providing information, 7% had data suggesting history of heart disease, less than one percent stroke, and 4% diabetes. The proportions in each study are given in Table 5. However, because the methods used to ascertain medical histories differed between studies (from brief questioning to detailed physical examinations), these reported proportions will not adequately reflect the true differences between studies.

3 Cause of death data

The numbers of deaths for various specific groups of causes (which are the primary outcomes of interest) are given for each study in Table 6. All studies provided data on stroke mortality, and most studies provided cause of death as a 3-digit ICD code (versions 6 to 10). However, some studies provided the cause of death only if it was one of a list of pre-defined causes of interest (although all but two of these studies provided cause of death data for each of the primary outcomes of interest given in Chapter 3, Table 2). As a consequence, the proportion of all deaths attributed to

some diseases, for example stomach cancer, is less than the true proportion of all participants who died from those diseases.

Cardiovascular diseases and cancer accounted for the majority of deaths. Overall, 64,000 deaths – almost half of all deaths - were due to vascular diseases. 40,000 of these deaths were ascribed to ischaemic heart disease (IHD) and 13,000 to cerebrovascular diseases ("stroke"). Some of the remaining 11,000 vascular deaths were ascribed to pulmonary embolus, but most were probably due to heart disease of some sort (including the late effects of IHD). Of the 13,000 stroke deaths, only 41% were ascribed to specific types of stroke, i.e. cerebral ischaemia (subsequently referred to as "ischaemic"), primary intra-cerebral haemorrhage (subsequently referred to as "haemorrhagic") or subarachnoid haemorrhage. Of those so classified, 40% were attributed to ischaemic stroke, 45% to haemorrhagic stroke and 15% to subarachnoid haemorrhage. Unclassified strokes consisted both of stroke deaths from studies that did not provide data on stroke sub-types, and of stroke deaths coded as "ill-defined". Furthermore, even where the stroke sub-type was provided, misclassification of stroke sub-types will cause systematic underestimation of any differences between haemorrhagic and ischaemic strokes in their associations with blood pressure and cholesterol. 30,000 deaths were ascribed to cancer, the largest proportion of which were lung cancer (6,500 deaths).

There were substantial variations between studies and regions in the proportions of deaths ascribed to various causes. In Europe, vascular diseases caused half of all deaths, and the majority of these (about 70%) were ascribed to ischaemic heart disease. 40% of all deaths in North America and Australia were caused by vascular

diseases: 56% IHD and 20% stroke. The proportion of vascular deaths was lower in East Asia, and the ratio of IHD to stroke deaths was reversed. In East Asia, just over one-third of all deaths were attributed to vascular diseases, and over half of these were due to stroke.

Figures and tables

Figure 1: Frequency distribution of (a) SBP and (b) DBP

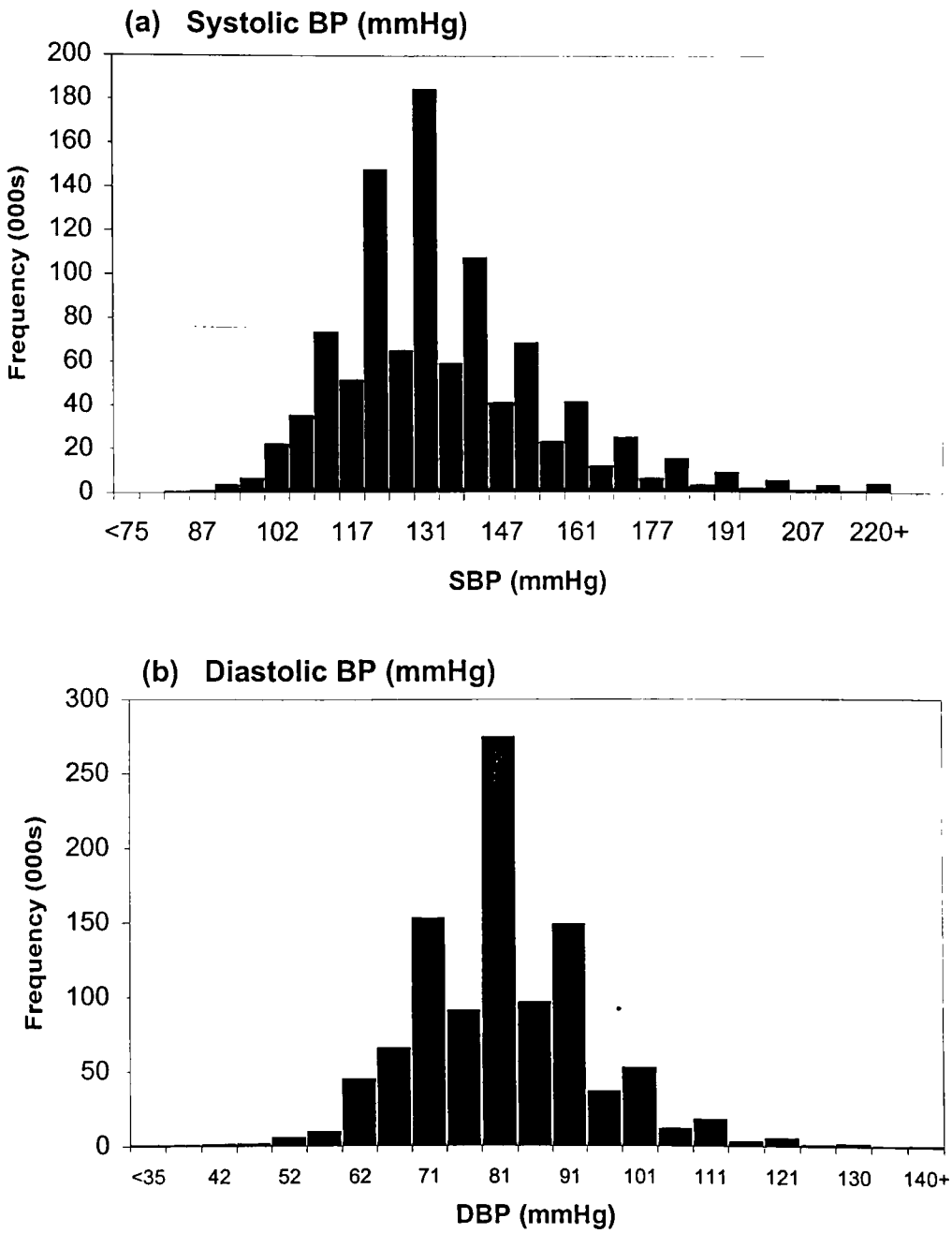


Figure 2: Frequency distribution of (a) total cholesterol and (b) HDL cholesterol

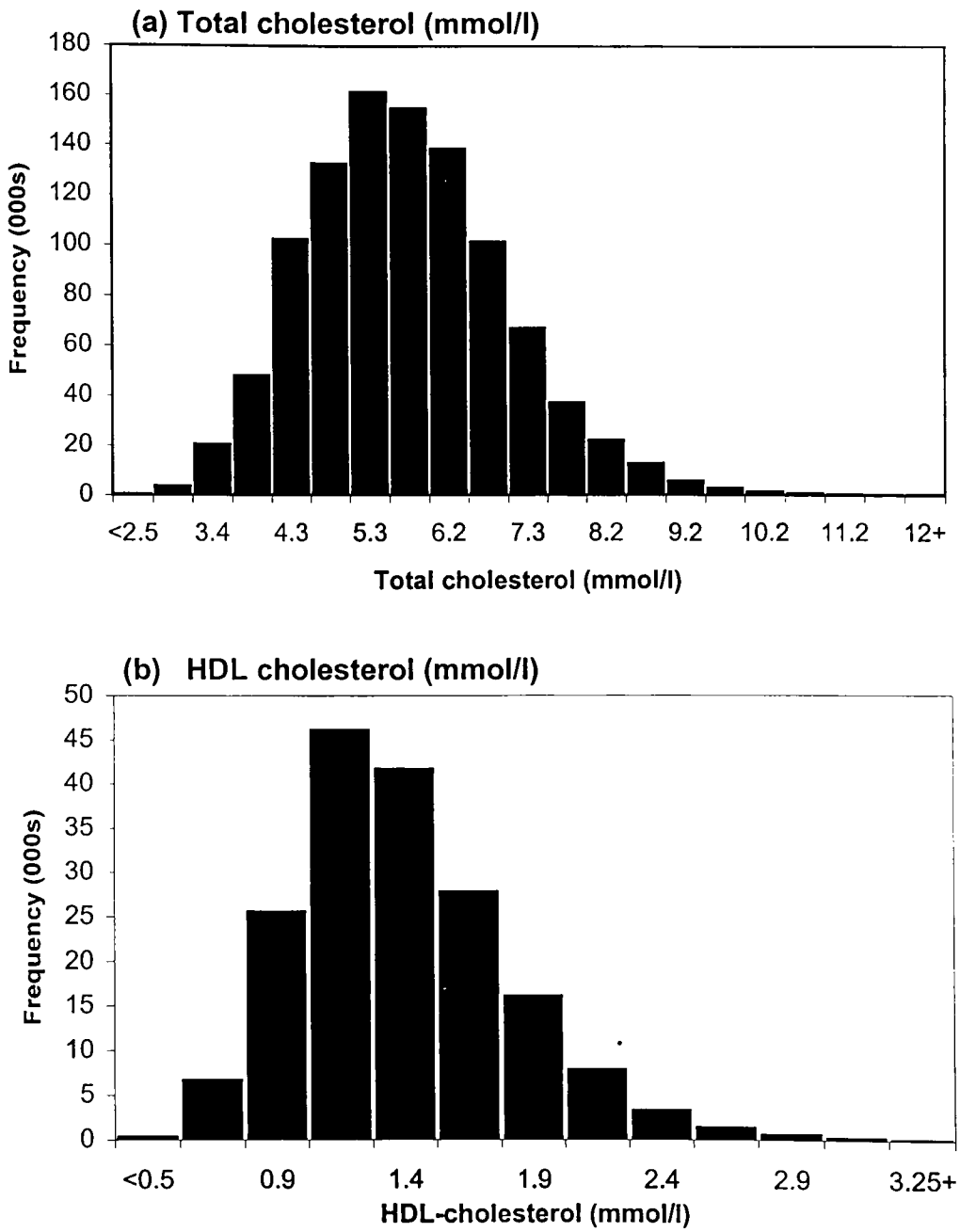


Table 1: Age at screening and at death, and time to death and last follow-up

Study (year screening began)	Number analysed*	Mean age at baseline (range)	Mean age at death (range)	Mean time to death (max)	Mean follow-up (max)
EUROPE					
BIRNH (1980)	10,611	49 (25-74)	67 (26-83)	5 (10)	10 (10)
BRHS (1978)	7,681	50 (38-61)	62 (40-76)	9 (16)	13 (16)
BUPA (1975)	21,508	47 (35-64)	64 (37-83)	10 (19)	15 (19)
Caerphilly (1979)	2,313	52 (44-60)	62 (46-73)	7 (15)	12 (15)
CB project (1974)	48,835	39 (30-54)	50 (33-69)	9 (17)	11 (22)
Copenhagen (1976)	13,762	53 (20-93)	70 (27-104)	10 (18)	14 (18)
Finnish Mobile Clinic (1966)	51,250	40 (15-92)	69 (15-105)	12 (26)	20 (26)
Finrisk (1972)	38,658	43 (24-64)	61 (26-81)	10 (22)	15 (22)
Glostrop (1977)	10,133	48 (29-80)	69 (31-96)	6 (16)	8 (16)
Gothenburgh Women (1968)	1,456	47 (37-61)	68 (42-86)	17 (26)	24 (27)
IPC-PARIS (1978)	221,738	41 (16-95)	61 (20-100)	8 (16)	11 (16)
Israel (1963)	9,900	49 (40-75)	67 (41-90)	14 (24)	20 (24)
Leuven (1985)	1,059	49 (19-90)	73 (42-92)	4 (9)	7 (10)
Norwegian (~1975)	48,613	42 (35-49)	54 (35-67)	10 (8)	16 (18)
NPHS (1972)	3,188	46 (17-87)	67 (25-95)	11 (22)	18 (23)
OG-Rome (1979)	3,344	55 (46-66)	63 (48-75)	6 (11)	9 (12)
Oslo (1972)	16,203	45 (39-50)	60 (41-67)	11 (18)	16 (18)
Paris (1967)	7,540	47 (43-53)	62 (44-76)	14 (26)	21 (26)
PROCAM (1978)	13,613	47 (35-68)	56 (36-75)	4 (11)	7 (16)
Renfrew/Paisley (1972)	15,262	54 (45-64)	66 (46-84)	9 (19)	15 (19)
SC Croatia (1958)	1,336	50 (39-59)	67 (42-82)	15 (24)	19 (24)
SC Finland (1959)	1,660	49 (39-59)	65 (42-84)	14 (24)	19 (25)
SC Greece (1960)	1,176	49 (39-59)	68 (42-82)	16 (24)	21 (24)
SC Italy (1960)	2,435	49 (39-59)	66 (41-82)	7 (24)	20 (25)
SC Netherlands (1960)	828	49 (39-60)	67 (44-84)	15 (24)	20 (24)
SC Serbia (1962)	1,559	48 (39-59)	66 (42-83)	15 (24)	20 (25)
SHHS (1984)	11,688	49 (25-65)	59 (36-74)	5 (11)	9 (11)
Speedwell (1979)	2,241	54 (45-63)	64 (49-76)	7 (15)	12 (15)
Tromso (1979)	16,595	34 (20-54)	51 (21-65)	6 (11)	11 (11)
UK HDPP (1971)	13,594	51 (39-66)	66 (41-83)	13 (24)	19 (24)
Varmland (~1970)	96,137	49 (22-89)	72 (25-89)	12 (24)	19 (24)
West Scotland (1970)	6,961	48 (21-75)	65 (37-90)	13 (24)	20 (24)
Whitehall (1967)	18,155	52 (39-69)	64 (41-82)	8 (14)	14 (15)
SUBTOTAL	**721,032**	**44 (15-95)**	**67 (15-105)**	**11 (26)**	**14 (27)**
N AMERICA & AUSTRALIA					
ARIC (1986)	15,526	54 (44-66)	62 (45-74)	4 (9)	7 (10)
Busselton (1966)	6,895	44 (16-93)	75 (21-103)	14 (29)	21 (30)
Charleston (1960)	2,161	50 (35-97)	72 (36-100)	17 (34)	23 (34)
CHS (1989)	5,158	73 (64-104)	79 (66-104)	2 (4)	4 (5)
Evans County (1960)	3,101	47 (15-79)	71 (19-100)	15 (30)	22 (30)
Framingham (1949)	3,172	44 (29-62)	72 (32-100)	25 (40)	28 (41)
Honolulu (1965)	7,961	54 (45-68)	73 (47-93)	16 (27)	23 (30)
LRC (1972)	8,649	47 (28-96)	69 (30-104)	9 (19)	17 (20)
MHHP (1980)	17,257	45 (24-75)	66 (26-84)	4 (10)	5 (10)
MHS (1981)	9,925	45 (24-85)	65 (35-81)	3 (9)	3 (9)
NHEFS (1971)	13,903	49 (24-77)	74 (30-95)	10 (21)	15 (22)
Perth (1979)	9,735	46 (17-90)	69 (29-102)	9 (18)	12 (18)
Puerto-Rico HHP (1965)	9,762	54 (35-79)	64 (40-85)	7 (13)	11 (13)
Rancho Bernardo (1971)	6,089	52 (12-99)	79 (18-102)	11 (20)	16 (21)
Tecumseh (1959)	4,129	47 (35-91)	71 (38-101)	14 (28)	20 (28)
US Health Profs (1986)	24,264	55 (39-77)	67 (41-82)	4 (8)	7 (8)
US Nurses (1988)	51,069	55 (41-69)	62 (43-72)	3 (5)	7 (5)
US Physicians (1981)	7,713	54 (39-85)	70 (45-96)	7 (12)	12 (14)
SUBTOTAL	**206,469**	**52 (12-100)**	**71 (18-104)**	**12 (40)**	**10 (41)**
ASIA					
Akita (1975)	8,728	54 (40-69)	69 (42-87)	9 (18)	13 (19)
Japan Railway: EJR (1973)	54,828	43 (6-58)	53 (26-66)	5 (11)	7 (12)
Ohasama (1990)	2,666	56 (26-86)	69 (37-91)	2 (5)	4 (5)
Saitama (1986)	3,623	54 (20-94)	71 (37-95)	4 (8)	6 (19)
SC Japan (1958)	917	50 (39-60)	67 (43-82)	15 (24)	19 (24)
Seven Cities:China (1986)	8,372	57 (35-90)	67 (51-84)	2 (4)	3 (4)
Shanghai (1972)	9,017	48 (30-78)	66 (41-92)	11 (20)	15 (20)
Shibata (1977)	2,328	57 (40-89)	76 (42-96)	8 (15)	13 (15)
SUBTOTAL	**90,479**	**47 (6-94)**	**65 (11-26)**	**8 (24)**	**8 (24)**
TOTAL	**1,017,980**	**46 (6-100)**	**68 (15-105)**	**11 (40)**	**13 (41)**

*Individuals with data available for each of sex, status, SBP, DBP and total cholestrol

Table 2: Mean (SD) systolic and diastolic BP at baseline

Study (year screening began)	Number analysed*	Mean (SD) systolic BP at baseline	Mean (SD) diastolic BP at baseline
EUROPE			
BIRNH (1980)	10,611	134 (19)	81 (12)
BRHS (1978)	7,681	145 (21)	82 (13)
BUPA (1975)	21,508	132 (18)	82 (12)
Caerphilly (1979)	2,313	141 (19)	89 (13)
CB project (1974)	48,835	130 (17)	79 (11)
Copenhagen (1976)	13,762	137 (22)	83 (12)
Finnish Mobile Clinic (1966)	51,250	141 (24)	78 (14)
Finrisk (1972)	38,658	143 (21)	87 (12)
Glostrop (1977)	10,133	127 (20)	80 (11)
Gothenburgh Women (1968)	1,456	133 (22)	86 (11)
IPC-PARIS (1978)	221,738	132 (15)	81 (11)
Israel (1963)	9,900	135 (21)	84 (11)
Leuven (1985)	1,059	129 (19)	76 (9)
Norwegian (~1975)	48,613	134 (17)	83 (11)
NPHS (1972)	3,188	139 (23)	86 (15)
OG-Rome (1979)	3,344	135 (19)	85 (11)
Oslo (1972)	16,203	136 (16)	87 (11)
Paris (1967)	7,540	141 (22)	80 (14)
PROCAM (1978)	13,613	131 (19)	85 (11)
Renfrew/Paisley (1972)	15,262	149 (24)	86 (13)
SC Croatia (1958)	1,336	138 (20)	83 (12)
SC Finland (1959)	1,660	144 (21)	86 (11)
SC Greece (1960)	1,176	136 (21)	82 (11)
SC Italy (1960)	2,435	143 (20)	86 (12)
SC Netherlands (1960)	828	145 (20)	89 (13)
SC Serbia (1962)	1,559	133 (19)	84 (11)
SHHS (1984)	11,688	132 (20)	82 (12)
Speedwell (1979)	2,241	140 (23)	87 (14)
Tromso (1979)	16,595	126 (14)	80 (10)
UK HDPP (1971)	13,594	139 (19)	84 (12)
Varmland (~1970)	96,137	155 (27)	87 (13)
West Scotland (1970)	6,961	134 (18)	83 (10)
Whitehall (1967)	18,155	136 (21)	85 (14)
SUBTOTAL	**721,032**	**137 (21)**	**83 (12)**
N AMERICA & AUSTRALIA			
ARIC (1986)	15,526	121 (19)	74 (11)
Busselton (1966)	6,895	133 (23)	78 (14)
Charleston (1960)	2,161	146 (29)	86 (12)
CHS (1989)	5,158	136 (21)	70 (11)
Evans County (1960)	3,101	149 (32)	91 (17)
Framingham (1949)	3,172	138 (24)	86 (13)
Honolulu (1965)	7,961	134 (21)	82 (12)
LRC (1972)	8,649	127 (20)	80 (11)
MHHP (1980)	17,257	121 (17)	74 (11)
MHS (1981)	9,925	121 (17)	74 (11)
NHEFS (1971)	13,903	135 (24)	83 (13)
Perth (1979)	9,735	130 (20)	81 (12)
Puerto-Rico HHP (1965)	9,762	132 (23)	82 (12)
Rancho Bernardo (1971)	6,089	131 (24)	78 (12)
Tecumseh (1959)	4,129	138 (22)	83 (13)
US Health Profs (1986)	24,264	129 (13)	81 (17)
US Nurses (1988)	51,069	126 (14)	79 (9)
US Physicians (1981)	7,713	126 (12)	79 (8)
SUBTOTAL	**206,469**	**128 (19)**	**79 (11)**
ASIA			
Akita (1975)	8,728	137 (21)	81 (12)
Japan Railway: EJR (1973)	54,828	130 (17)	81 (12)
Ohasama (1990)	2,666	130 (17)	73 (11)
Saitama (1986)	3,623	135 (20)	80 (12)
SC Japan (1958)	917	135 (25)	76 (14)
Seven Cities:China (1986)	8,372	131 (24)	82 (12)
Shanghai (1972)	9,017	125 (22)	79 (13)
Shibata (1977)	2,328	131 (21)	78 (12)
SUBTOTAL	**90,479**	**130 (19)**	**81 (12)**
TOTAL	**1,017,980**	**135 (21)**	**82 (12)**

*Individuals with data available for each of sex, status, SBP, DBP and total cholestrol

Table 3: Mean (SD) total cholesterol at baseline, and total and HDL-cholesterol in those individuals with HDL-cholesterol measurements

Study (year screening began)	Number analysed *	Mean (SD) total cholesterol baseline	Number with HDL**	Mean (SD) total cholesterol baseline	Mean (SD) HDL cholesterol baseline
EUROPE					
BIRNH (1980)	10,611	6.08 (1.20)	10,522	6.07 (1.2)	1.39 (0.37)
BRHS (1978)	7,681	6.30 (1.04)			
BUPA (1975)	21,508	6 26 (1.13)	8,972	6.25 (1.16)	1.34 (0.37)
Caerphilly (1979)	2,313	5.71 (1.14)	2,305	5.72 (1.14)	1.12 (0.33)
CB project (1974)	48,835	5.36 (1.14)			
Copenhagen (1976)	13,762	6.12 (1.24)	1,608	5.96 (1.15)	1.45 (0.40)
Finnish Mobile Clinic (1966)	51,250	6.41 (1.44)			
Finrisk (1972)	38,658	6.43 (1.34)	15,551	6.08 (1.27)	1.38 (0.36)
Glostrop (1977)	10,133	6.12 (1.23)	10,109	6.12 (1.23)	1.47 (0.41)
Gothenburgh Women (1968)	1,456	6.82 (1.20)			
IPC-PARIS (1978)	221,738	5.53 (1.12)			
Israel (1963)	9,900	5.41 (1.04)	6,545	5.38 (1.02)	0.95 (0.24)
Leuven (1985)	1,059	5.87 (1.37)	708	5.93 (1.42)	1.29 (0.42)
Norwegian (~1975)	48,613	6.33 (1.26)			
NPHS (1972)	3,188	5.98 (1.24)			
OG-Rome (1979)	3,344	5.54 (1.04)	3,107	5.55 (1.03)	1.22 (0.28)
Oslo (1972)	16,203	6.94 (1.27)			
Paris (1967)	7,540	5.76 (1.12)			
PROCAM (1978)	13,613	5.81 (1.20)	13,613	5.81 (1.10)	1.27 (0.36)
Renfrew/Paisley (1972)	15,262	6.16 (1.07)			
SC Croatia (1958)	1,336	5.02 (1.07)			
SC Finland (1959)	1,660	6.75 (1.35)			
SC Greece (1960)	1,176	5.31 (1.12)			
SC Italy (1960)	2,435	5.26 (1.06)			
SC Netherlands (1960)	828	6.09 (1.18)			
SC Serbia (1962)	1,559	4.66 (1.08)			
SHHS (1984)	11,688	6.43 (1.26)	11,197	6.43 (1.25)	1.52 (0.42)
Speedwell (1979)	2,241	5.88 (1.21)	2,179	5.86 (1.20)	1.10 (0.37)
Tromso (1979)	16,595	5.84 (1.25)	16,584	5.84 (1.25)	1.59 (0.46)
UK HDPP (1971)	13,594	5.62 (1.04)			
Varmland (~1970)	96,137	6.46 (1.07)			
West Scotland (1970)	6,961	5.86 (1.04)			
Whitehall (1967)	18,155	5.11 (1.22)			
SUBTOTAL	**721,032**	**5.94 (1.26)**	**103,000**	**5.99 (1.22)**	**1.38 (0.42)**
N AMERICA & AUSTRALIA					
ARIC (1986)	15,526	5.56 (1.09)	15,525	5.56 (1.09)	1.33 (0.44)
Busselton (1966)	6,895	5.82 (1.29)	946	5.67 (1.36)	1.51 (0.38)
Charleston (1960)	2,161	6.09 (1.25)			
CHS (1989)	5,158	5.55 (1.01)	5,152	5.55 (1.01)	1.39 (0.41)
Evans County (1960)	3,101	5.40 (1.20)			
Framingham (1949)	3,172	5.73 (1.17)			
Honolulu (1965)	7,961	5.65 (0.99)			
LRC (1972)	8,649	5.84 (1.29)	8,606	5.84 (1.29)	1.32 (0.43)
MHHP (1980)	17,257	5.30 (1.06)	16,841	5.29 (1.06)	1.24 (0.37)
MHS (1981)	9,925	5.23 (1.05)	9,897	5.23 (1.05)	1.25 (0.39)
NHEFS (1971)	13,903	5.71 (1.25)	0		
Perth (1979)	9,735	5.77 (1.22)	6,501	5.27 (1.16)	1.36 (0.37)
Puerto-Rico HHP (1965)	9,762	5.22 (1.07)			
Rancho Bernardo (1971)	6,089	5.32 (1.07)			
Tecumseh (1959)	4,129	5.36 (1.04)			
US Health Profs (1986)	24,264	5.25 (0.94)			
US Nurses (1988)	51,069	5.35 (1.24)			
US Physicians (1981)	7,713	5.48 (1.16)			
SUBTOTAL	**206,469**	**5.45 (1.16)**	**63,468**	**5.55 (1.13)**	**1.30 (0.41)**
ASIA					
Akita (1975)	8,728	4.85 (0.91)			
Japan Railway: EJR (1973)	54,828	4.59 (0.89)			
Ohasama (1990)	2,666	5.06 (0.95)	341	4.91 (0.92)	1.34 (0.36)
Saitama (1986)	3,623	4.99 (0.96)	3,546	4.99 (0.96)	2.00 (0.41)
SC Japan (1958)	917	4.26 (0.9)			
Seven Cities:China (1986)	8,372	4.86 (1.22)	7,867	4.86 (1.20)	1.39 (0.46)
Shanghai (1972)	9,017	4.19 (0.86)		-	
Shibata (1977)	2,328	4.61 (1.21)		-	
SUBTOTAL	**90,479**	**4.63 (0.96)**	**11,754**	**4.90 (1.13)**	**1.58 (0.52)**
TOTAL	**1,017,980**	**5.72 (1.28)**	**178,222**	**5.73 (1.23)**	**1.37 (0.43)**

*Individuals with data available for each of sex, status, SBP, DBP and total cholestrol

** Together with each of sex, status, SBP, DBP and total cholesterol

Table 4: Proportion[#] of never smokers and non-drinkers

		Males			Females		
Study (year screening began)	**Number in analyses***	**% male**	**% never smokers**	**% not current drinkers**	**% female**	**% never smokers**	**% not current drinkers**
EUROPE							
BIRNH (1980)	10,611	53	21	38	47	74	61
BRHS (1978)	7,681	100	24	30	0	-	-
BUPA (1975)	21,508	100	30	42	0	-	-
Caerphilly (1979)	2,313	100	16	6	0	-	-
CB project (1974)	48,835	48	64	x	52	48	x
Copenhagen (1976)	13,762	46	11	12	54	27	30
Finnish Mobile Clinic (1966)	51,250	53	30	x	47	78	-
Finrisk (1972)	38,658	49	21	x	51	67	x
Glostrop (1977)	10,133	50	19	9	50	36	26
Gothenburgh Women (1968)	1,456	0	-	-	100	52	48
IPC-PARIS (1978)	221,738	57	(y/n)	x	43	(y/n)	x
Israel (1963)	9,900	100	31	x	0	-	x
Leuven (1985)	1,059	50	(y/n)	x	50	(y/n)	x
Norwegian (~1975)	48,613	51	23	-	49	50	-
NPHS (1972)	3,188	71	27	64	29	49	71
OG-Rome (1979)	3,344	100	16	0	0	-	-
Oslo (1972)	16,203	100	19	x	0	-	-
Paris (1967)	7,540	100	26	x	0	-	-
PROCAM (1978)	13,613	76	34	x	24	65	x
Renfrew/Paisley (1972)	15,262	46	17	x	54	46	x
SC Croatia (1958)	1,336	100	28	x	0	-	-
SC Finland (1959)	1,660	100	19	x	0	-	-
SC Greece (1960)	1,176	100	24	x	0	-	-
SC Italy (1960)	2,435	100	23	x	0	-	-
SC Netherlands (1960)	828	100	8	x	0	-	-
SC Serbia (1962)	1,559	100	34	x	0	-	-
SHHS (1984)	11,688	51	22	18	49	41	36
Speedwell (1979)	2,241	100	16	6	0	-	-
Tromso (1979)	16,595	51	50	x	49	53	x
UK HDPP (1971)	13,594	100	8	x	0	-	x
Varmland (~1970)	96,137	50	n/a	x	50	n/a	x
West Scotland (1970)	6,961	100	23	35	0	-	-
Whitehall (1967)	18,155	100	18	x	0	-	-
SUBTOTAL	**721,032**	**61**	**27 (16)**	**27 (5)**	**39**	**56 (27)**	**41 (7)**
N AMERICA & AUSTRALIA							
ARIC (1986)	15,526	45	23	35	55	52	51
Busselton (1966)	6,895	48	31	19	52	65	42
Charleston (1960)	2,161	45	(ever)	-	55	(ever)	-
CHS (1989)	5,158	43	32	41	57	57	54
Evans County (1960)	3,101	48	38	-	52	83	-
Framingham (1949)	3,172	47	(y/n)	-	53	(y/n)	-
Honolulu (1965)	7,961	100	30	37	0	-	-
LRC (1972)	8,649	53	37	10	47	34	21
MHHP (1980)	17,257	46	35	42	54	51	60
MHS (1981)	9,925	47	37	-	53	50	-
NHEFS (1971)	13,903	41	29	37	59	58	60
Perth (1979)	9,735	53	40	-	47	61	(ever)
Puerto-Rico HHP (1965)	9,762	100	34	-	0	-	-
Rancho Bernardo (1971)	6,089	46	35	-	54	53	-
Tecumseh (1959)	4,129	49	(ever)	16	51	(ever)	45
US Health Profs (1986)	24,264	100	46	23	0	-	-
US Nurses (1988)	51,069	0	-	-	100	45	40
US Physicians (1981)	7,713	100	52	16	0	-	-
SUBTOTAL	**206,469**	**48**	**38 (36)**	**28 (20)**	**52**	**50 (47)**	**45 (34)**)
ASIA							
Akita (1975)	8,728	42	20	33	58	93	95
Japan Railway: EJR (1973)	54,828	100	21	22	0	-	-
Ohasama (1990)	2,666	39	38	34	61	97	92
Saitama (1986)	3,623	38	16	23	62	90	67
SC Japan (1958)	917	100	16	-	0	-	-
Seven Cities:China (1986)	8,372	43	40	57	57	80	96
Shanghai (1972)	9,017	69	39	x	31	93	x
Shibata (1977)	2,328	42	22	26	58	95	91
SUBTOTAL	**90,479**	**80**	**24 (19)**	**32 (22)**	**20**	**90 (90)**	**91 (91)**
TOTAL	**1,017,980**	**60**	**29 (12)**	**28 (6)**	**40**	**56 (35)**	**50 (17)**

*Individuals with data available for each of sex, status, SBP, DBP and total cholestrol
(y/n) Only information on current smoking was available
(ever) Only information on ever smoking available
x No data available
% of those providing information (% of all included in analyses), with "never smokers" and "not current drinkers" determined from the data available within each study

Table 5: Prior# history of disease

Study (year screening began)	Number analysed*	% with possible# prior history of Heart disease	Stroke	Diabetes
EUROPE				
BIRNH (1980)	10,611	4	-	2
BRHS (1978)	7,681	4	1	1
BUPA (1975)	21,508	2	-	1
Caerphilly (1979)	2,313	22	2	2
CB project (1974)	48,835	0	0	1
Copenhagen (1976)	13,762	16	8	2
Finnish Mobile Clinic (1966)	51,250	8	-	2
Finrisk (1972)	38,658	2	1	3
Glostrop (1977)	10,133	6	-	2
Gothenburgh Women (1968)	1,456	4	-	1
IPC-PARIS (1978)	221,738	13	-	6
Israel (1963)	9,900	2	-	6
Leuven (1985)	1,059	-	-	-
Norwegian (~1975)	48,613	1	0	1
NPHS (1972)	3,188	3	1	0
OG-Rome (1979)	3,344	4	1	-
Oslo (1972)	16,203	2	0	1
Paris (1967)	7,540	1	0	3
PROCAM (1978)	13,613	0	0	5
Renfrew/Paisley (1972)	15,262	27	1	1
SC Croatia (1958)	1,336	3	0	-
SC Finland (1959)	1,660	9	3	-
SC Greece (1960)	1,176	4	1	-
SC Italy (1960)	2,435	6	1	-
SC Netherlands (1960)	828	5	2	-
SC Serbia (1962)	1,559	38	33	-
SHHS (1984)	11,688	3	1	1
Speedwell (1979)	2,241	20	1	2
Tromso (1979)	16,595	0	0	0
UK HDPP (1971)	13,594	3	0	1
Varmland (~1970)	96,137	-	-	-
West Scotland (1970)	6,961	6	-	1
Whitehall (1967)	18,155	-	-	1
SUBTOTAL	**721,032**	**8%**	**1%**	**3%**
N AMERICA & AUSTRALIA				
ARIC (1986)	15,526	5	2	7
Busselton (1966)	6,895	9	1	1
Charleston (1960)	2,161	4	-	3
CHS (1989)	5,158	10	4	11
Evans County (1960)	3,101	3	2	5
Framingham (1949)	3,172	2	1	1
Honolulu (1965)	7,961	4	1	26
LRC (1972)	8,649	8	-	
MHHP (1980)	17,257	2	1	-
MHS (1981)	9,925	2	1	4
NHEFS (1971)	13,903	6	2	6
Perth (1979)	9,735	4	1	2
Puerto-Rico HHP (1965)	9,762	6	-	-
Rancho Bernardo (1971)	6,089	5	2	4
Tecumseh (1959)	4,129	3	-	10
US Health Profs (1986)	24,264	9	1	4
US Nurses (1988)	51,069	2	1	5
US Physicians (1981)	7,713	-	-	-
SUBTOTAL	**206,469**	**5%**	**1%**	**6%**
ASIA				
Akita (1975)	8,728	1	2	4
Japan Railway: EJR (1973)	54,828	-	-	2
Ohasama (1990)	2,666	7	1	16
Saitama (1986)	3,623	3	1	2
SC Japan (1958)	917	0	0	-
Seven Cities:China (1986)	8,372	13	-	-
Shanghai (1972)	9,017	-	-	-
Shibata (1977)	2,328	1	2	1
SUBTOTAL	**90,479**	**6%**	**2%**	**3%**
TOTAL	**1,017,980**	**7%**	**1%**	**4%**

*Individuals with data available for each of sex, status, SBP, DBP and total cholestrol

- Information was not available

% of those providing information, with "possible" prior disease determined from the data available within each study within each study

Table 6: Number of deaths due to the primary outcomes of interest

(a) Vascular deaths

Study (year screening began)	Number analysed*	All cardio-vascular	IHD	MI	Cerebro-vascular	Sub. Haemo-rrhage	Haemo-rrhagic stroke	Isch-aemic stroke #	Other cardio-vascular
EUROPE									
BIRNH (1980)	10,611	400	158	131	78	3	30	19	164
BRHS (1978)	7,681	624	546	-	78				-
BUPA (1975)	21,508	1,029	760	487	118	5	21	17	151
Caerphilly (1979)	2,313	243	195	136	24	1	4	4	24
CB project (1974)	48,835	392	231	203	72	26	17	11	89
Copenhagen (1976)	13,762	1,050	633	393	183	9	47	22	234
Finnish Mobile Clinic (~1965)	51,250	6,811	4,172	-	1,483	203	257	801	1,156
Finrisk (1972)	38,658	2,740	1,903	1,343	488	86	132	200	349
Glostrop (1977)	10,133	447	268	153	76	7	18	14	103
Gothenburgh Women (1968)	1,456	109	82	-	27	-			-
IPC-PARIS (1978)	221,738	1,462	633	470	356	28	117	20	473
Israel (1963)	9,900	1,669	1,014	786	355	10	46	12	300
Leuven (1985)	1,059	44			-	-	-	-	-
Norwegian (~1975)	48,613	1,504	1,009	748	202	70	56	21	293
NPHS (1972)	3,188	217	147	75	18	7	0	1	52
OG-Rome (1979)	3,344	164	127	79	29	1	9	2	8
Oslo (1972)	16,203	1,057	833	603	88	17	25	20	136
Paris (1967)	7,540	616	202	149	105	5	46	11	309
PROCAM (1978)	13,613	121	94	-	17	0	9	5	10
Renfrew/Paisley (1972)	15,262	2,268	1,553	1,304	469	23	51	48	246
SC Croatia (1958)	1,336	322	95	85	138	1	1	0	89
SC Finland (1959)	1,660	524	368	322	79	4	14	31	77
SC Greece (1960)	1,176	176	44	40	80	1	0	0	52
SC Italy (1960)	2,435	473	217	181	131	0	14	8	125
SC Netherlands (1960)	828	213	148	131	31	2	4	0	34
SC Serbia (1962)	1,559	368	135	123	129	0	4	0	104
SHHS (1984)	11,688	332	239	177	56	8	7	4	37
Speedwell (1979)	2,241	266	196	170	43	1	8	12	27
Tromso (1979)	16,595	150	123	91	11	3	3	2	16
UK HDPP (1971)	13,594	1,978	1,479	968	252	32	29	26	247
Varmland (~1970)	96,137	20,278	12,926		4,371	182	1,247	508	2,981
West Scotland (1970)	6,961	1,033	743	572	179	8	19	20	111
Whitehall (1967)	18,155	1,771	1,306	1,024	214	25	41	26	251
SUBTOTAL	**721,032**	**50,851**	**32,579**	**10,944**	**9,980**	**768**	**2,276**	**1,865**	**8,248**
N AMERICA & AUSTRALIA									
ARIC (1986)	15,526	372	204	104	42	4	8	6	126
Busselton (1966)	6,895	1,122	652	500	256	10	34	44	214
Charleston (1960)	2,161	775	430	197	152	7	36	39	193
CHS (1989)	5,158	277	152	81	49	3	8	9	76
Evans County (1960)	3,101	261	141	80	63	0	1	11	57
Framingham (1949)	3,172	803	486	-	148		-	-	169
Honolulu (1965)	7,961	1,080	556	367	361	36	85	100	163
LRC (1972)	8,649	214	114	47	33	2	8	3	67
MHHP (1980)	17,257	237	144	76	32	1	9	4	61
MHS (1981)	9,925	60	33	14	11	1	0	2	16
NHEFS (1971)	13,903	2,209	1,245	722	388	14	58	74	576
Perth (1979)	9,735	255	160	123	50	4	8	4	45
Puerto-Rico HHP (1965)	9,762	837	374	175	186	6	43	123	277
Rancho Bernardo (1971)	6,089	1,108	617	294	218	6	31	37	273
Tecumseh (1959)	4,129	829	452	328	155	10	17	26	222
US Health Profs (1986)	24,264	450	347	58	44	6	• 16	12	59
US Nurses (1988)	51,069	222	134	34	49	12	14	8	39
US Physicians (1981)	7,713	240	119	51	38	3	11	7	83
SUBTOTAL	**206,469**	**11,351**	**6,360**	**3,251**	**2,275**	**125**	**387**	**509**	**2,716**
ASIA									
Akita (1975)	8,728	351	50	44	152	13	59	55	149
Japan Railway: EJR (1973)	54,828	509	120	106	221	42	120	14	168
Ohasama (1990)	2,666	23	6	3	11	5	3	0	6
Saitama (1986)	3,623	76	14	14	27	5	8	14	35
SC Japan (1958)	917	150	32	29	97	1	16	21	21
Seven Cities:China (1986)	8,372	58			58	-			-
Shanghai (1972)	9,017	373	81		249	-	-	-	43
Shibata (1977)	2,328	260	45	30	159	13	27	62	56
SUBTOTAL	**90,479**	**1,800**	**348**	**226**	**974**	**79**	**233**	**166**	**478**
TOTAL	**1,017,980**	**64,002**	**39,287**	**14,421**	**13,229**	**972**	**2,896**	**2,540**	**11,442**

* Individuals with all of duration, sex, status, SBP, DBP and total cholesterol

- data not available

Only ICD-9 codes 433-434 are classified as ischaemic strokes. 436-438 are classified as "of unknown type".

Table 6

(b) Cancer deaths

Study (year screening began)	Number analysed*	All neoplasms	Large Intestine	Liver	Lung Cancer	Female Breast
EUROPE						
BIRNH (1980)	10,611	313	35	7	98	10
BRHS (1978)	7,681	0		-	-	#
BUPA (1975)	21,508	864	117	13	197	0
Caerphilly (1979)	2,313	135	16	1	46	#
CB project (1974)	48,835	555	45	8	134	95
Copenhagen (1976)	13,762	917	101	16	249	92
Finnish Mobile Clinic (~1965)	51,250	2,604	-	-	-	-
Finrisk (1972)	38,658	1,237	95	34	313	106
Glostrop (1977)	10,133	349	46	7	107	18
Gothenburgh Women (1968)	1,456	108	-			-
IPC-PARIS (1978)	9,900	689	72	13	120	#
Israel (1963)	221,738	3,059	327	102	661	195
Leuven (1985)	1,059	32	2	1	0	0
Norwegian (~1975)	48,613	1,212	162	9	191	111
NPHS (1972)	3,188	143	14	2	49	7
OG-Rome (1979)	3,344	169	14	9	46	#
Oslo (1972)	16,203	614	68	9	176	#
Paris (1967)	7,540	915	73	34	213	1
PROCAM (1978)	13,613	151	13	4	29	13
Renfrew/Paisley (1972)	15,262	1,391	154	13	482	104
SC Croatia (1958)	1,336	140	12	4	36	#
SC Finland (1959)	1,660	213	9	5	98	#
SC Greece (1960)	1,176	110	7	12	29	#
SC Italy (1960)	2,435	331	31	11	61	1
SC Netherlands (1960)	828	145	16	0	61	#
SC Serbia (1962)	1,559	167	20	12	30	#
SHHS (1984)	11,688	312	33	7	91	27
Speedwell (1979)	2,241	170	25	1	53	#
Tromso (1979)	16,595	141	13	0	19	18
UK HDPP (1971)	13,594	1,296	157	13	459	1
Varmland (~1970)	96,137		-	-	-	-
West Scotland (1970)	6,961	658	73	4	235	8
Whitehall (1967)	18,155	1,024	121	8	375	#
SUBTOTAL	**721,032**	**20,164**	**1,871**	**359**	**4,658**	**807**
N AMERICA & AUSTRALIA						
ARIC (1986)	15,526	395	31	3	128	35
Busselton (1966)	6,895	519	81	5	85	35
Charleston (1960)	2,161	279	24	2	81	21
CHS (1989)	5,158	157	17	4	36	11
Evans County (1960)	3,101	80	7	2	25	6
Framingham (1949)	3,172	522	-	-	-	-
Honolulu (1965)	7,961	1,173	169	40	297	#
LRC (1972)	8,649	121	17	4	30	9
MHHP (1980)	17,257	172	23	0	48	19
MHS (1981)	9,925	77	11	1	24	4
NHEFS (1971)	13,903	991	131	18	232	85
Perth (1979)	9,735	264	48	3	51	23
Puerto-Rico HHP (1965)	9,762	376	32	6	52	1
Rancho Bernardo (1971)	6,089	558	65	6	142	45
Tecumseh (1959)	4,129	351	47	7	83	26
US Health Profs (1986)	24,264	401	58	6	60	#
US Nurses (1988)	51,069	700	60	2	131	195
US Physicians (1981)	7,713	261	39	5	41	-
SUBTOTAL	**206,469**	**7,397**	**860**	**114**	**1,546**	*** 515**
ASIA						
Akita (1975)	8,728	344	27	22	52	8
Japan Railway: EJR (1973)	8,372	-	-		-	-
Ohasama (1990)	54,828	501	56	55	54	0
Saitama (1986)	2,666	29	7	1	6	0
SC Japan (1958)	3,623	88	6	8	17	1
Seven Cities:China (1986)	917	144	13	20	18	#
Shanghai (1972)	9,017	415	33	75	105	7
Shibata (1977)	2,328	145	10	6	17	3
SUBTOTAL	**90,479**	**1,666**	**152**	**187**	**269**	**19**
TOTAL	**1,017,980**	**29,227**	**2,883**	**660**	**6,473**	**1,341**

* Individuals with all of duration, sex, status, SBP, DBP and total cholesterol

- data not available

not applicable (all male cohort)

Study characteristics

Table 6

(c) Other deaths

Study (year screening began)	Number analysed*	COPD and related	All liver disease	All renal causes	Violence, suicide, trauma	All causes
EUROPE						
BIRNH (1980)	10,611	31	13	2	0	995
BRHS (1978)	7,681	-		-	-	1,389
BUPA (1975)	21,508	53	29	15	85	2,269
Caerphilly (1979)	2,313	24	6	3	0	448
CB project (1974)	48,835	17	25	2	0	1,286
Copenhagen (1976)	13,762	105	53	34	0	4,577
Finnish Mobile Clinic (~1965)	51,250	221	80	109	1,029	12,355
Finrisk (1972)	38,658	100	72	33	215	5,131
Glostrop (1977)	10,133	37	17	10	0	1,069
Gothenburgh Women (1968)	1,456	-	-	9	9	300
IPC-PARIS (1978)	9,900	79	152	33	0	7,187
Israel (1963)	221,738	61	40	55	0	3,403
Leuven (1985)	1,059	0	0	1	0	101
Norwegian (~1975)	48,613	67	35	21	0	3,573
NPHS (1972)	3,188	13	1	5	0	519
OG-Rome (1979)	3,344	3	17	2	0	399
Oslo (1972)	16,203	47	52	9	0	2,114
Paris (1967)	7,540	19	120	10	0	2,161
PROCAM (1978)	13,613	5	7	-	38	363
Renfrew/Paisley (1972)	15,262	163	41	40	49	4,406
SC Croatia (1958)	1,336	57	50	4	47	744
SC Finland (1959)	1,660	27	2	8	50	912
SC Greece (1960)	1,176	21	12	8	18	412
SC Italy (1960)	2,435	45	45	15	52	1,068
SC Netherlands (1960)	828	14	1	2	9	407
SC Serbia (1962)	1,559	57	8	8	50	713
SHHS (1984)	11,688	23	8	3	0	774
Speedwell (1979)	2,241	19	2	6	0	499
Tromso (1979)	16,595	4	1	2	0	410
UK HDPP (1971)	13,594	161	23	24	0	3,956
Varmland (~1970)	96,137	-	-	-	-	34,422
West Scotland (1970)	6,961	66	22	9	0	2,046
Whitehall (1967)	18,155	93	17	32	0	3,346
SUBTOTAL	**721,032**	**1,632**	**951**	**514**	**1,651**	**103,754**
N AMERICA & AUSTRALIA						
ARIC (1986)	15,526	30	9	10	18	1,061
Busselton (1966)	6,895	109	12	41	0	2,115
Charleston (1960)	2,161	37	14	25	0	1,397
CHS (1989)	5,158	22	4	12	0	570
Evans County (1960)	3,101	17	9	14	0	1,616
Framingham (1949)	3,172	-		-	-	1,980
Honolulu (1965)	7,961	144	60	38	0	3,423
LRC (1972)	8,649	14	5	5	0	1,458
MHHP (1980)	17,257	22	5	6	0	547
MHS (1981)	9,925	5	2	0	0	176
NHEFS (1971)	13,903	172	55	85	0	4,412
Perth (1979)	9,735	23	11	3	0	771
Puerto-Rico HHP (1965)	9,762	26	89	24	153	1,725
Rancho Bernardo (1971)	6,089	89	18	28	0	2,185
Tecumseh (1959)	4,129	56	19	17	0	1,505
US Health Profs (1986)	24,264	21	14	8	71	1,135
US Nurses (1988)	51,069	21	17	8	0	1,165
US Physicians (1981)	7,713	10	8	2	46	662
SUBTOTAL	**206,469**	**818**	**351**	**326**	**288**	**27,903**
ASIA						
Akita (1975)	8,728	12	21	27	0	1,019
Japan Railway: EJR (1973)	8,372	-	-	-	-	58
Ohasama (1990)	54,828	6	87	26	57	1,362
Saitama (1986)	2,666	0	0	0	1	72
SC Japan (1958)	3,623	3	4	0	14	222
Seven Cities:China (1986)	917	6	13	6	24	421
Shanghai (1972)	9,017	85	32	5	22	1,046
Shibata (1977)	2,328	19	8	10	0	609
SUBTOTAL	**90,479**	**131**	**165**	**74**	**118**	**4,809**
TOTAL	**1,017,980**	**2,581**	**1,467**	**914**	**2,057**	**136,466**

* Individuals with all of duration, sex, status, SBP, DBP and total cholesterol
- data not available

Chapter 5: Regression dilution in the Prospective Studies Collaboration - the size of the problem

Assessing the magnitude of the regression dilution ratio within each study with re-survey data

Chapter summary

The purpose of this chapter is to quantify the magnitude of the regression dilution ratios for blood pressure and cholesterol over prolonged follow-up, using repeated measurements of these factors from all the studies in the Prospective Studies Collaboration able to provide such data. This will help to determine whether, for these particular factors, regression dilution is importantly different for participants in different populations.

1 The data

Twenty-five of the 58 studies comprising the Prospective Studies Collaboration (PSC) provided re-survey data on at least a sample of their original cohort, and these are included in the analyses of regression dilution. Table 1 shows the distribution of repeat blood pressure measurements by region and by study. Data on blood pressure at an initial screening visit and at least one re-survey were available from approximately 100,000 individuals from Europe, 60,000 from North America & Australia and 50,000 from East Asia (Table 1). The numbers were smaller for cholesterol because some studies took blood from only a sample of those re-surveyed, and other studies did not take any blood at the re-surveys. Although only individuals with both blood pressure *and* cholesterol at the baseline visit were

included in the outcome analyses, participants with at least one re-measurement of either blood pressure *or* cholesterol were included in the regression dilution analyses. Therefore, some participants who were not eligible for the outcome analyses could still contribute to the regression dilution analyses. The proportion of participants contributing to both the outcome analyses and the regression dilution analyses was greatest in East Asia (approximately half) and least in Europe (only 15%).

The number of re-surveys within any one study ranged from 1 to 10. However, whilst there were two measurements of blood pressure (i.e., baseline and one re-survey) for over 200,000 individuals, there were three measurements for only 40,000 and five or more measurements for less than 10,000 (Figure 1). A frequency distribution of the longest time interval between measurements for each individual (Figure 2) also shows that the majority of the re-survey data came from the period 3 to 6 years after the initial measurement, with only 65,000 individuals providing any information beyond 5 years. Yet, the average time to death in these studies was about 10 years (11 years in Europe, 12 years in North America & Australia, and 8 years in East Asia). Thus, *regression dilution ratios* over a period of up to 15 years are required to interpret the outcome analyses appropriately.

1.1 Baseline data

Blood pressure

The mean age at baseline and the mean values and interquintile ranges (the absolute difference between the top and bottom fifths of the distribution: see Materials and methods) for systolic blood pressure (SBP), diastolic blood pressure (DBP) and total cholesterol in each study are shown in Table 2. In the European studies, the participants were, on average, 5 years younger at the time of the initial measurement than were participants from the North American & Australian studies or East Asian studies. SBP was higher, by about 10 mmHg, in the European studies (mean=136mmHg) than in the North American & Australian studies (mean=125mmHg), with intermediate values found in East Asian studies (mean=131mmHg). Whether these differences between regions in the mean SBP were due to ethnic differences is impossible to tell from these data, where the comparisons are between different generations in different countries with very different life experiences. Differences in measurement techniques, blood pressure management and various other lifestyle factors could also have contributed substantially to differences between the cohorts. The mean SBP ranged from 132 mmHg (IPC-Paris) to 149 mmHg (Paisley/Renfrew) in Europe; from 121 mmHg (ARIC and MHHP) to 149 mmHg (Evans County) in North America & Australia; and with a much narrower range from 125 mmHg (Shanghai) to 131 mmHg (Shibata) in East Asia. Thus, the differences between cohorts were most extreme in North America/Australia and least extreme in East Asia. But, the interquintile ranges were

remarkably similar between the regions: 50 mmHg in Europe, 54 mmHg in North America & Australia and 53 mmHg in East Asia. So, although the absolute levels of SBP varied considerably between studies and between regions, the differences between the top and bottom fifths were more consistent.

Similar observations were made for DBP. On average, this was highest in the European studies (mean=83mmHg), and lowest in the North American & Australian studies (mean=75 mmHg). The mean DBP in East Asian studies was 81mmHg. As for SBP, the mean DBP varied less between studies in Europe (range between study means = 8 mmHg) and East Asia (range between study means = 5 mmHg) than in North America & Australia (range between study means = 21 mmHg). Despite these differences in absolute values, the interquintile ranges were again remarkably similar between the regions (33 mmHg in Europe, 30 mmHg in North America & Australia and 35 mmHg in East Asia), which should simplify the investigation of regression dilution.

Total cholesterol

Higher mean total cholesterol levels were seen in participants from the European studies (5.9 mmol/litre), than the North American & Australian studies (5.2 mmol/litre), with the lowest levels observed in participants from the East Asian studies (4.7 mmol/litre). The interquintile ranges for total cholesterol were slightly larger in the European studies (3.2 mmol/l) than in the North American & Australian studies (2.8 mmol/l) or the East Asian studies (2.6 mmol/litre).

1.2 Re-survey data

Mean values over time within each cohort are shown for SBP and DBP in Figure 3 and for total cholesterol in Figure 4. There were modest changes in the mean values over time in some cohorts. In some cases, these changes may have been due to documented changes in the measurement methodology (e.g., in the Caerphilly/Speedwell studies, a Hawksley random zero sphygmomanometer was used initially but later replaced by a regular sphygmomanometer), but in most cases, the reasons for change were obscure. Possible reasons could include changes in study personnel resulting in different measurement techniques, changes in lifestyle, or changes over time in blood pressure and cholesterol management (including the introduction of more effective treatments during the course of the study). Selective mortality was not a major reason. These changes may also, at least in the smaller studies, reflect random fluctuations in the population mean caused by the play of chance. If anything, there was a tendency for SBP to increase slightly with age during follow-up, but only by a few mmHg, and not in all cohorts. There were no systematic temporal changes for DBP or total cholesterol, with the recorded mean levels consistently rising in some cohorts, consistently falling in some and fluctuating in others (Figure 3 and Figure 4).

2 Usual levels: shrinkage of the ranges and comparisons with such shrinkage in the Framingham study

In contrast to any modest changes in the overall means over time, substantial systematic artefactual changes were, as expected, seen at subsequent follow-up in the mean values of the top and bottom groups defined by baseline measurements. This was true for each of SBP, DBP and total cholesterol, and in every study. Serial changes in the mean values of groups defined by baseline measurements are shown for each study in Figure 5 (SBP), Figure 6 (DBP), and Figure 7 (total cholesterol). In these plots, the area of each square is proportional to the amount of statistical information (that is, to the number of participants in the group with a value both at baseline and at that re-survey). 95% confidence intervals are represented by vertical bars. Mean values in the top and bottom groups, and absolute differences between them (ranges) are shown at baseline (interval = 0) and at the final re-survey. Each figure is divided into three parts (or pages) with cohorts from each region represented on a separate page.

The most striking observation from these figures was the consistency between the cohorts in the magnitude and the pattern of convergence of the top and bottom groups. This convergence was somewhat more extreme for DBP than for SBP, and may have been slightly more extreme for SBP than for total cholesterol. The convergence was not necessarily uniform in all cohorts, but the consistency in the pattern was more remarkable than any differences. This was despite the diversity of populations from which these cohorts were selected, and the very different conditions under which the studies were conducted. In every cohort and for each

parameter, the range (i.e. the absolute difference between the means of the top and bottom groups) at the initial survey ("r_b") was greater than the range at any subsequent survey ("r_u"). Hence, the *regression dilution ratio* (estimated by r_u/r_b) was always less than unity. In nearly all cohorts, the range narrowed with increasing follow-up, and so the regression dilution ratio became more extreme with increasing follow-up. Thus, for every study, uncorrected risk relationships based on analyses using the means of the baseline measurements would *under-estimate* the true relationships. Although the consistency of this phenomenon between studies is remarkable, the analysis of individual studies is informative.

2.1 Systolic Blood Pressure

The interquintile ranges of the baseline values seemed to show greater variation than the subsequent interquintile ranges of the usual values. For example, the interquintile range at baseline in MHHP was 46 mmHg , whereas in the CHS the interquintile range at baseline was 59 mmHg (Table 2). After 2 years, the range in MHHP had reduced to 35 mmHg (that is, 75% of the baseline range), and in CHS to 37 mmHg (that is, 56% of the baseline range). Thus, although the baseline levels had been very different, the "usual" levels after 2 years were more similar. After a further 4 years, the range in MHHP was 32 mmHg (70% of the baseline range), compared with 28 mmHg in CHS (48% of the baseline range). The Norwegian Counties study had the narrowest range at baseline (only 41 mmHg), but still the group means converged with increasing follow-up. Thus, by the 5-year re-survey point, the difference between the means of the top and bottom groups was 31 mmHg (which is similar to MHHP and CHS after 6 years).

Two other studies warranting discussion are Speedwell and the UK Heart Disease Prevention Project (UKHDPP). The pattern of the convergence of the group means was noticeably different in the Speedwell study due to a dramatic increase in the overall mean by year 3, and subsequent fall over future visits (140 mmHg at baseline; 151 mmHg at year 3; 145 mmHg at year 7; and 143 mmHg at year 9: Figure 3). The UKHDPP re-surveyed only a small sample (under 500 at years 1 and 3) of their cohort every year for 4 years, then re-surveyed all survivors after 5 years. Consequently, the results from years 1 to 4 are relatively uninformative, being dominated by the play of chance.

Despite these few anomalies, there were no striking differences between studies or between regions.

2.2 Diastolic Blood Pressure

Many of the observations about DBP were similar to those for SBP. The serial changes in DBP within groups (defined by baseline measurements) were less uniform than those observed for SBP, and appeared generally more extreme. In Seven Cities: China, there was no discernible difference in the means of groups III and IV by 3 years - that is, the "exposure" to "risk" appeared to be the same in both groups, despite having appeared different at baseline (means of 80 mmHg and 85 mmHg).

2.3 Total cholesterol

Only 22 studies were able to provide re-survey data on total cholesterol, and even in these studies there were fewer participants with cholesterol data than there were with blood pressure data.

The differences between the studies in the overall means of cholesterol were more marked than any differences in blood pressure. In particular, all of the East Asian cohorts had lower mean cholesterol levels than any of the studies from Europe, North America or Australia. In Finrisk and Copenhagen, the overall means were particularly high (over 6 mmol/litre), which is reflected in the means of the top groups being around 8 mmol/litre in both cohorts, while even the means of the bottom groups were over 4.5 mmol/litre. In East Asia, because the mean cholesterol levels were all low, the group means were also low, so that the means of group II in Finrisk and Copenhagen were greater than the means of group IV in all of the East Asian cohorts. Despite these differences in absolute values, the range between top and bottom became narrower between baseline and re-survey in every cohort. However, although the regression dilution bias was an issue for all these studies, the implications for combining data from these cohorts, and then correcting for this bias, are considerable. Suppose one chose to group individuals by absolute levels of cholesterol, and one of these groups contained all individuals with a *baseline* cholesterol measurement of 4.5 – 5.4 mmol/litre. Then, of the individuals allocated to this group in Finrisk and Copenhagen, some would have usual values of 4.5 - 5.4 mmol/l, a few may even have usual values below 4.5 mmol/l, but a disproportionate number would have usual values of over 5.5 mmol/l. Conversely, of the individuals

allocated to this group from East Asian studies, some would have usual values of 4.5 - 5.4 mmol/l, a few may even have usual values above 5.5 mmol/l, but a disproportionate number would have usual values below 4.5 mmol/l (see Epidemiology Gremlin, Introduction: Figure 2).

Two studies in which the convergence was somewhat irregular were Shanghai and Caerphilly. The Shanghai cohort consisted of workers from two separate factories in Shanghai. Workers from one factory were re-surveyed after 3 years and workers from the other after 5 years. Thus, the participants contributing to the 3-year values were not the same as those contributing to the 5-year values. In Caerphilly, the means converged between years 0 and 5, but between years 5 and 10 the mean of each group rose (including the top group) and between years 10 and 15 the mean of each group fell (including the bottom group). This was due to the overall mean cholesterol measurements fluctuating between visits. However, at every re-survey the difference between the top and bottom group was less than at baseline.

2.4 Comparison with Framingham "multiple pairs"

The results from studies involved in the PSC were consistent with those observed by Clarke et al. in their analyses of the Framingham study and the Whitehall study (Clarke et al., [in press]) (Figure 8). The analysis of individuals in the Framingham study showed some fluctuations over the 30-year follow-up period (Figures 5-7) similar to those seen, for example, in Caerphilly and Speedwell. However, the use of "multiple pairs" largely smoothed out these fluctuations. The Framingham study was unique in re-surveying all the survivors of the cohort every two years, thus allowing

for the calculation of the multiple pairs. No other cohort in the PSC provided comparable data - either re-surveys were not conducted at regular intervals (e.g. 5,10, 14 in Caerphilly), or not all participants could provide data from every re-survey (e.g. IPC-Paris and UKHDPP had re-surveys every year, but only re-surveyed a sample each time).

Although the convergence was not necessarily uniform in all cohorts, many of the irregularities could possibly be explained by temporal changes in the overall mean.

3 Regression dilution ratios: Ratio of ranges, *r*, and self-correlation, ρ

The *regression dilution ratios* (*RDR*s) provide a means of quantifying the magnitude of the convergence of the top and bottom groups while disregarding the shape or pattern of that convergence. These *RDR*s provide a useful summary of the *magnitude* of the bias.

Figure 9 shows the estimates of the *RDR* at each re-survey in every cohort. Data from all cohorts are plotted together with each re-survey represented by a square whose area is proportional to the amount of statistical information (that is, the number of participants who contributed values from both the baseline survey and that re-survey). Estimates of the *RDR* using the *ratio of ranges* are down the left-hand side of the figure, and estimates using the *self-correlation* are down the right-hand side. Because calculation of the variance of the ratios of ranges can be somewhat complex (for example, using a non-central F-distribution), confidence intervals for the self-correlation have also been used for the ratios of ranges. Although this will underestimate the uncertainty around these estimates (see Chapter

3: Data collection and statistical methods) where the data are sparse, it will make no qualitative difference to the conclusions. The top third of the figure shows these estimates for SBP, the middle shows them for DBP and the bottom third shows them for total cholesterol. Thus, viewing plots across the page allows a comparison of the two estimates of the *RDR*s for each parameter, while viewing down the page allows a comparison of the magnitude of the *RDR*s between different parameters.

There was a general trend for more extreme *RDR*s with increasing interval between measurements, whether estimated by the *ratio of ranges* or the *self-correlation.* This trend was seen for each of SBP, DBP and total cholesterol. However, there was considerable scatter on these plots. For example, after a two-year interval the *ratio of ranges* for SBP varied from 0.75 in MHHP down to 0.57 in CHS. But, the interquintile range at baseline in MHHP was much greater (46 mmHg) than in CHS (59 mmHg). So the more extreme *ratio of ranges* in MHHP after 2 years only reduced the range to 35 mmHg, which is comparable to 34 mmHg (= 0.57x59 mmHg) in CHS. Thus, although the *RDR*s were quite different after 2 years, this translated into very similar *usual* ranges.

When regression lines, of the type $y = at + b$, weighted by the amount of statistical information, were fitted to the *RDR* estimates, the following models were obtained (where t = time in years between baseline and re-measurement):

	Ratio of ranges (%)	*Self-correlation* (%)
SBP:	71 – 1.1*time	69 – 1.3*time
DBP:	63 – 1.8*time	63 – 1.6*time
Total cholesterol:	70 – 0.7*time	70 – 0.6*time

None of these models provided a particularly good fit to the data, so the standard errors of the slopes were very large (of the order of 0.5 for each slope). The fit was particularly bad for total cholesterol due to a cluster of very low estimates of the *RDR* (below 0.50) within 10 years. These low values were all from East Asian cohorts, and when the East Asian cohorts were removed, the corresponding regression equations were:

	Ratio of ranges (%)	*Self-correlation* (%)
Total cholesterol:	73– 1.0*time	74 – 1.0*time

The fit of these models was much improved, explaining 28% and 44%, respectively, of the variation in the estimated *RDR*s. Cholesterol levels in East Asia are known to be substantially different from those typically seen in Western populations, providing some justification for treating them separately. (However, this may also suggest that the East Asian cohorts should be treated separately in the outcome analyses.)

These equations imply that the regression coefficient relating risk to a *baseline* measurement of SBP is only around 65% as steep as the corrected regression coefficient relating risk to the *usual* SBP during the first decade, 55% during the second decade and 45% during the third decade. Or, by relating risk to *baseline* measurements the true relationship for SBP would be underestimated by 35%, 45% and 55% in each decade. The true relationships between DBP and risk would be underestimated by 45%, 65% and 80%, and between total cholesterol and risk by 30%, 40% and 50% (using the equations without East Asian cohorts).

3.1 Comparisons between Europe, North America & Australia and East Asia

When studies were divided into the three regions - Europe, North America & Australia and East Asia - and regression lines (of the type $y = at + b$) weighted by the amount of statistical information, fitted separately to the *ratios of ranges* from each region, the following models were obtained:

	Europe	N America & Australia	East Asia
SBP:	72 – 1.3*time	73 – 1.3*time	52 – 2.7*time
DBP:	62 – 1.6*time	67 – 2.2*time	39 – 3.5*time
Total cholesterol:	74 – 1.0*time	70 – 0.9*time	_

Data from 25 re-surveys conducted in 10 studies contributed data to the regression lines for Europe and from 24 re-surveys conducted in 9 studies for North America & Australia. This compared with only 10 re-surveys from 6 studies in East Asia. There were even less data for total cholesterol. Consequently, the regression lines for East Asia were based on too few data for the results to be reliable. However, using these regression equations to predict *RDR*s and interquintile ranges gave reassuringly compatible values in Europe, North America & Australia and All PSC for each of the risk factors studied (Table 3). After five years (i.e. the middle of the first decade), the predicted *RDR*s for SBP were 0.66 in Europe, 0.67 in North America & Australia, and 0.65 in All PSC. The corresponding values for DBP were 0.54, 0.56 and 0.54, and for total cholesterol were 0.69, 0.66 and 0.68.

Thus, these analyses suggest that in all regions the use of *baseline* measurements would underestimate the true risk relationships by about one-third in the first decade and by about one-half in the second decade. This means that appropriate correction

for regression dilution increases the strength of the relationship of risk to disease by 50% in the first decade of follow-up and by 100% in the second decade. For DBP, the underestimation was more extreme: about 45% in the first decade and about 65% in the second decade

3.2 Comparison with Framingham "multiple pairs"

Figure 10 shows estimates of the *RDR* between pairs of measurements against the interval between measurements, for participants in the Framingham study. Estimates derived non-parametrically (using *ratio of ranges*) are on the left-hand side and parametrically (using *self-correlations*) are on the right-hand side. For both blood pressure and cholesterol, the *RDR*s became progressively more extreme with increasing interval between measurements. The weighted regression lines through these data are given below:

	Ratio of ranges (%)	*Self-correlation (%)*
SBP:	80 – 1.4*time	77 – 1.5*time
DBP:	73 – 1.7*time	70 – 1.8*time
Total cholesterol:	81 – 1.5*time	80 – 1.4*time

These models all provided particularly good fits to the data – explaining 99% of the variation in the estimates of *RDR* for both SBP and DBP, and 95% of the variation in the estimates of *RDR* for total cholesterol. They suggested that, in an analysis of the Framingham study relating risk to a *baseline* measurement of SBP, the regression coefficient would underestimate the true relationship by 30% during the first decade, 40% during the second decade and 55% during the third decade. The corresponding

underestimation for DBP would be 40%, 55% and 75%, and for total cholesterol it would be 30%, 45% and 60%.

If baseline measurements are more erratic than repeat measurements in experimental subjects (as both patients and health professionals become more familiar with the procedures) and thus more prone to the effects of regression dilution, use of multiple pairs will underestimate the real *RDR* needed to convert baseline measurements to usual values. This is particularly so for short intervals between measurements when re-surveys from the second and even third decades were contributing to the "initial" values for each pair. For example, for the 2-year interval, not only was the baseline (year 0) survey contributing to the "initial" values, but also the 2-year re-survey, the 4-year re-survey, and so on up to the 28-year re-survey (acting as the "initial" value for the 30-year re-survey). Furthermore, in the Framingham study, health professionals were aware of previous measurements at each survey. Thus in the Framingham study, the estimated *RDR*s for short intervals between measurements were generally less extreme than those observed in the remaining studies in the PSC (including those from the USA). However, for total cholesterol they became slightly more extreme over prolonged follow-up.

4 Implications for the analysis of observational studies and for an overview of such studies

The regression dilution bias needs to be corrected for in the analysis of all these studies. Fortunately, uncertainties about absolute values, distributions and trends have little effect on the non-parametric assessment of the regression dilution bias

(indeed, they are part of the broader definition of regression dilution). Parametric methods, however, rely on certain assumptions being met, such as the means and variances being constant over time, which these data suggest are not appropriate - thus arguing for the use of robust, non-parametric methods (such as the ratio of ranges).

There were differences in the size of the *RDR*s at any given time between the studies, but the ratios were always substantially less than unity. Indeed, any differences between studies in the magnitude of the estimated *RDR* were always less than the differences between each *RDR* and unity. Thus, any analysis making no attempt to account for the regression dilution bias would grossly underestimate the importance of these risk factors, and would **always be more biased** than an analysis using an *RDR* estimated, in any of the fore-going ways, from whatever study. Furthermore, the general trend within cohorts for a decrease in the *RDR* with increasing follow-up confirmed the findings from the Framingham and Whitehall studies (Clarke et al. [*in press*]) of a need for period-specific corrections. The average time to death in the Prospective Studies Collaboration was 11 years, ranging from 8 years in those aged 40-59 when dying to 14 years in those aged 70-79 when dying. Therefore, period-specific correction factors were needed for both the first and second decades of follow-up. That is, a "*generic period-specific correction procedure*" was required which could be adapted for any risk factor in different periods of follow-up, and which accounted for the risk factor's distribution.

Figures and tables

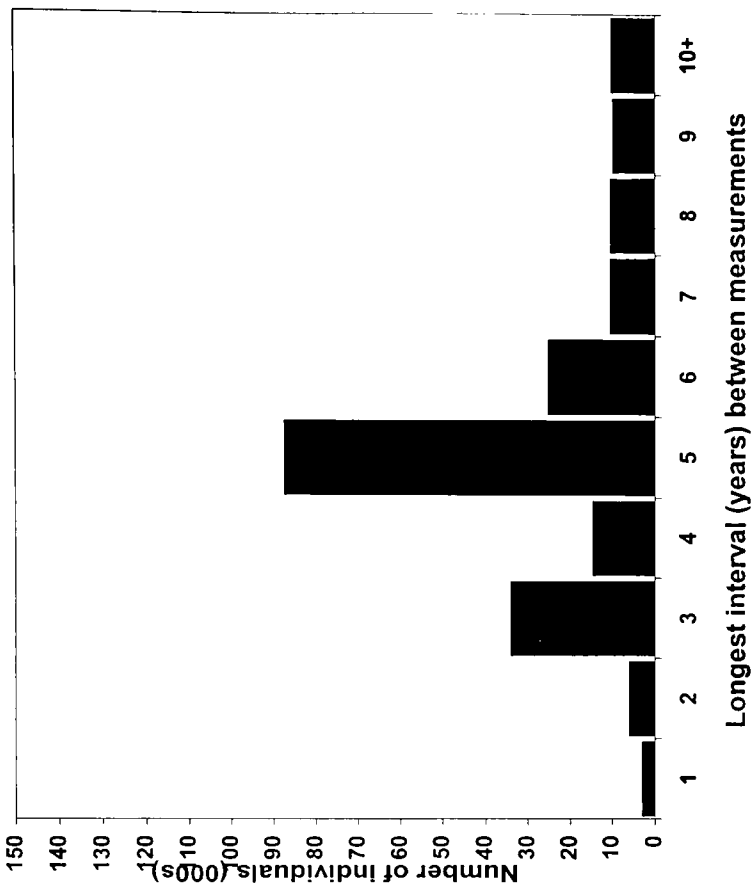


Figure 1: Frequency distribution of the maximum number of re-surveys provided by each individual (Number screened = 1 million)

Figure 2: Frequency distribution of the longest interval between measurements in each individual (Number screened = 1 million)

Figure 3: Mean values of systolic and diastolic BP over time in each cohort
(a) Europe

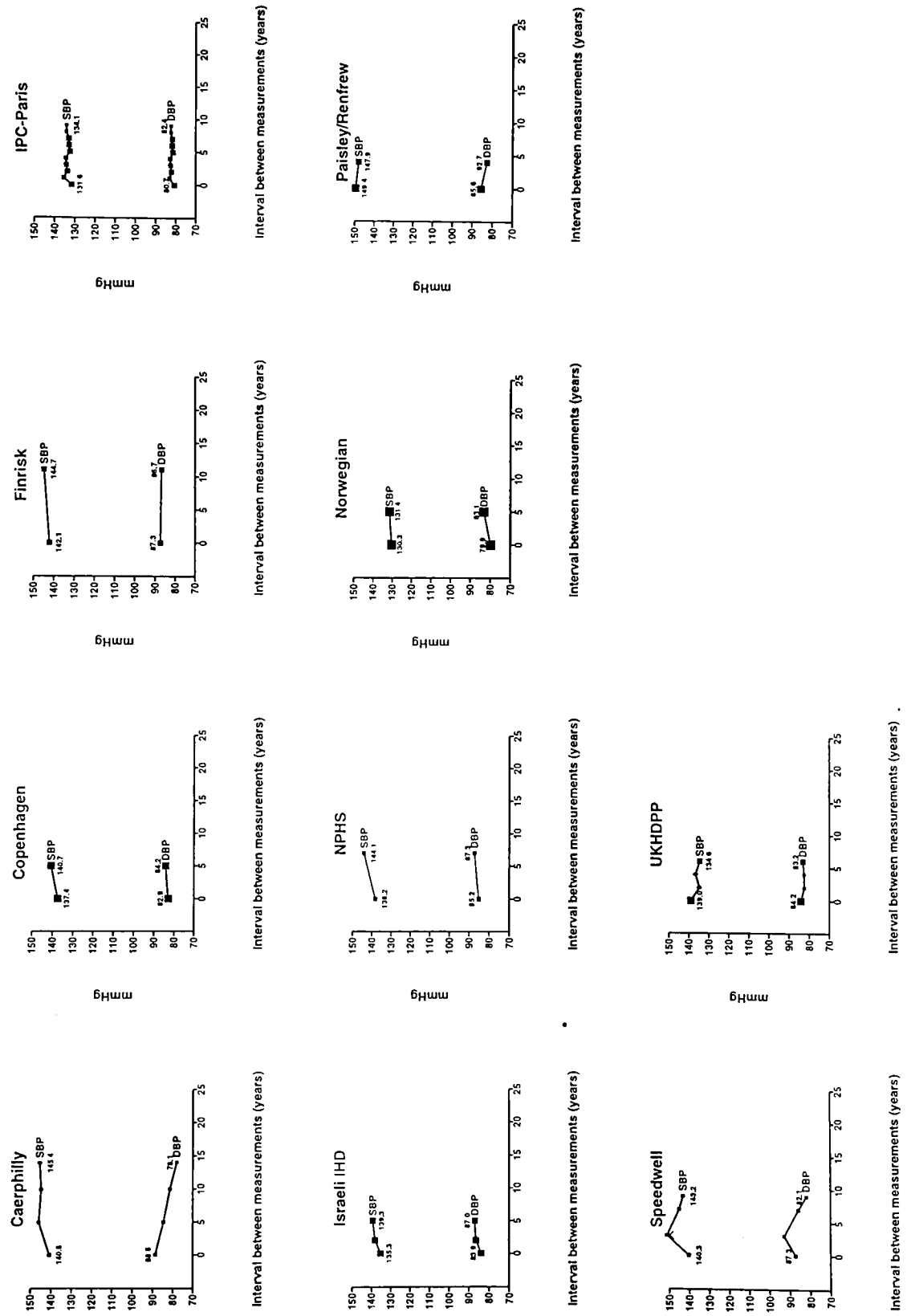


Figure 3: Mean values of systolic and diastolic BP over time in each cohort
(b) North America and Australia

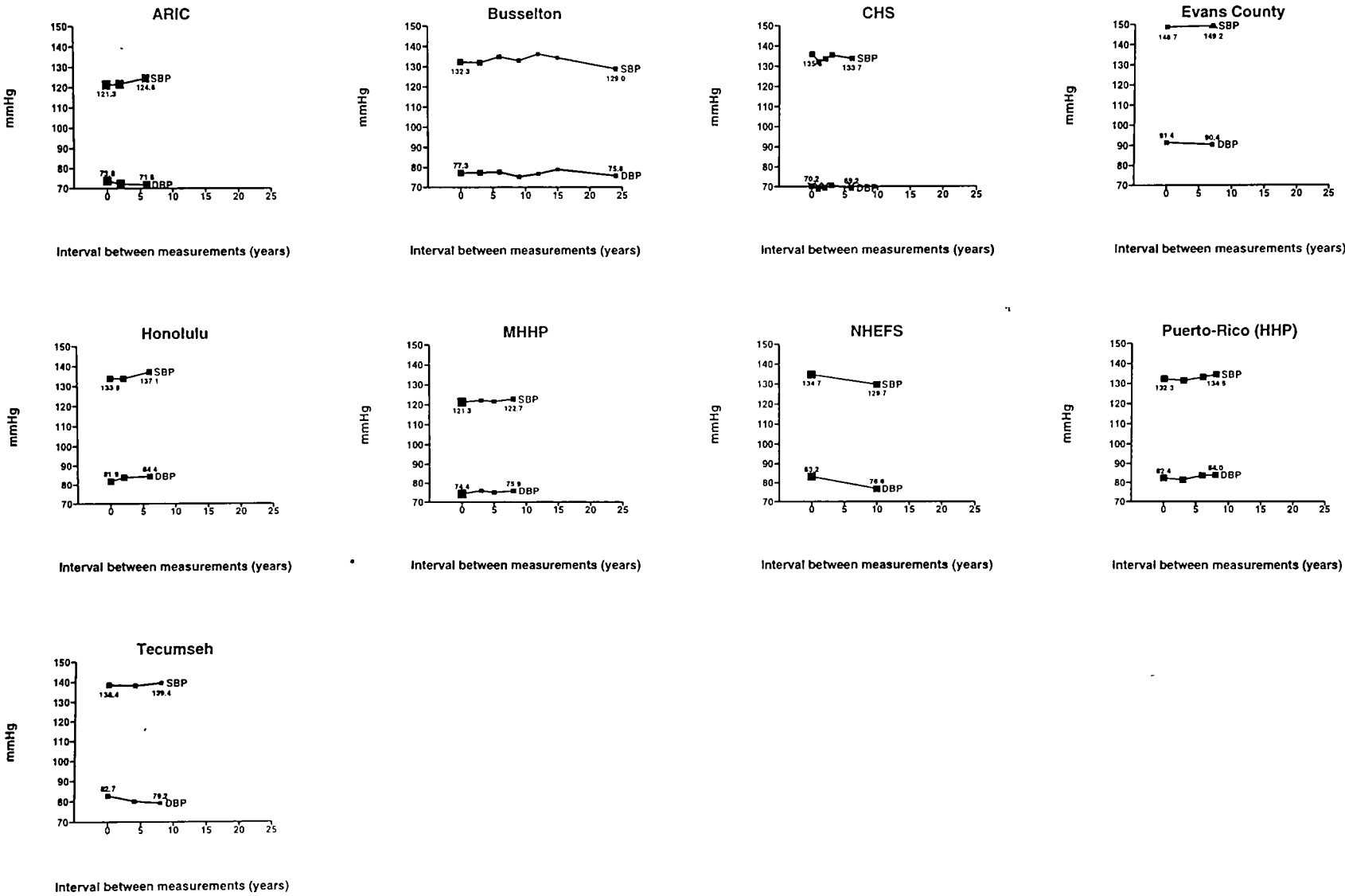


Figure 3: Mean values of systolic and diastolic BP over time within each cohort
(c) East Asia

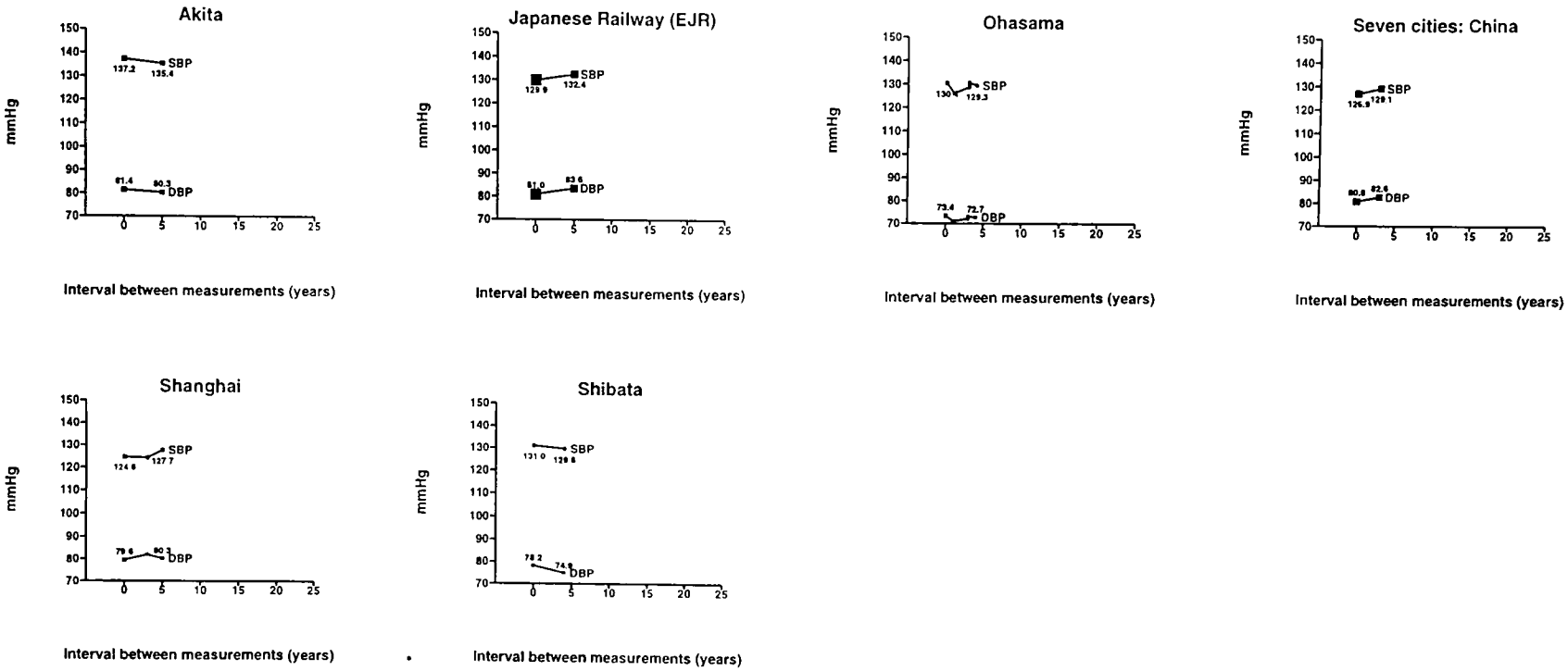


Figure 4: Mean values of total cholesterol over time within each cohort
(a) Europe

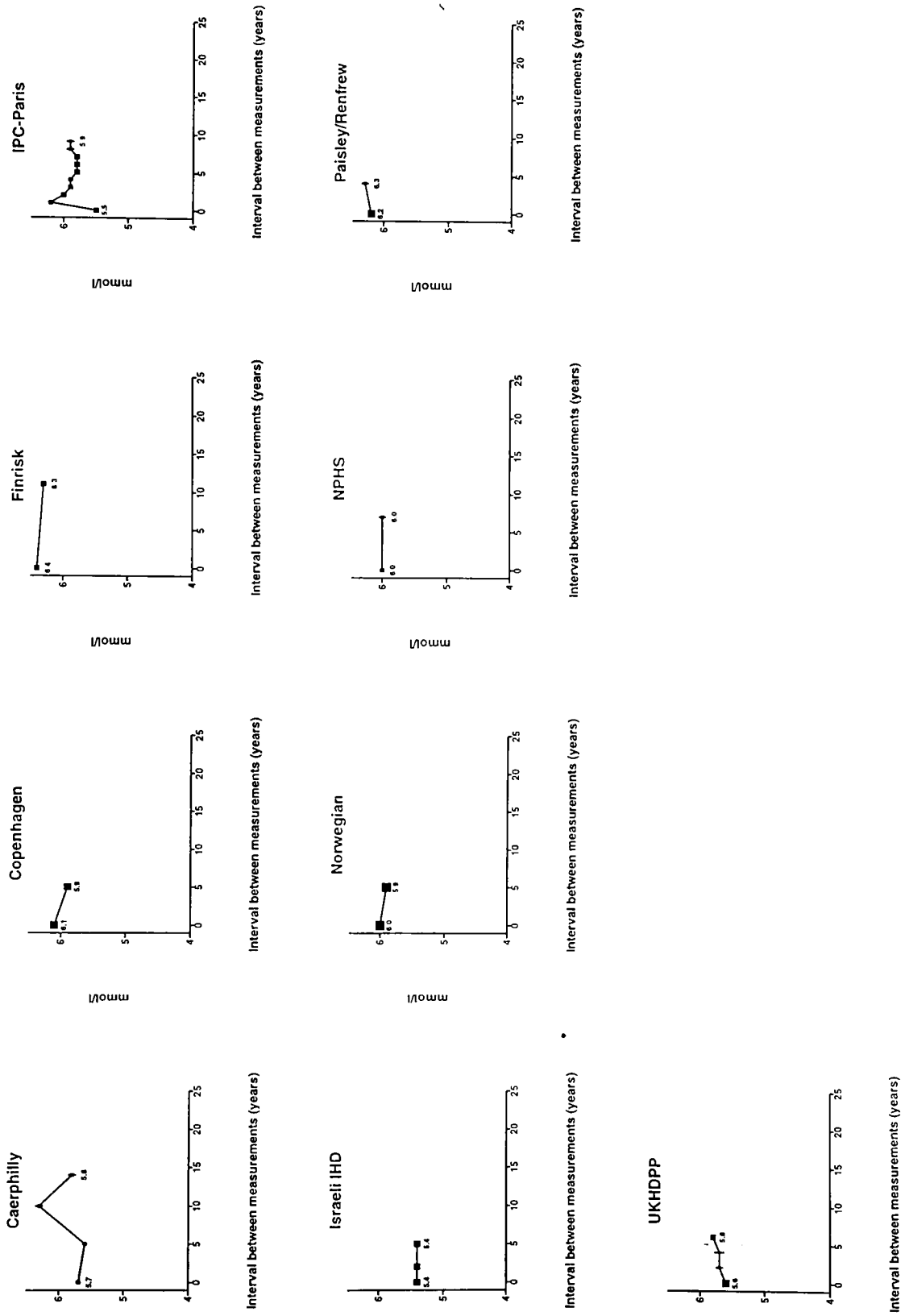


Figure 4: Mean values of total cholesterol over time within each cohort
(b) North America, Australia and East Asia

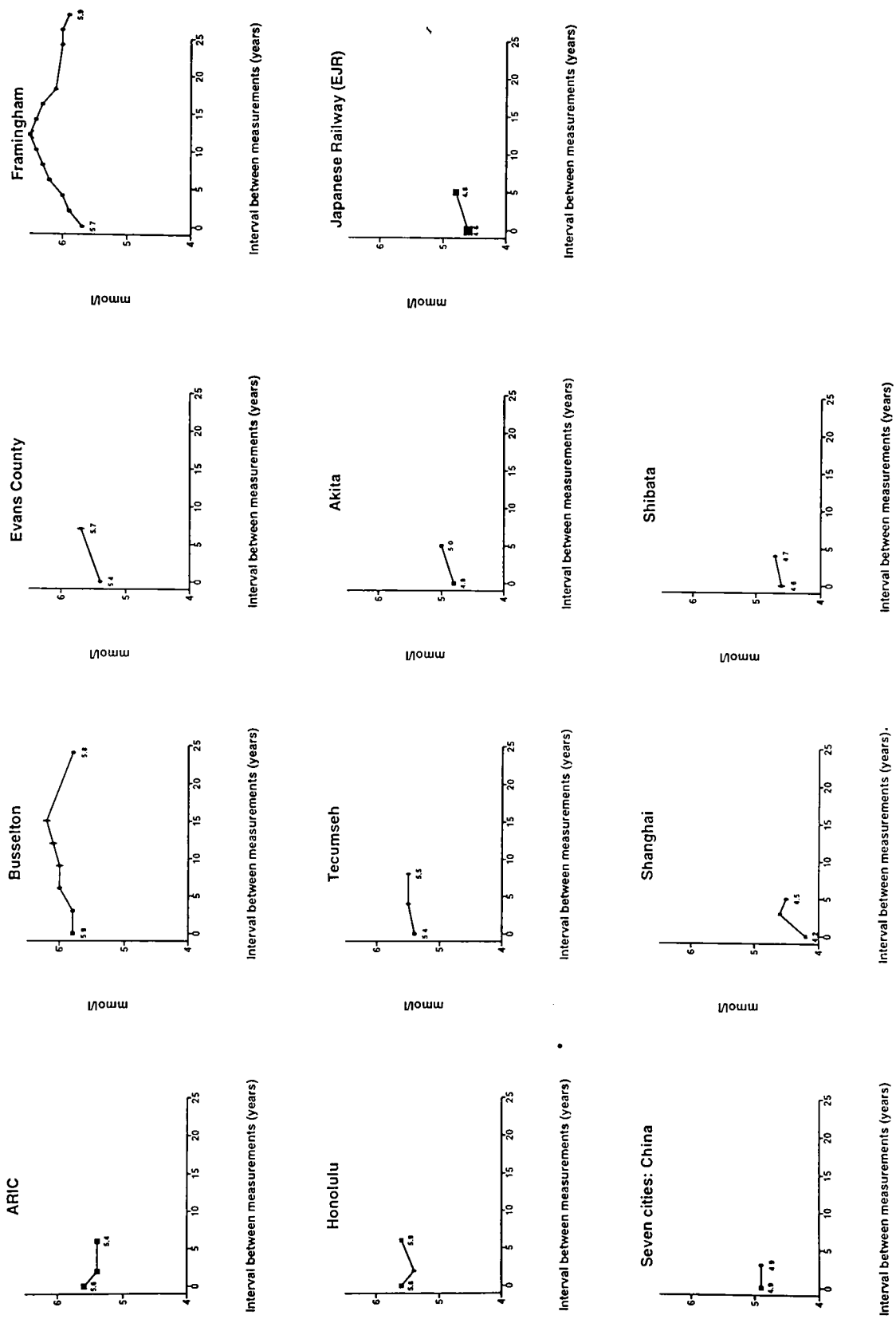


Figure 5: Serial changes in systolic BP within groups defined by baseline measurements (a) Europe

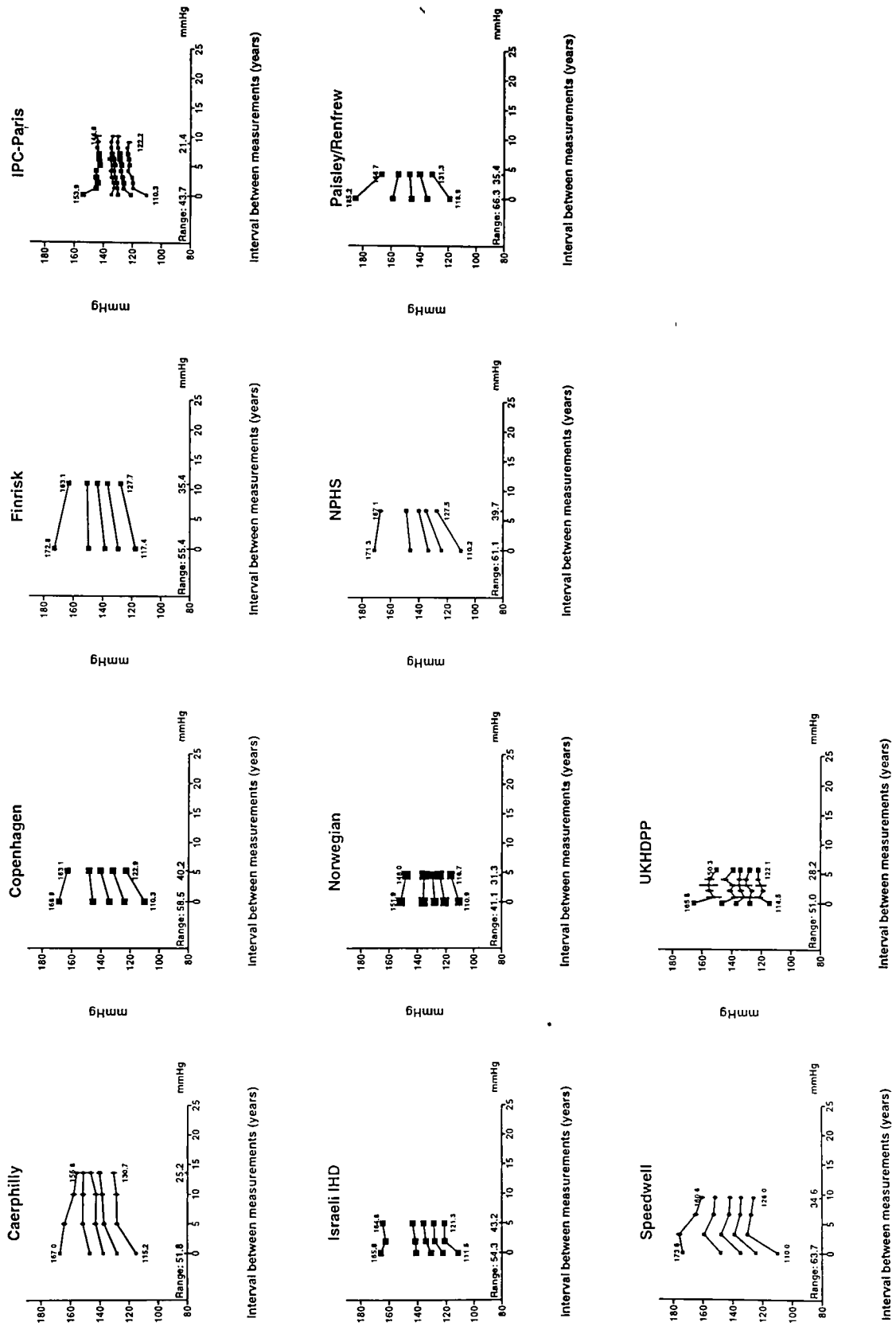


Figure 5: Serial changes in systolic BP within groups defined by baseline measurements
(b) North America & Australia

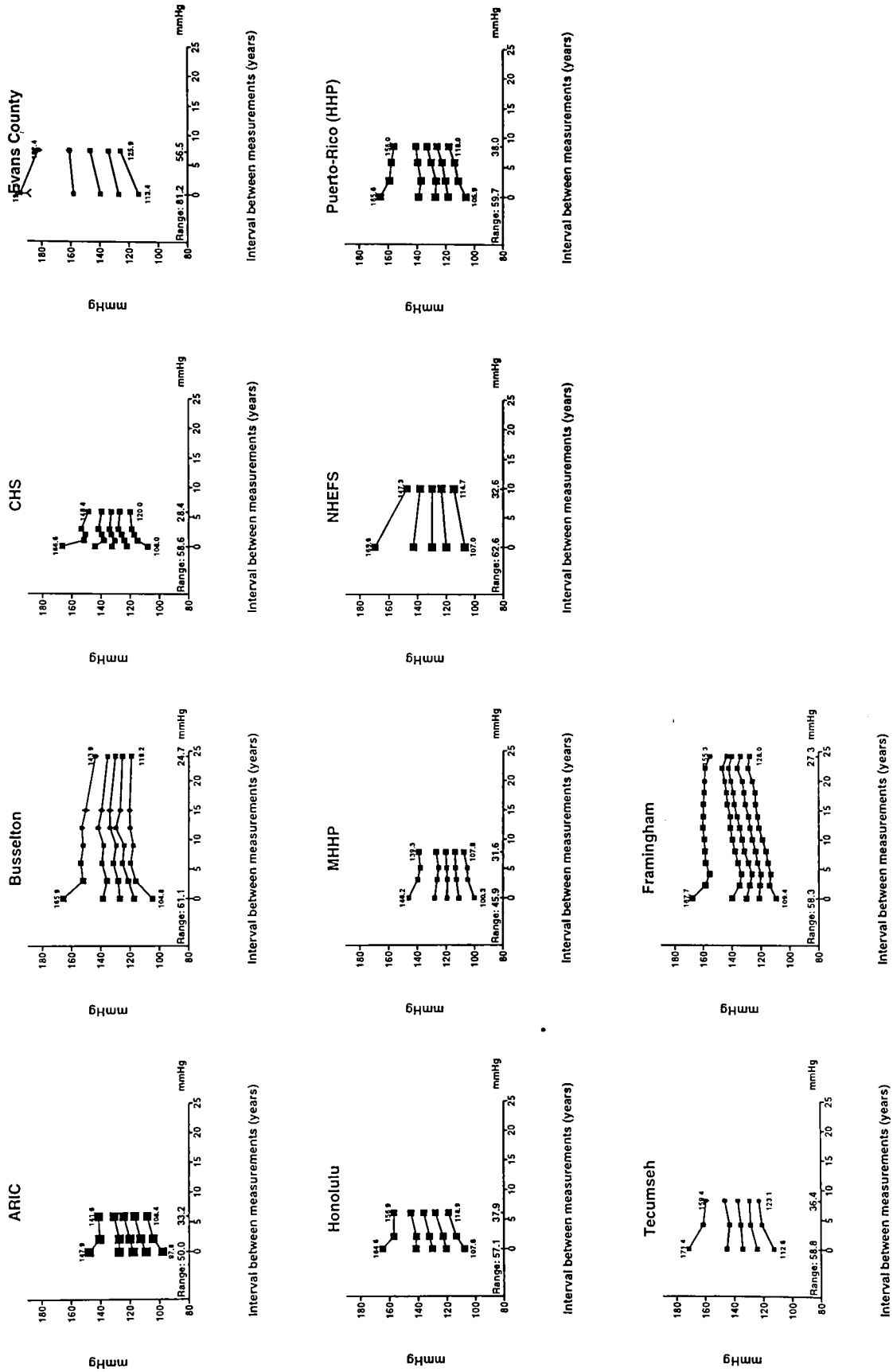


Figure 5: Serial changes in systolic BP within groups defined by baseline measurements
(c) South East Asia

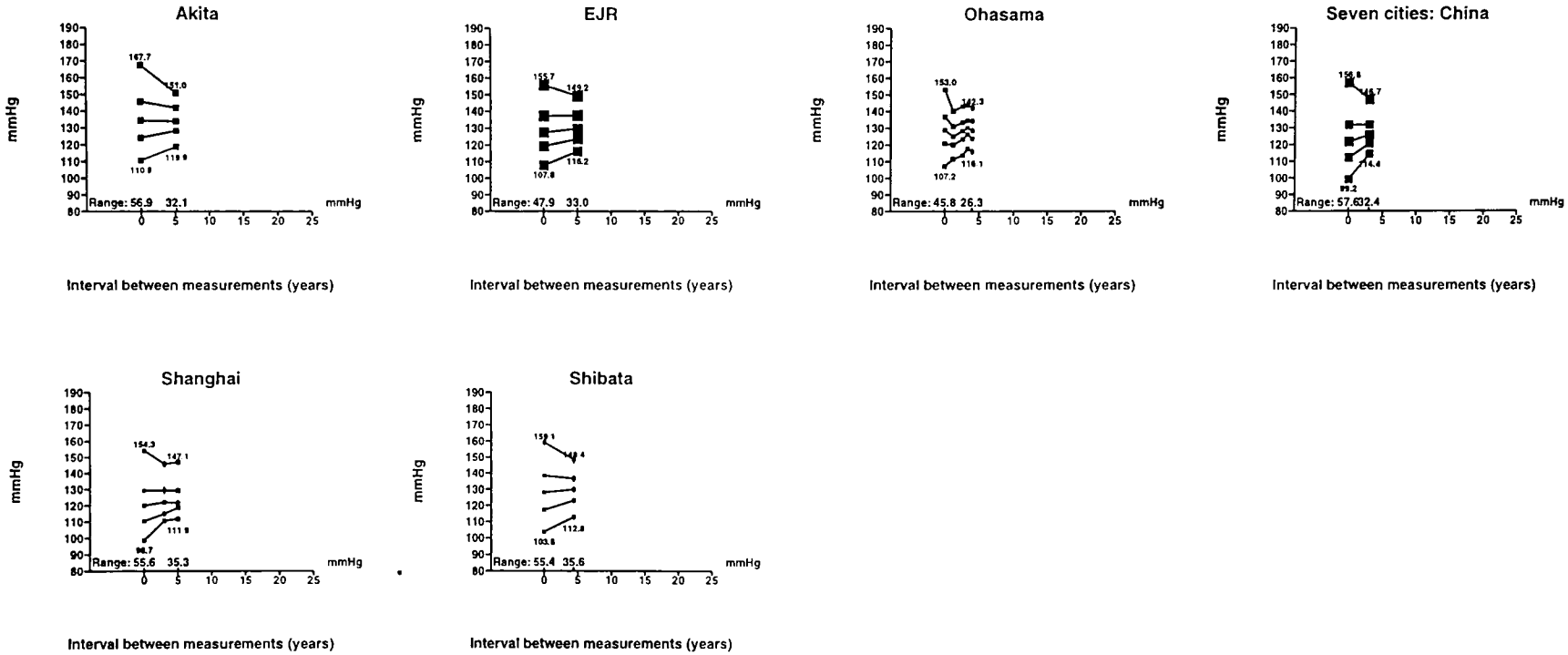


Figure 6: Serial changes in diastolic BP within groups defined by baseline measurements
(a) Europe

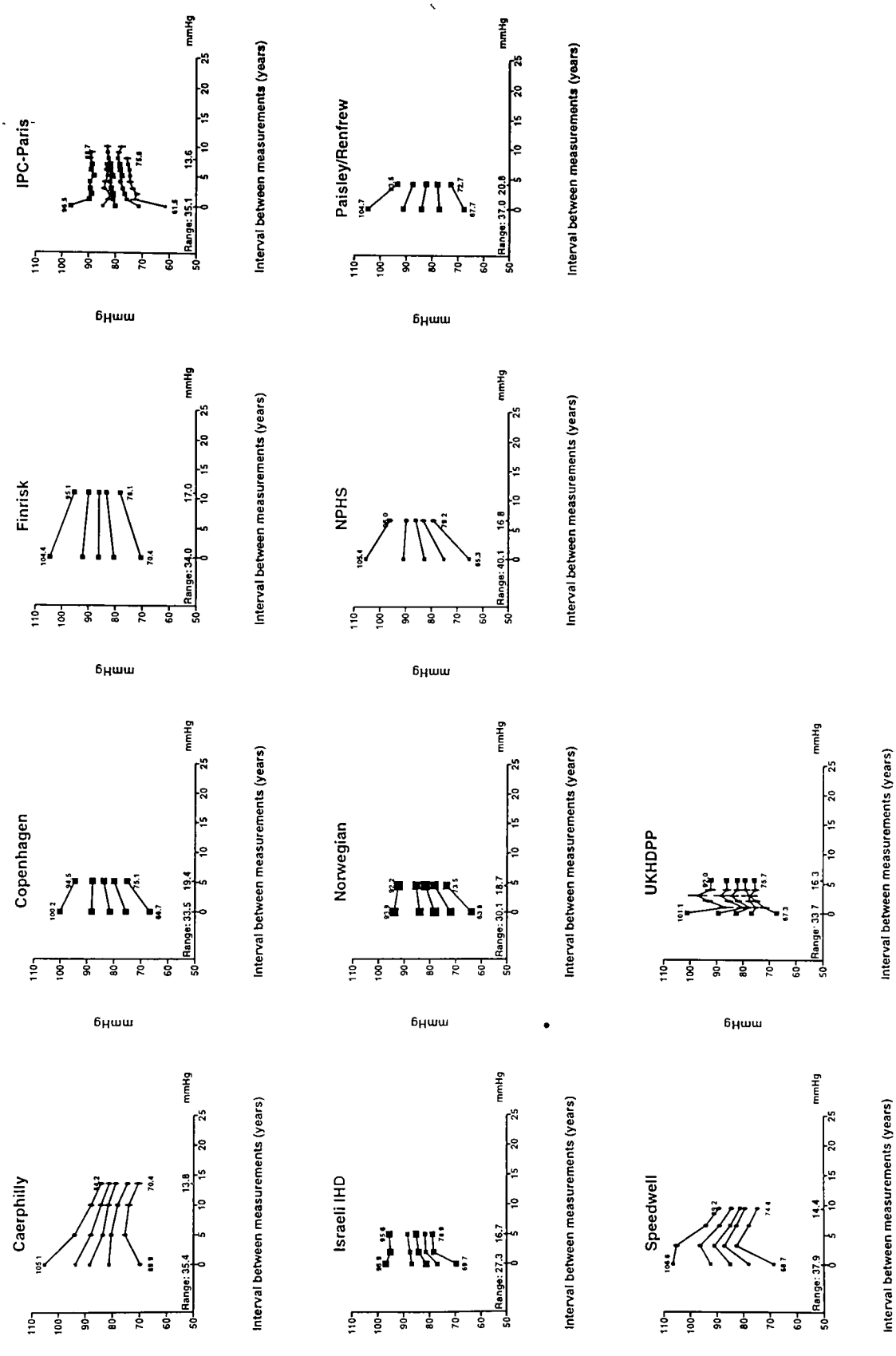


Figure 6: Serial changes in diastolic BP within groups defined by baseline measurements (b) North America & Australia

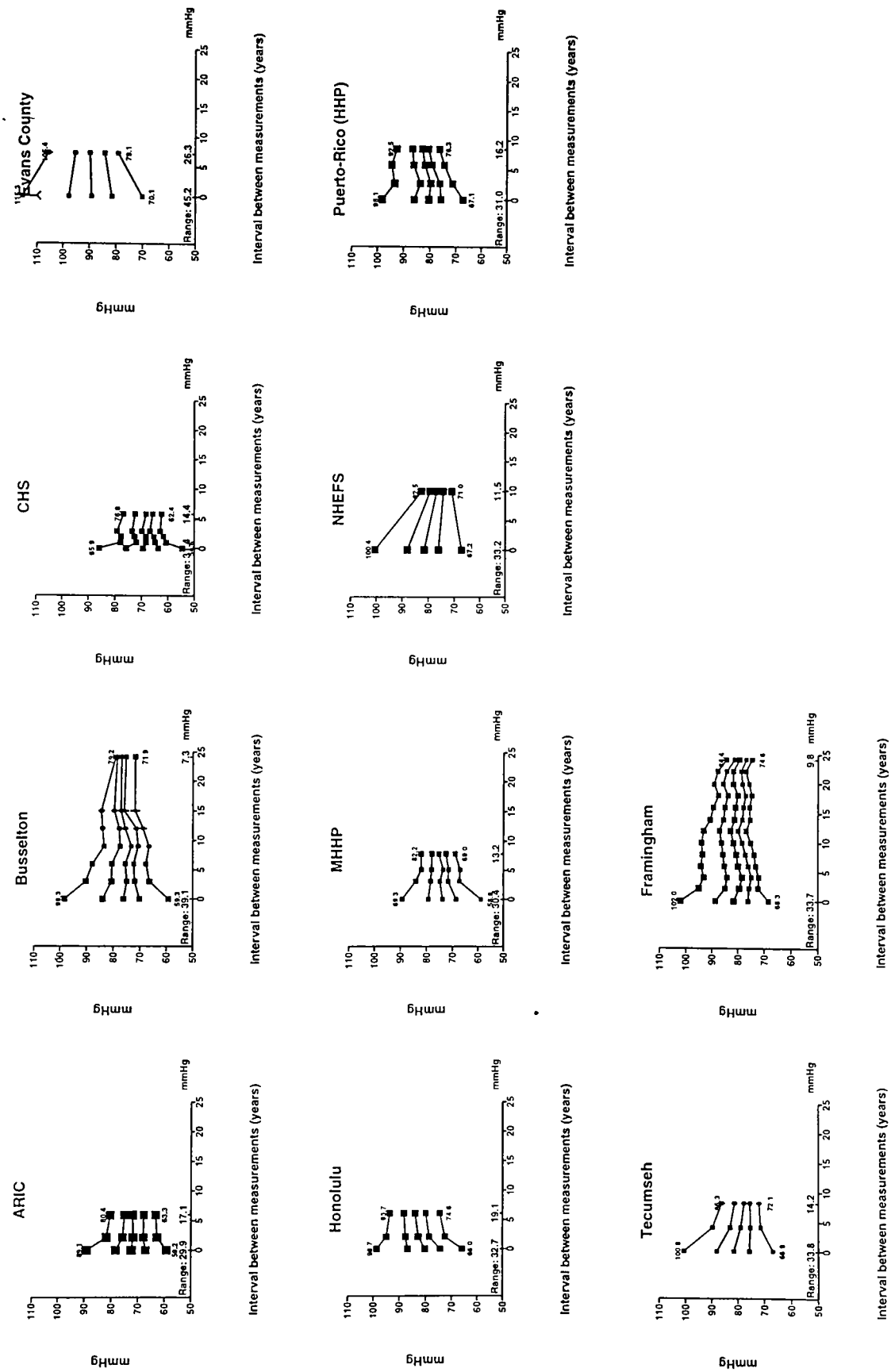


Figure 6: Serial changes in diastolic BP within groups defined by baseline measurements. (c) South East Asia

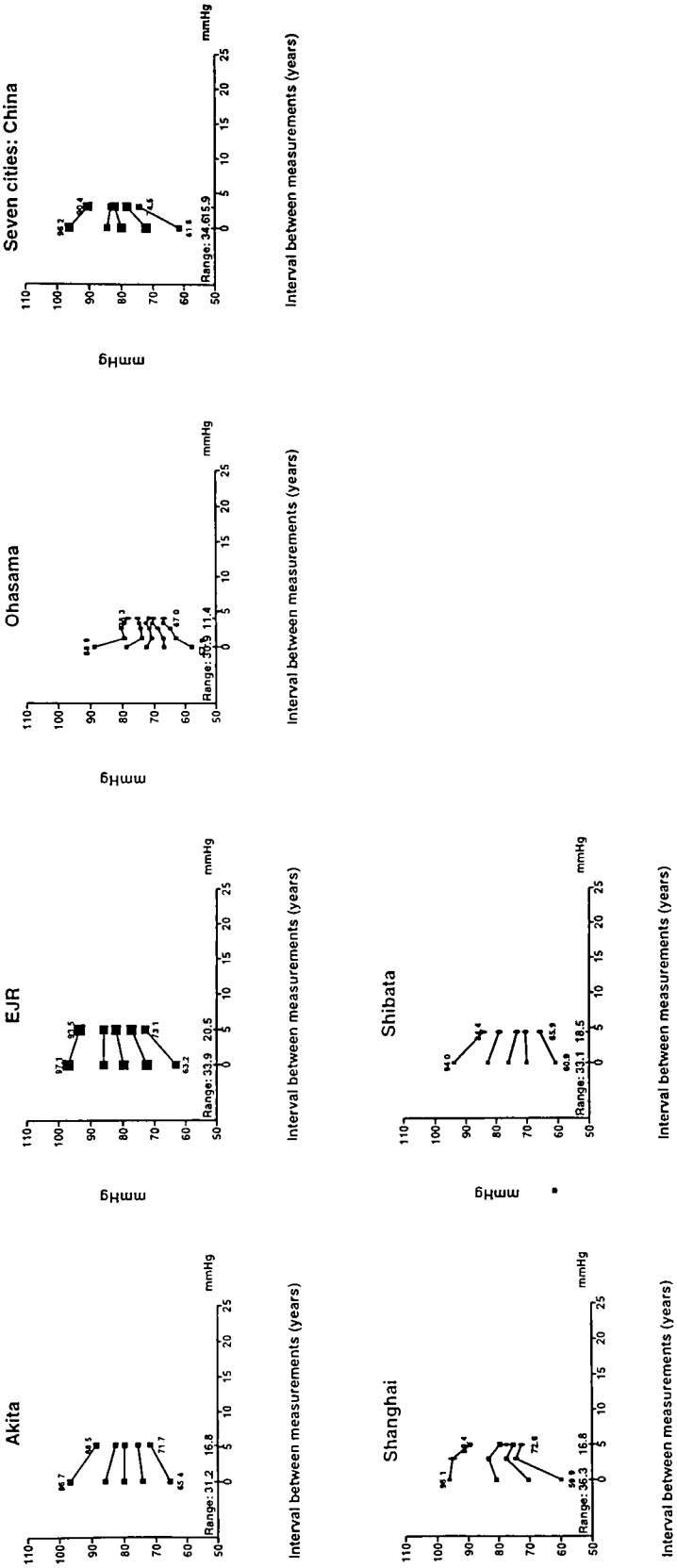


Figure 7: Serial changes in total cholesterol within groups defined by baseline measurements
(a) Europe

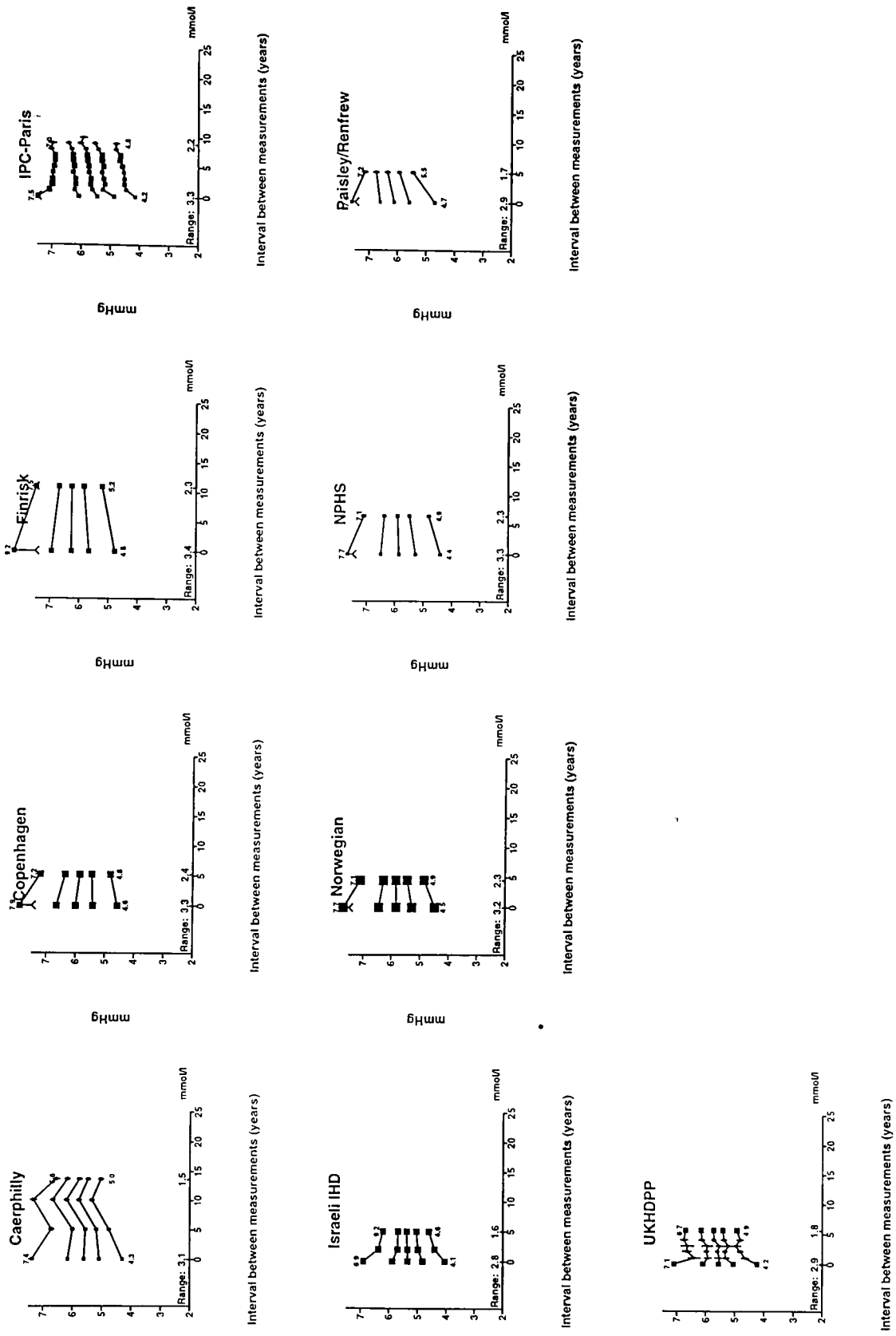


Figure 7: Serial changes in total cholesterol within groups defined by baseline measurements (b) North America, Australia and South East Asia

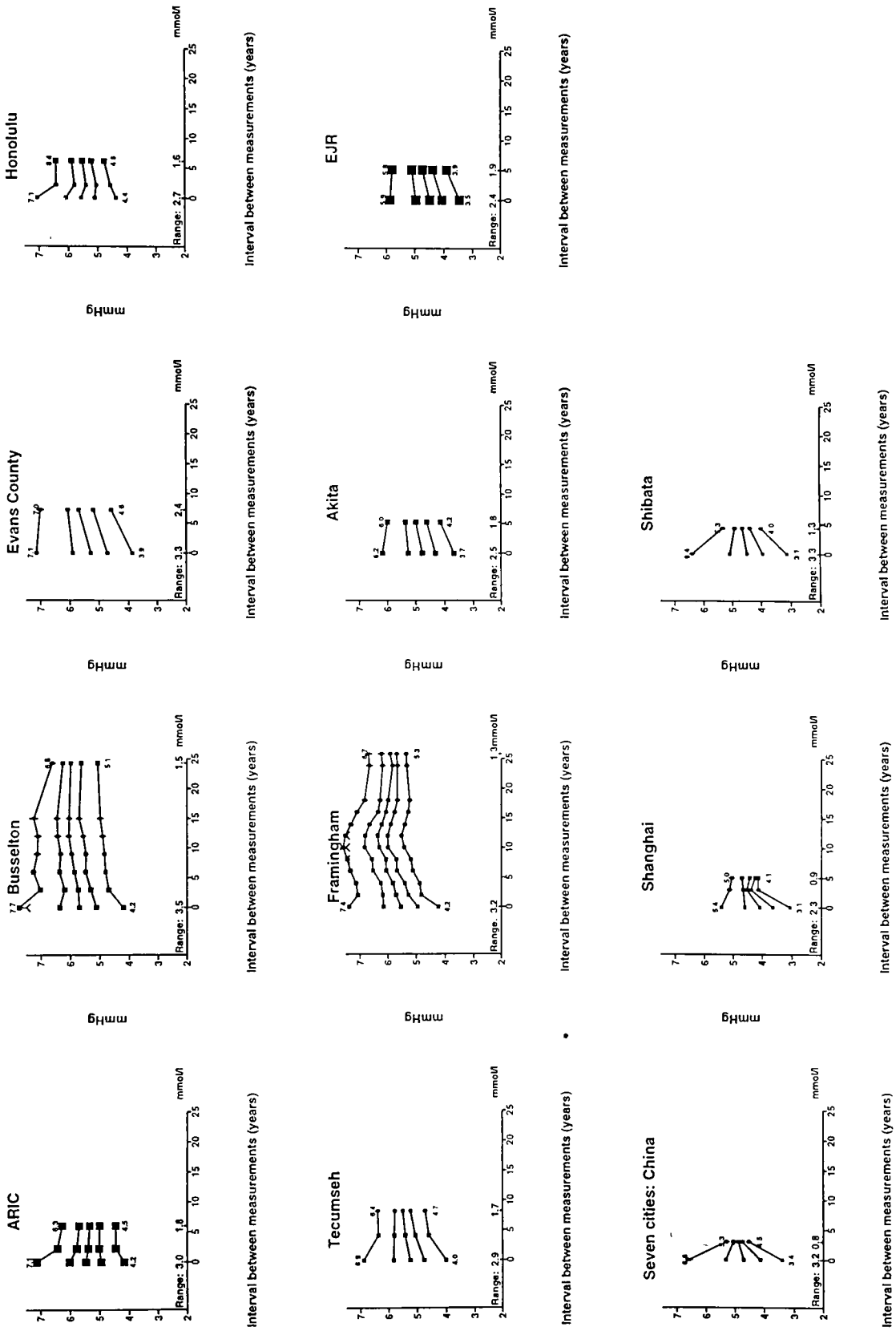


Figure 8: Serial changes in blood pressure and cholesterol values within groups defined by the initial measurements

Framingham "multiple pairs"

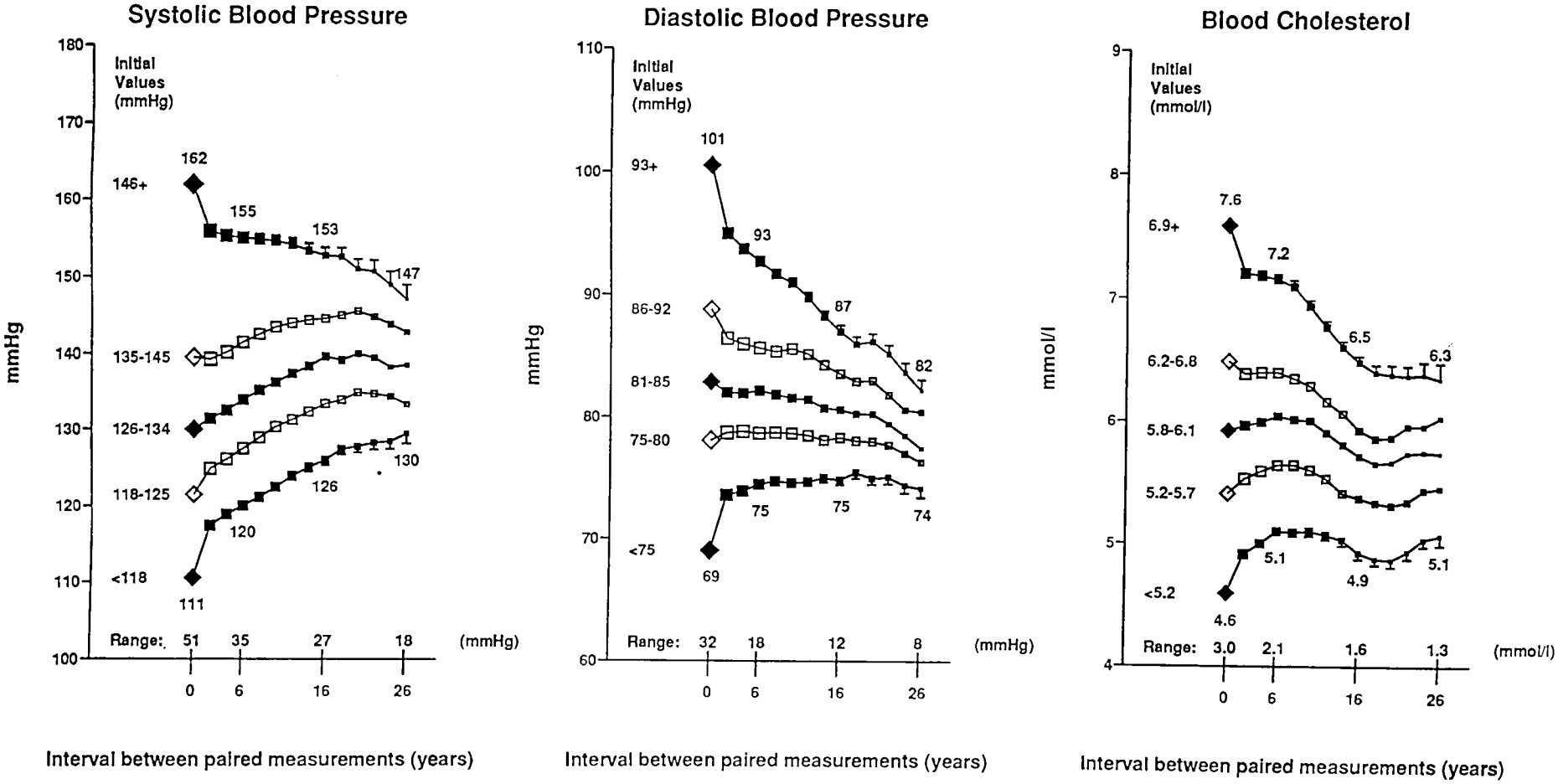


Figure 9: Serial changes in the regression dilution ratios for (a) systolic BP, (b) diastolic BP and (c) total cholesterol: ratios of Ranges and self-correlations

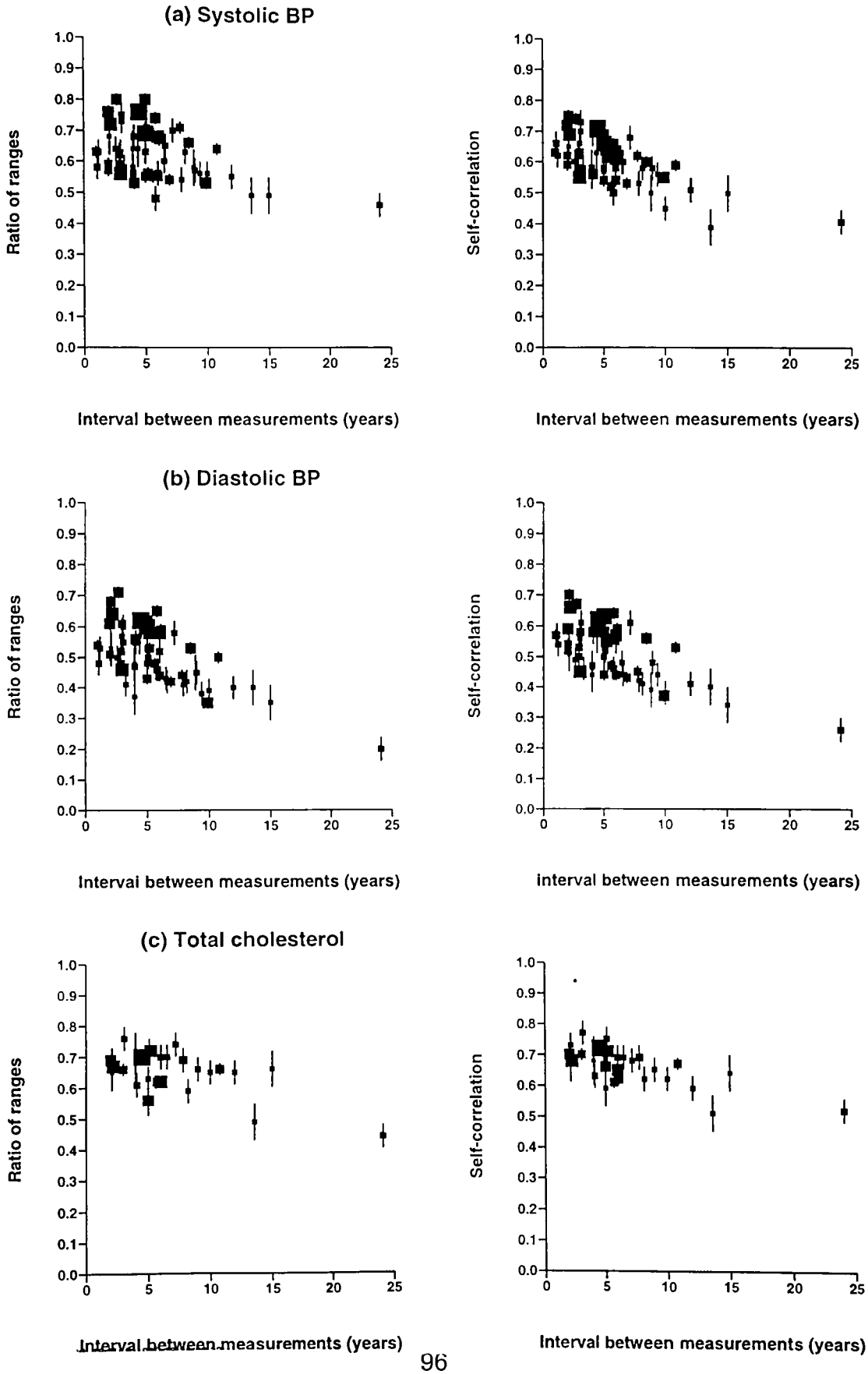


Figure 10: Serial changes in the regression dilution ratios within Framingham "multiple pairs" for systolic BP, diastolic BP and total cholesterol: ratios of ranges and self-correlations

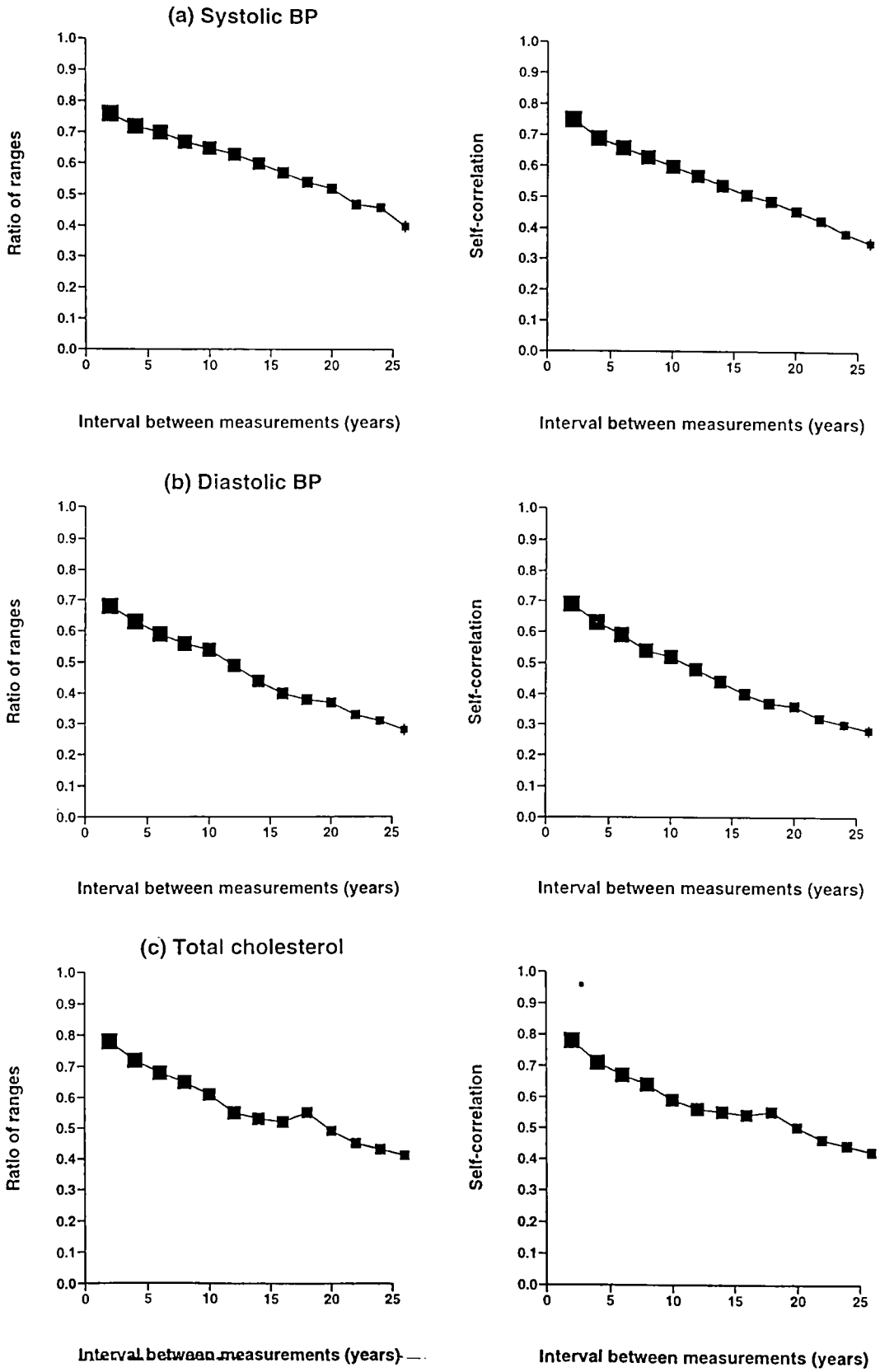


Table 1: Distribution of repeat blood pressure measurements by region and by study

Region and study name (year screening began)	Baseline population[1] (000)	No. with at least one re-survey[2] (000)	Mean time to death (years)	Repeat blood pressure and cholesterol measurements	
				Number of re-surveys[3]	Time interval of first and last re-survey from baseline (years)
Europe					
Caerphilly (1979)	2.5	2.0	7	3	5,14
Copenhagen (1976)	14.0	14.0	10	1	5
Finrisk (1974)	38.8	5.7	10	2	11, 15
IPC-Paris (1978)	221.8	28.2	8	10	1, 10
Israeli IHD (1963)	10.0	10.0	14	2	2,5
Norwegian (1974)	48.6	24.8	10	1	5
NPHS (1972)	3.5	2.5	11	1	7
Paisley/Renfrew (1972)	15.4	8.2	9	1	4
Speedwell (1979)[4]	2.3	2.1	7	3	3, 9
UK HDPP (1971)	13.9	6.4	13	5	1,6
Sub-total	**370.8**	**103.9**	**9**		**1, 15**
(Outcome analyses[5]	721		11		)
N. America/Australia					
ARIC (1987)	15.8	14.6	4	2	3, 6
Busselton (1966)	7.3	5.8	14	6	3, 24
CHS (1989)	5.2	5.0	2	4	1, 6
Evans County (1960)	3.1	2.5	15	1	7
Honolulu (1965)	8.0	7.6	16	2	2, 6
MHHP (1980)[4]	17.9	5.5	4	3	3,8
NHEFS (1971)[4]	14.3	10.0	10	1	10
Puerto-Rico HHP (1965)[4]	9.8	9.2	7	3	3, 8
Tecumseh (1959)	4.2	3.0	14	2	4, 8
Sub-total	**85.6**	**63.2**	**8**		**1, 24**
(Outcome analyses[5]	206	-	12		)
East Asia					
Akita (1975)	8.7	6.2	9	1	5
Japanese Railway (1973)	55.0	22.6	5	1	5
Ohasama (1990)[4]	3.1	2.6	2	4	1, 4
Seven cities: China (1986)	18.9	14.7	2	1	3
Shanghai (1972)	9.3	4.4	11	2	3, 5
Shibata (1977)	2.4	1.4	8	1	4
Sub-total	**97.4**	**51.9**	**5**		**1, 5**
(Outcome analyses[5]	90		8	-	)
Framingham (multiple pairs)	53.2	53.2		15	2,30

1 Numbers given are for individuals with at least blood pressure measurements at baseline. Numbers with cholesterol are similar (but sometimes slightly less).

2 Numbers given are for individuals with data from at least one re-survey of blood pressure. Numbers with cholesterol are similar (but sometimes slightly less).

3 Not all participants had re-measurements at each re-survey.

4 No blood cholesterol values available at re-survey.

5 Only individuals with data on both blood pressure and cholesterol at baseline are included in the outcome analyses.

Table 2: Mean values and interquintile range (absolute difference between the top and bottom quintiles) for systolic BP, diastolic BP and total cholesterol at baseline, in studies with repeat measurements

Region and study name (year screening began)	Age[1] (years)	Systolic BP (mmHg)		Diastolic BP (mmHg)		Total Cholesterol (mmol/ litre)	
	Mean	Mean	Interquintile range	Mean	Interquintile range	Mean	Interquintile range
Europe							
Caerphilly (1979)	52	141	52	89	35	5.7	3.1
Copenhagen (1976)	53	137	59	83	33	6.1	3.3
Finrisk (1974)	43	142	55	87	34	6.4	3.4
IPC-Paris (1978)	41	132	44	81	35	5.5	3.3
Israeli IHD (1963)	49	135	54	84	27	5.4	2.8
Norwegian (1974)	42	130	41	80	30	6.0	3.2
NPHS (1972)	46	138	61	85	40	6.0	3.3
Paisley/Renfrew (1972)	54	149	66	86	36	6.2	2.9
Speedwell (1979)[2]	54	140	64	87	38		
UK HDPP (1971)	51	139	51	84	34	5.6	2.9
Sub-total	**46**	**136**	**50**	**83**	**33**	**5.9**	**3.2**
(Outcome analyses[3]	44	137	54	83	32	5.9	3.2
N. America/Australia							
ARIC (1987)	54	121	50	74	30	5.6	3.0
Busselton (1966)	44	132	61	77	39	5.8	3.5
CHS (1989)[2]	73	136	59	70	31	-	
Evans County (1960)	47	149	81	91	45	5.4	3.3
Honolulu (1965)	54	134	57	82	33	5.6	2.7
MHHP (1980)[2]	45	121	46	74	30		
NHEFS (1971)[2]	49	135	63	83	33		
Puerto-Rico HHP (1965)[2]	54	132	60	82	31	-	
Tecumseh (1959)	47	138	59	83	34	5.4	2.9
Sub-total	**51**	**125**	**54**	**75**	**30**	**5.2**	**2.8**
(Outcome analyses[3]	52	128	51	79	29	5.5	3.1
East Asia							
Akita (1975)	54	137	57	81	31	4.8	2.5
Japanese Railway (1973)	43	130	48	81	34	4.6	2.4
Ohasama (1990)[2]	56	130	46	73	31	-	-
Seven cities: China (1986)	57	127	58	81	35	4.9	3.2
Shanghai (1972)	48	125	56	80	36	4.2	2.3
Shibata (1977)	57	131	54	78	33	4.6	3.3
Sub-total	**50**	**131**	**53**	**81**	**35**	**4.7**	**2.6**
(Outcome analyses[3]	47	135	53	82	33	4.6	2.5
Framingham (multiple pairs)	44	135	51	85	32	5.7	3.0

1 Age at initial measurement of blood pressure and cholesterol.
2 No blood cholesterol values available at re-survey.
3 Only individuals with data on both blood pressure and cholesterol at baseline are included in the outcome analyses.

Table 3: Predicted regression dilution ratios and interquintile ranges within each region for studies with re-survey data: systolic BP, diastolic BP and total cholesterol

Region	Interval between re- survey and baseline measurements	Systolic BP (mmHg)		Diastolic BP (mmHg)		Total Cholesterol (mmol/ litre)	
		RDR	Interquintile range	RDR	Interquintile range	RDR	Interquintile range
Europe							
	0	1.0	50	1.0	33	1.0	3.2
	5	0.66	33	0.54	18	0.69	2.2
	10	0.59	30	0.46	15	0.64	2.0
	15	0.53	27	0.38	13	0.59	1.9
	20	0.46	23	0.30	10	0.54	1.7
N. America/Australia							
	0	1.0	54	1.0	30	1.0	2.8
	5	0.67	36	0.56	17	0.66	1.8
	10	0.61	33	0.45	14	0.61	1.7
	15	0.54	29	0.34	10	0.57	1.6
	20	0.48	26	0.23	7	0.52	1.5
East Asia							
	0	1.0	53	1.0	35	1.0	2.6
	5	0.39	21	0.22	8	*	*
	10	0.25	13	0.04	1	*	*
	15	0.12	6	0	0	*	*
	20	0	0	0	0	*	*
All PSC							
(excl. Framingham)	**0**	**1.0**	**53**	**1.0**	**31**	**1.0**	**3.1**
	5	**0.65**	**34**	**0.54**	**17**	**0.68**	**2.1**
	10	**0.60**	**32**	**0.45**	**14**	**0.63**	**2.0**
	15	**0.55**	**29**	**0.36**	**11**	**0.58**	**1.8**
	20	**0.49**	**26**	**0.27**	**8**	**0.53**	**1.6**
Framingham **							
(multiple pairs)	0	1.0	58	1.0	31	1.0	3.2
	5	0.71	41	0.61	19	0.70	2.2
	10	0.65	38	0.54	17	0.61	2.0
	15	0.58	34	0.42	13	0.52	1.7
	20	0.52	30	0.37	11	0.49	1.6

* There were only six re-surveys from 5 studies, so predicting regression dilution ratios was not possible.
** Framingham values are observed values. Years 5 and 15 are interpolated from years 4 & 6 and years 14 & 16 respectively.

Chapter 6: New techniques to account for the regression dilution bias over prolonged follow-up in individual observational studies and in overviews of such studies

Chapter summary

This chapter describes a simple method to correct for the regression dilution bias in individual observational studies with prolonged follow-up and in an overview of many such studies. It provides a reasonably reliable estimate of the regression dilution ratio (*RDR*) that can be utilised even in studies where later re-measurements are unavailable from the appropriate interval of follow-up or absent altogether.

1 Objective

The chief objective of this thesis is to establish a unique centralised collection of data on individuals from prospective observational studies worldwide; and to establish and to use appropriate statistical methods to assess the true age- and sex-specific relationships between blood pressure, cholesterol and cause-specific mortality. Appropriate techniques to correct for the time-dependent regression dilution bias in individual studies and, particularly, in an overview of studies (where cohorts vary substantially in risk factor distributions and in duration of follow-up) have not previously been available.

We have shown previously that although the regression dilution ratio (*RDR*) depends strongly on the duration of follow-up (Clarke et al., [in press]), it is largely independent of age and sex. The previous chapter showed how the regression dilution bias is relevant in all observational studies, irrespective of quality or size of study

populations. Observations from those studies confirmed the findings from Framingham and Whitehall that the magnitude of this bias generally increases with follow-up. Yet, only half the studies within the Prospective Studies Collaboration (PSC) had re-measured risk factor levels in their cohort, and often these re-measurements were carried out within only a few years of the initial survey. Consequently, there was very little information available from later periods of follow-up.

2 Statistical methods

Data on repeat measurements performed 10 years or more after the baseline survey were limited from studies in the PSC (and therefore worldwide). Consequently, estimates of the long-term *RDR*s were confined to the few studies where such later re-measurements were available. Nevertheless, the many studies with data from at least one re-survey within a few years of the baseline (that is, about half of all studies) were able to provide some estimate of the short-term *RDR* within that study. The subsequent magnitude of the *RDR* could be imputed from those studies in which measurements were available from later re-surveys. This assumed that, although the short-term effects of the regression dilution bias (or magnitude of the *RDR*) may have varied between studies, any subsequent medium or longer-term effects were likely to be similar - an assumption that was tested in those studies with re-survey data more than ten years after the baseline survey.

2.1 Standardising for different distributions of risk factor levels

Since the overall mean levels of risk factors varied substantially between different cohorts (Chapter 4, Tables 3 and 4), the quintile values used as cut-offs when defining group membership differed. Consequently, the mean values of the top and bottom groups (fifths) varied substantially between cohorts. For example, the mean systolic blood pressure (SBP) at baseline in the ARIC study from North America (121 mmHg) was much lower than that in the Caerphilly study from Europe (141 mmHg). Consequently, the mean values at baseline and after 5 or 6 years of the top and bottom groups also differed:

	ARIC		Caerphilly	
Years between measurements:	0	6	0	5
Baseline groups I (bottom)	98	108	115	128
V (top)	149	142	167	164
Range (V-I)	51	34	52	36

However, the *difference* between the top and bottom groups, both at baseline and after 5 or 6 years of follow-up, was very similar. In order to combine such studies in an overview, a common position at which to plot each group was required, appropriate for both studies.

In addition to differences in mean values between cohorts at baseline, the mean values of blood pressure and cholesterol varied substantially during prolonged follow-up within individual studies (Chapter 5, Figures 3 and 4). For example, in Speedwell the mean SBP increased between baseline and the first re-survey after 3 years, presumably due to a change in the measurement technique. This was reflected in the changes in the mean values of each group, so that even the mean of the top group increased during this period (rather than decreased as anticipated). However,

because there was an even greater proportionate increase in the mean of the bottom group, the difference between the top and bottom groups was still substantially less than it had been at baseline.

Thus, it was necessary to standardise for differences in mean values of risk factor levels in different cohorts, and for any changes within cohorts during follow-up, before defining a simple method for estimating the magnitude of the *RDR* over prolonged follow-up. This was achieved by subtracting each individual's risk factor level from the mean or median of their cohort at each survey. Participants were first divided into 5 groups within each cohort according to increasing values of the baseline measurements, as in the previous chapter. The means of the standardised values within each of these groups were calculated at each survey, and plotted against the time between the baseline and the re-survey. The effects of standardisation on diastolic blood pressure (DBP) within three cohorts (Framingham, Caerphilly and Speedwell) with prolonged follow-up of 15 to 30 years are shown in Figure 1. The mean values of DBP fluctuated markedly during follow-up (Chapter 5, Figure 3), but this was greatly attenuated after standardisation (Figure 1). In Framingham the overall mean increased between years 2 and 12, decreased between years 12 and 14 and then increased again between years 18 and 20 (Chapter 5, Figure 3). Consequently, the means of the measured values in each of the five groups followed a similar pattern (Figure 1(a)). However, after standardising at each visit for these differences in the overall mean, the fluctuations in the group means largely disappeared. The mean of the top group decreased consistently and the mean of the bottom group increased over the 30-year follow-up. Hence, the interquintile ranges decreased progressively after a sharp initial convergence between baseline and the

first re-survey after 2 years. A similar pattern was observed in the Caerphilly and Speedwell studies (Figure 1(b)).

If the parameter of interest were normally distributed, the mean of the central group would be equivalent to the overall mean, and after standardisation would, subject to the play of chance, be zero. However, although blood pressure and cholesterol are approximately normally distributed, there was some slight skewness. Consequently, although the narrowing of the ranges was essentially symmetrical after standardisation, the mean of the central group was usually below zero and still tended to drift slightly over prolonged follow-up (Figure 1(b)). It might be anticipated that, in making no assumptions of normality, standardising for the median would account for this skew in the distributions, and hence the mean of the standardised values in the central group might be closer to zero. However, Figure 1(c) shows that the means of the central groups continued to deviate from zero, and to drift during follow-up.

When risk factor levels were standardised by subtracting each individual's risk factor level from the mean of the central group then, irrespective of the shape of the distribution, the means of the standardised values in the central groups were always zero (by definition): see Figure 1(d). This had the advantage that the artefact was obvious - that these standardised values were not "real" measurements - while the symmetry around the central group was largely maintained. Further, the rate of convergence also appeared similar between studies. [Although standardising for the mean of the central group was considered most appropriate for these data, it is the principle of standardising for each study's distribution that is important and not the exact method used to do this.]

2.2 Regression dilution over prolonged follow-up: a generally applicable correction procedure

The general principle adopted throughout was to use data from re-measurements within individual studies (where possible) when estimating the magnitude of the *RDR* for that particular study. Half of the studies had some repeat measurements, but most of these were within a few years of baseline. However, these early re-measurements could still be used to estimate the magnitude of the **short-term** *RDR* in those studies. The magnitude of the study-specific short-term *RDR* was estimated using the baseline measurement and the first re-measurement obtained within 10 years of the baseline from each individual. The baseline measurements *within each cohort* were divided into 5 groups, and individuals allocated a group number (I to V) according to the fifth in which their baseline measurement fell (with the lowest values in group I and the highest values in group V). All measurements were then standardised at each survey by the mean of the central group in their cohort (at that survey). Once standardised, data from all cohorts included in the analysis were combined. The means of the standardised values at baseline (Z_{0i}) and at the re-survey (Z_{1i}) within each of the groups (i=I to V) were calculated, the mean of group III being zero by definition. The absolute difference between the top and bottom groups (the "interquintile range") was calculated for the baseline and the re-survey. The ratio of the interquintile range at the first re-survey to that at baseline (the *ratio of ranges*) was then used to estimate the *RDR*, which was attributed to the average time to re-survey. This provided a study-specific estimate of the magnitude of the short-term *RDR* (an "*intercept*").

In the previous chapter, the regression lines (or "*slopes*") fitted to the estimates of the *RDR*s (for re-surveys with 500 participants or more) suggested that the *RDR* for SBP declined by 1.1% per year, and that for DBP by 1.8% per year in all regions. There were too few data to estimate reliably the decline in the *RDR* for total cholesterol in East Asia. However, since it declined by 1% per year in Europe, North America and Australia, this value was used for East Asia until sufficient data are obtained from this region (Eastern Stroke and Coronary Heart Disease Collaborative Research Group, 1998). These slopes were used to estimate *RDR*s over prolonged follow-up. The predicted *RDR* beyond the time of the *intercept* was estimated by assuming that, beyond the time of this *intercept*, the *RDR* subsequently declined according to the *slope* appropriate to that risk factor. For any study with no re-survey data, or with fewer than 500 participants re-surveyed, the *intercept* was assumed to be the average for its region (that is, Europe, North America or East Asia). The means of the *usual* standardised values for each group during any given exposure period could also be estimated using these slopes: -

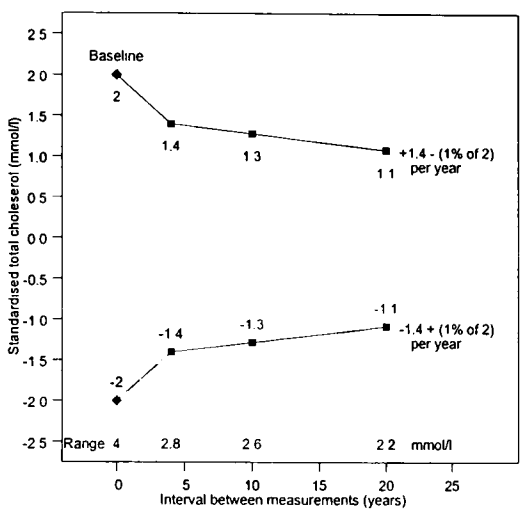
If t_1 denotes the average time to the first re-survey within 10 years, and Z_{1i} is the mean of the standardised values in group i at t_1, the mean of the *usual* standardised values at any time, t is: -

$$\{(Z_{1i}/Z_{0i})-(slope/100)(t-t_1)\}*Z_{0i}$$

Example:

Suppose the mean of the standardised values in group V was **+2** mmol/l at *baseline* (i.e., 2 mmol/l above the mean), and had fallen to **+1.4** mmol/l (i.e. 70% of baseline) after **4** years. Then, assuming the *RDR* for total cholesterol declined at a rate of **1.0%** per year, mean of the *usual* standardised values in group V after **10** years would be estimated as

$$\{(1.4/2.0) - 0.01(10-4)\} * 2.0 = +1.28 \text{ mmol/l}$$



For any study with no re-survey data, the means of the standardised values at baseline were calculated, and the predicted *usual* values for subsequent surveys estimated by multiplying the observed baseline means by the predicted estimate of the *RDR* for their region at that time.

The standardised *usual* mean values could be back-transformed for presentation: for example, an estimate of the overall mean during the risk period of interest could be added to each standardised mean value. Then, the estimated hazard ratio for the central group would be plotted against this overall mean of the population under study. However, particularly for cholesterol, where the mean values varied substantially between studies, this would disguise the true breadth of the cholesterol

values being studied. For the purposes of this thesis, risk will be plotted against the standardised *usual* mean values (so the hazard ratio for the central group will be plotted against zero).

Comparing predicted estimates with those observed more than a decade after the baseline survey

Only three studies (Caerphilly study in the UK, Busselton in Australia and Framingham in North America) had re-survey data from more than a decade after the baseline survey. The first survey within a decade of screening (for each participant) was used to calculate the *intercept* for that study, and then predicted estimates of the *RDR* were extrapolated beyond that time using the methods described above. The observed *ratios of ranges* (after 14 years in Caerphilly, 24 years in Busselton and 30 years in Framingham) were then compared with the predicted estimates. Since the Framingham study had measurements every two years, it could be used to assess whether deviations from the predicted lines were due to real discrepancies, or merely the play of chance, at a particular survey. In the absence of re-survey data from Framingham, the predicted estimates from North America would have been used to impute *usual* mean values in this study for any given period of follow-up beyond the baseline. Hence, the validity of the method could be assessed by comparing the observed estimates of the *RDR* (that is, the observed *ratio of ranges*) in Framingham at each survey with the predicted estimates.

3 Results

3.1 Predicting regression dilution ratios over prolonged follow-up

Systolic Blood Pressure

On average, the first re-measurements were collected four years after the baseline survey. At this time, the mean *RDR* for SBP was similar in Europe, North America and Australia at 70%, but was slightly more extreme in East Asia at 65%. However, the magnitude of the estimated *RDR* in individual studies varied within each region (see Table 1). Estimates of the *RDR* varied from 53% to 75% in studies from Europe, 51% to 79% in those from North America & Australia and 54% to 69% in those from East Asia. So, although the *RDR* was generally, if only slightly, more extreme in East Asia, there was considerable overlap between the regions. The short-term estimate of the *RDR* after four years was 70% in all individuals with available data.

The observed *ratios of ranges* from each survey, when plotted separately for each of the three regions and for "All PSC", were compared with biennial surveys in Framingham (Figure 2(a)). The predicted declines in the *RDR* with increasing follow-up are indicated in each panel by a line starting at the region-specific "*intercept*" and decreasing by 1.1% per year. The final plot compares data from Framingham with a line representing the predicted *RDR*s from "All PSC".

In North America & Australia, the predicted line closely matched the observed *ratios of ranges* for up to 24 years. In Europe and in "All PSC", although the short-term predicted *RDR* was slightly less extreme than the observed short-term *ratios of ranges*, the subsequent declines in each were roughly parallel. There were no data

beyond five years from East Asia to make any such comparisons. The predicted *RDR*s in each region after 5, 10, 15 and 20 years of follow-up are shown in Table 2. In "All PSC", the predicted *RDR*s for SBP were 0.68 after 5 years, 0.63 after 10 years and 0.52 after 20 years compared with *ratios of ranges* of 0.71, 0.65 and 0.52 observed in the Framingham multiple pairs. Table 2 also shows the expected interquintile ranges using these predicted *RDR*s. At baseline, the interquintile ranges varied from 48 mmHg in Europe to 57 mmHg in North America & Australia. Since the same decline was imposed on all regions, the proportional differences in the ranges persisted.

The predicted *RDR*s calculated using this method suggest that the biased risk relationships observed using only baseline measurements to define exposure, would underestimate the real relationships with *usual* SBP by about one-third after 10 years and one-half after 20 years.

Diastolic Blood Pressure

Estimates of the short-term *RDR*s for DBP were 10% more extreme than those for SBP in all regions. When the first re-measurements of participants from each study were considered together, the short-term estimate of the *RDR* after four years was 60%, compared with 70% for SBP at this time. As with SBP, the effects of regression dilution in DBP appeared to be similar in the European, North American and Australian studies and slightly more extreme (that is, the *RDR* was smaller) in the East Asian studies. Again, the magnitude of the estimated *RDR*s in individual studies varied within each region: Table 1.

Not only were the short-term *RDR*s more extreme for DBP, the decline in the *RDR* with increasing follow-up was also more extreme for DBP at 1.8% per year (see previous chapter). For many studies, the predicted *RDR*s appeared slightly less extreme than the observed *ratios of ranges*, that is they were underestimating the magnitude of the regression dilution bias. This was due to less extreme estimates of the magnitude of the **short-term** *RDR*s in each region (Figure 2(b)). Even these apparently conservative predictions of the magnitude of the *RDR*, however, suggested that use of baseline measurements would underestimate any risk relationships with DBP by about half after 10 years and by two-thirds after 20 years (Table 2). The observed *ratios of ranges* for DBP in Framingham were consistent with those predicted using all data from the PSC: Figure 2(b) and Table 2.

Although the predicted *RDR*s were similar between Europe, North America and Australia, the interquintile range at baseline was slightly smaller for Europe. By ten years, these differences had largely disappeared, and the predicted ranges were about 15 mmHg in all regions. (Table 2).

Total cholesterol

The estimates of the short-term *RDR*s for total cholesterol were similar to those for SBP (and hence less extreme than those for DBP). The decline in the *RDR* with increasing follow-up was 1.0% per year. The observed short-term *ratios of ranges* ranged from 57% to 76% in Europe, 61% to 74% in North America & Australia and from 23% to 78% in East Asia. The observed *ratios of ranges* in East Asia were particularly disparate with three estimates below 40% and two estimates over 70% (Table 1). Due to this wide variation in values from East Asia, the regional estimate of

the magnitude of even the short-term *RDR* should be viewed with caution. Indeed, the estimates of the *RDR* from "All PSC" - the totality of the available evidence - may well provide a better indication of the magnitude of the *RDR* in East Asia than relying on the six estimates available here, i.e. on a sub-group of the available evidence. The predicted *RDR*s for total cholesterol provided a reasonably good fit to the observed *ratios of ranges* (Figure 2(c)) in Europe, North America and Australia.

Although the observed *ratios of ranges* in Framingham were less extreme than in the rest of the PSC within the first five years (see previous chapter), they declined at a faster rate (1.5% per year). Therefore, after 20 years the observed *ratios of ranges* in Framingham were 5% more extreme than the predicted *RDR*s from "All PSC" (Table 1). However, after 20 years the predicted interquintile ranges in Europe and North America & Australia and the observed ranges in Framingham were all 1.6-1.7 mmol/litre.

In the PSC, the mean time from the baseline measurements of blood pressure and cholesterol to death was 11 years. For those dying before age 60, the mean time was 8 years; for those dying in their sixties it was 10 years, and for those dying in their seventies it was 13 years. Therefore, for the outcome analyses, estimates of the *RDR* at each of 8, 10, 11 and 13 years were required:

Interval between re- survey and baseline measurements	SBP	DBP	Total cholesterol
Baseline	1	1	1
8 years	0.65	0.52	0.66
10 years	0.62	0.48	0.64
11 years	0.61	0.46	0.63
13 years	0.59	0.43	0.61

3.2 Predicting standardised usual values over prolonged follow-up

Systolic Blood Pressure

Table 3 gives the means of the standardised values at baseline and at the first re-survey within 10 years (for each participant) in each of the five groups defined by baseline measurements. In each region, the values in the bottom group were, on average, nearer to zero (the mean of the central group) than those in the top group. The most striking observation from this table was the similarity between regions (even at baseline) in the mean values within a group. Although the mean SBP for group V was somewhat higher at baseline in North America & Australia (by 6-8 mmHg) than in either Europe or Asia, by the time of the first re-measurements 3 or 4 years later, this difference had largely disappeared. The mean values in each group in Framingham were similar to those from the remaining cohorts from North America.

The predicted lines appeared to fit the data from North America & Australia slightly better than the data from Europe (Figure 3(a)). There were no data beyond five years in Asia for comparison. The "All PSC" plot suggests that, had no appropriate data been available, use of the predicted means would have provided reasonable estimates of the *usual* values for most studies when compared with the observed data. The predicted means in the top and bottom groups in each region after 5, 10, 15 and 20 years are given in Table 4. Use of "All PSC" mean values would have underestimated the mean *usual* values in the top group in North America & Australia by only about 3 mmHg after five years and 2 mmHg after 15 years. Thus, use of "All PSC" to predict mean standardised *usual* values would have provided acceptable estimates for each group over 20 years in all regions. In Framingham, the observed

means in the top group were higher (i.e. less extreme) in the first 10 to 15 years than those predicted from "All PSC", but were compatible thereafter. The observed means of the bottom groups were consistently, but only slightly, lower than those predicted from "All PSC".

Diastolic Blood Pressure

The results for DBP were even more consistent between the regions than SBP or cholesterol. By the first re-measurement 3 or 4 years after the baseline, there were no differences between the regions in the group means (Table 3). Consequently, the predicted means over prolonged follow-up were also very similar between regions (after the same convergence of 1.8% per year had been applied). The predicted lines in the top and bottom groups of DBP seemed to provide particularly good fits to the observed means (Figure 3(b)) – probably better than those for SBP. Furthermore, the observed means from Framingham were particularly close to the predicted lines in both the top and bottom group. This was reassuring, given the marked asymmetry in the means of the absolute values (Figure 1(a)).

Total cholesterol

The available data for cholesterol were limited, but even so, the predicted lines provided a good fit to the observed *usual* values in both Europe and North America & Australia (Figure 3(c)). The mean (standardised) cholesterol values in each group were similar in Europe and North America & Australia, both at baseline and by the time of the first re-measurement 3 or 4 years later (Table 3). The predicted mean *usual* values after 5,10, 15 and 20 years differed only by about 0.1 mmol/litre (Table 4). Although the observed mean *usual* values from Framingham were a little less

extreme than the predicted mean *usual* values during the first few years of follow-up, after less than 10 years the predicted *usual* values lay over the observed *usual* values. There were insufficient re-measurements of cholesterol in East Asia to assess the model reliably.

For multi-group plots in the outcome analyses (see Chapter 3: Methods) estimates of the *usual* values of SBP, DBP and cholesterol at each of 8, 10, 11 and 13 years were required. The following table gives approximate observed standardised mean values in the top and bottom groups at the time of the baseline measurements, and (rounded) estimates of the *usual* standardised mean values 8, 10, 11 and 13 years later:

Interval between re- survey and baseline measurements	SBP (mmHg)		DBP (mmHg)		Total Cholesterol (mmol/l)	
	Bottom	Top	Bottom	Top	Bottom	Top
Baseline	-21	30	-14	16	-1.4	1.8
8 years	-13	18	-7	9	-0.9	1.2
10 years	-13	19	-6	8	-0.9	1.1
11 years	-13	18	-6	7	-0.9	1.0
13 years	-12	18	-5	7	-0.8	1.0

3.3 Comparison of predicted values with those observed more than a decade after the baseline survey: a sensitivity analysis

In both Caerphilly and Busselton, the predicted *RDR*s tended to be less extreme than the observed *ratios of ranges* (that is, they underestimated the importance of regression dilution) for both SBP and DBP (Figure 4), although this was more acute in Caerphilly. For *usual* values, the procedure predicted quite well for the bottom group, but the rate of convergence of the predicted *usual* levels in the top group was slightly less extreme than that of the observed *usual* levels (Figure 5). For

cholesterol, the predicted lines provided good fits to the observed *RDR*s and the *usual* values in both studies.

In Framingham, use of the predicted *usual* values would have meant an overestimate of at most 3 mmHg in any group over 20 years for SBP. Furthermore, any discrepancy between observed differences between the top and bottom groups and the predicted differences were always less than 5 mmHg. For DBP the method accurately predicted not only the mean difference between the top and bottom groups but also where these groups should be plotted (the *usual* values). For cholesterol, use of the predicted *usual* values would have meant an overestimate in the range of at most 0.1 mmol/litre in any group over 20 years.

The general compatibility between the observed and predicted (standardised) *usual* values show that if late re-surveys had not been available from these studies, the predicted *usual* values calculated by the method described in this chapter would have provided better estimates of regression dilution at those times than use of data from the early re-surveys.

4 Discussion

4.1 Regression Dilution Ratios

When combining data from different studies in an overview, the general principle in the analysis is to compare participants within a study before pooling the results. The approach adopted here first allocated participants to groups within their own study and standardised the measured values from each individual for that study's distribution before pooling the values. Standardising values for the study's distribution

led to similar values from different studies within each group, and thus made studies sufficiently comparable for pooling to be appropriate.

Although estimates of the *RDR* from within a study will always remain the ideal, these data suggest that the effects of regression dilution are comparable in most studies, and these effects generally become more extreme (that is, the *RDR*s decrease) with increasing follow-up. Furthermore, use of the sparse data available from late re-surveys in other studies will provide better estimates of the *RDR* over prolonged follow-up than use of data from early re-surveys within the study. This was demonstrated using data from early and late re-surveys in Busselton, Caerphilly and Framingham.

Any estimate of the *RDR* in Europe was dominated by IPC-Paris and Norwegian Counties, which provided data from over 50,000 participants re-surveyed at least once in the first decade. However, removal of these two cohorts made little difference to the regional *intercepts* for either SBP or DBP, possibly because the estimated *RDR*s from IPC-Paris were generally low (i.e., more extreme), while those from the Norwegian Counties were generally high (i.e., less extreme). Both studies gave quite high estimates of the short term *RDR*s for total cholesterol, but the European *intercept* became only a few percent more extreme when they were removed.

From Table 2 it can be seen that, over prolonged follow-up, results from Framingham were compatible with those from Europe, North America and Australia, but may have underestimated the importance of regression dilution by 5% or more in East Asia. Beyond 15 years of follow-up, the Framingham estimates were almost identical to those from "All PSC" suggesting that use of Framingham would provide adequate

estimates of the *RDR*'s if no other data were available (despite concerns by some authors to the contrary (Carroll and Stefanski, 1994)).

4.2 Usual values

Inevitably, the predicted estimates of *usual* values were not ideal for all studies. However, the general convergence of the mean values towards zero in all studies suggested that use of these predicted estimates was better, not only than use of the baseline measurements, but also than use of re-measurements taken within only a few years of the baseline survey. Since Busselton was the only study, apart from Framingham, with data beyond 15 years, the regression lines were strongly influenced by data from its 24-year re-survey. The regression lines could have been refitted without data from this late re-survey, but this would have negated the purpose of this approach – that is, to use any available long-term data for the majority of the studies that did not have such data. Furthermore, since the analyses have shown that differences between studies and regions were quite small in the short term, and since Framingham and Busselton were the only studies with longer-term data, it was appropriate to use their data.

Data from Caerphilly and Busselton suggested that the assumption of symmetrical convergence of the means may be too strong, and that different *slopes* should be used for each group. For example, maybe the *slope* for the top group should have been steeper than that for the bottom group. In general, however, the predicted estimates fitted Caerphilly, Busselton and Framingham (the studies with prolonged follow-up) reasonably well, and in all cases the predicted estimates of *usual* values at the last re-survey were closer to the observed values at that survey than were the

observed values from earlier surveys within a few years of the baseline. Although different *slopes* could be fitted to each group, the procedure would no longer be generic because this would require different *slopes* being calculated according to the number of groups being analysed. Thus, the current approach, assuming identical convergence, was simple and could be easily incorporated into an appropriate analysis.

4.3 Incorporating these methods into an overview

These methods can easily be incorporated into an overview. Individuals should be divided into approximately equally sized groups according to increasing levels of baseline blood pressure or cholesterol, and hazard ratios calculated for each of these groups in the usual way. These hazard ratios are then plotted, not against the mean of the blood pressure or cholesterol in each group, but against the mean of the standardised *usual* values. If the central group had been used to standardise values, then the hazard ratio for this group would be plotted at zero.

4.4 Other risk factors

The PSC also planned to assess the role of HDL-cholesterol, non-HDL cholesterol and different indices of blood pressure such as pulse pressure, mean arterial pressure and average blood pressure in chronic disease. The *regression dilution ratios* for these composite variables are derived in the same manner as the original variables, except that the composite variables must first be calculated for each individual at every relevant survey.

By June 1998, data on at least one re-survey measurement of HDL were available from 30,000 individuals in 8 cohorts. Most of these re-surveys were performed within 5 years of the baseline measurement, and so, only the short-term effects of regression dilution could be estimated reliably. Furthermore, information from more than one re-survey, to allow prediction of the decline in the *RDR* with increasing follow-up, was available from less than 3,000 individuals. Therefore, the available data could be used to estimate only the short-term effects of the regression dilution bias (the *intercept*) for HDL and non-HDL. However, any subsequent medium and longer-term effects were predicted by assuming the *RDR* declined at 1% per year as for total cholesterol. Although this assumption may well not be valid on further investigation, these *usual* values still allow better estimates of the real strength and shape of the relationships of *usual* HDL and non-HDL cholesterol with risk than would use of uncorrected baseline measurements. Further data on HDL re-measurements have been sought from the principal investigators, and will be available for publications resulting from these analyses. Until these data are available, the following table gives the observed baseline values in the top and bottom groups together with the *regression dilution ratios* and (rounded) estimates of the *usual* values 8, 10, 11 and 13 years later:

Interval between re- survey and baseline	HDL cholesterol RDR	*Usual* values (mmol /litre)		Non-HDL cholesterol RDR	*Usual* values (mmol /litre)		Average BP RDR	*Usual* values (mm Hg)	
Baseline	1	-0.42	0.63	1	-1.41	1.86	1	16	23
8 years	0.69	-0.28	0.42	0.62	-0.85	1.09	0.62	-10	14
10 years	0.67	-0.27	0.40	0.60	-0.82	1.06	0.60	-9	14
11 years	0.66	-0.27	0.40	0.59	-0.81	1.04	0.58	-9	13
13 years	0.64	-0.26	0.39	0.57	-0.78	1.00	0.55	-8	12

5 Conclusions and implications for epidemiology

This chapter investigated whether a simple correction procedure could be developed that would be appropriate for the analysis of all observational studies and their overviews. The novelty of the methods described lies in their use of available data within a study together with information provided by all other studies with appropriate data. The magnitude of the short-term *RDR* (the "*intercept*") was calculated from short-term repeatability data within a study, while the subsequent decline in the *RDR*s was predicted using longer-term data from all relevant studies. This allowed for different amounts of measurement error and other study- or population-specific causes of short-term intra-individual fluctuations, but assumed that the medium-term or longer-term intra-individual fluctuations were similar between studies and populations.

Unless appropriate time-dependent corrections are made for regression dilution, the strength of associations of disease with *usual* blood pressure or cholesterol will be substantially underestimated. The following table shows the approximate magnitude of the short-term *RDR* (within about 5 years of the baseline measurement) in each region, and the subsequent decline in this ratio over prolonged follow-up:-

		Short-term *RDR* (within 5 years)	Subsequent decline
SBP	**All PSC**	**70 %**	**~ 1 % per year**
	Europe	70 %	"
	N America & Australia	70 %	"
	East Asia	65 %	"
DBP	**All PSC**	**60 %**	**~ 2 % per year**
	Europe	60 %	"
	N America & Australia	60 %	"
	East Asia	55 %	"
Total cholesterol	**All PSC**	**70 %**	**~ 1 % per year**
	Europe	70 %	"
	N America & Australia	70 %	"
	East Asia	65 %	"

These data suggest that, for SBP and cholesterol the underestimation of any risk relationships is about 30% within five years of measurement increasing by approximately 1% for each subsequent year of follow-up. DBP is affected rather more strongly, with underestimation of any risk relationships of about 40% within five years, increasing at a rate of nearly 2% per year thereafter. These estimates are independent of any assumptions about the constancy of risks over time, about the biological mechanisms by which exposures affect disease, or about whether changes in exposure are real or artefactual.

Figures and tables

Figure 1: Mean diastolic BP at baseline and subsequent surveys in groups defined by baseline measurements: (a) measured values; (b) values standardised towards the mean; (c) values standardised towards the median; (d) values standardised towards the mean of the central group

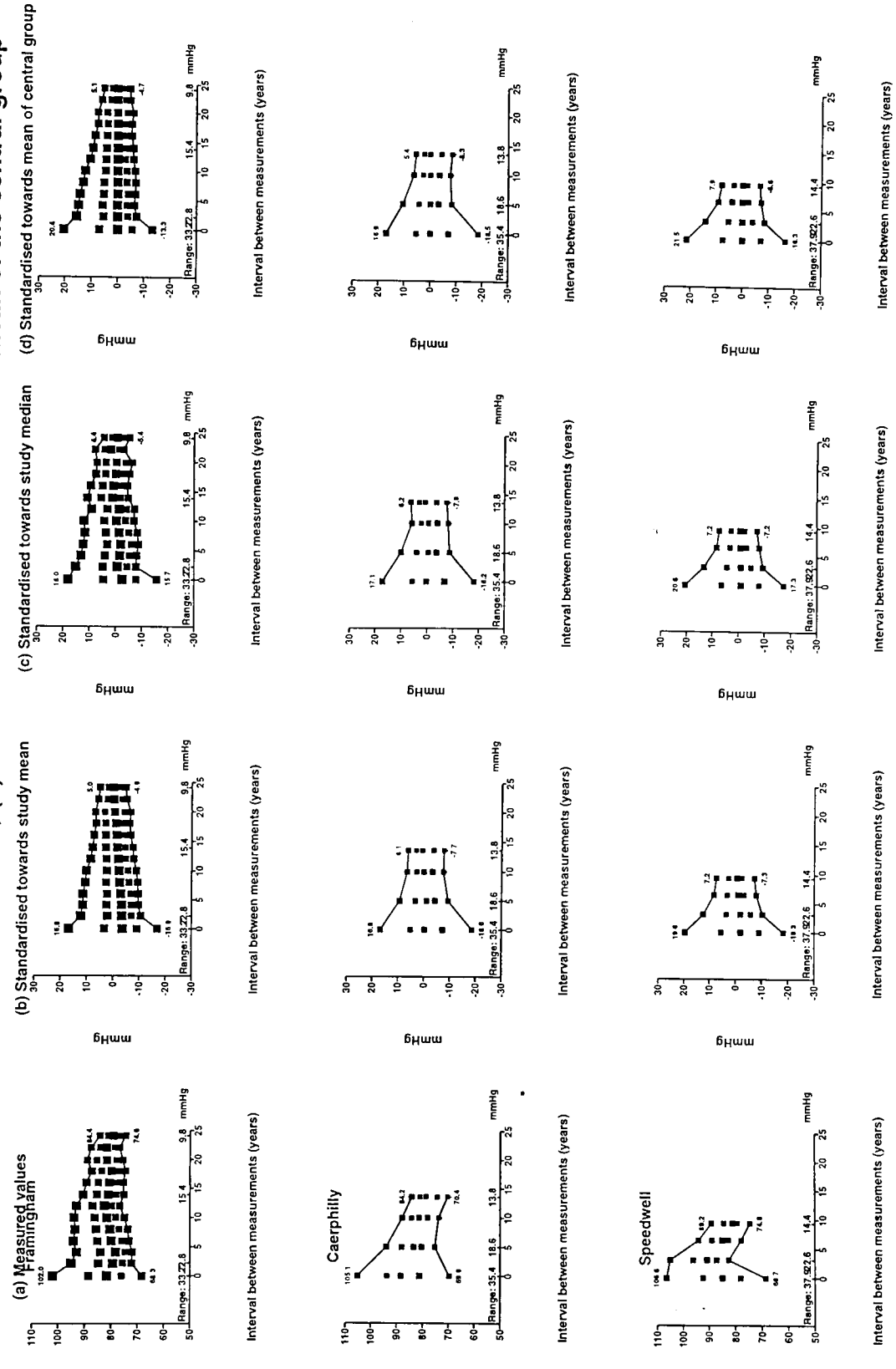


Figure 2: Observed and predicted regression dilution ratios based on ratio of ranges (a) Systolic BP

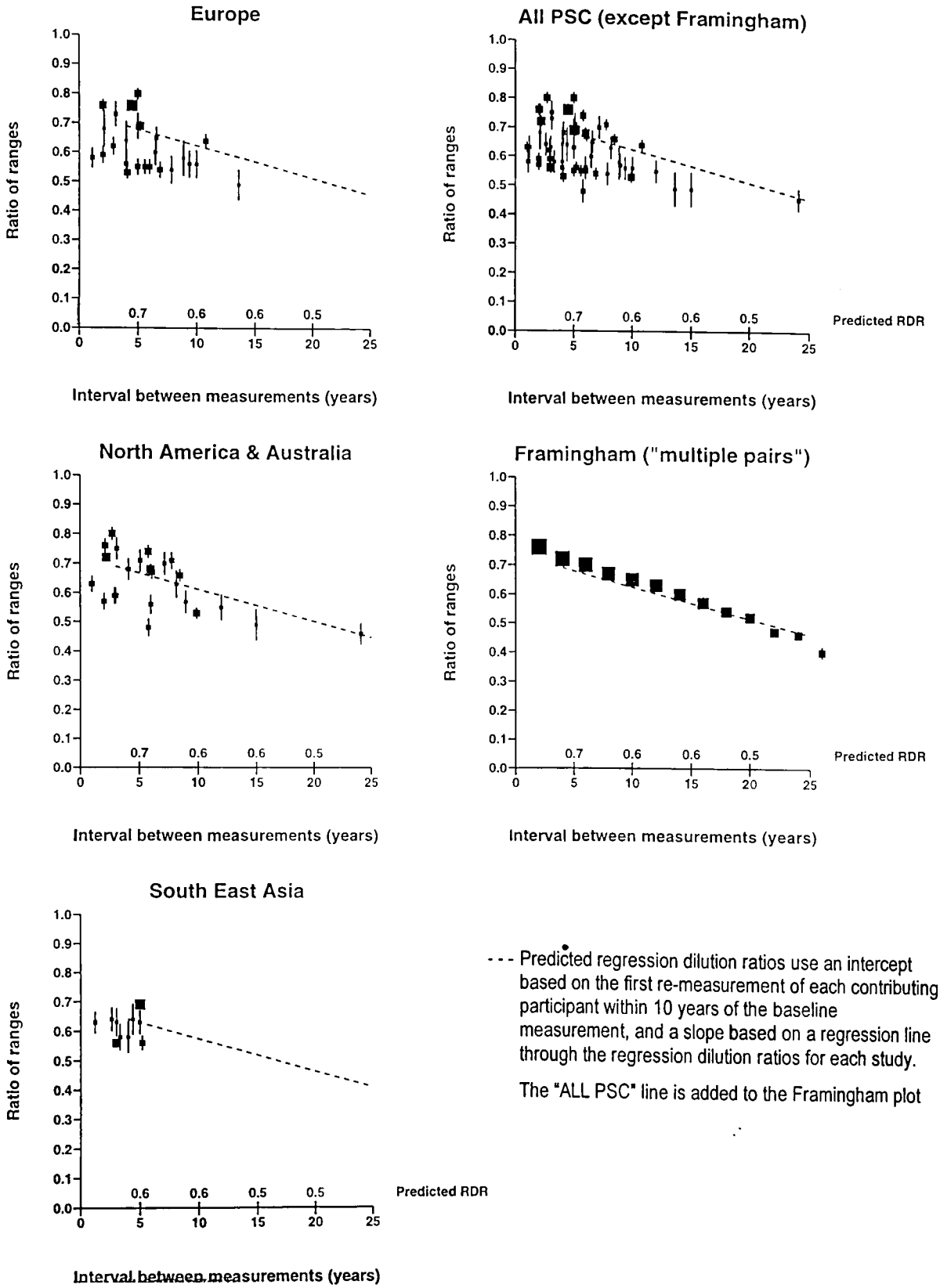
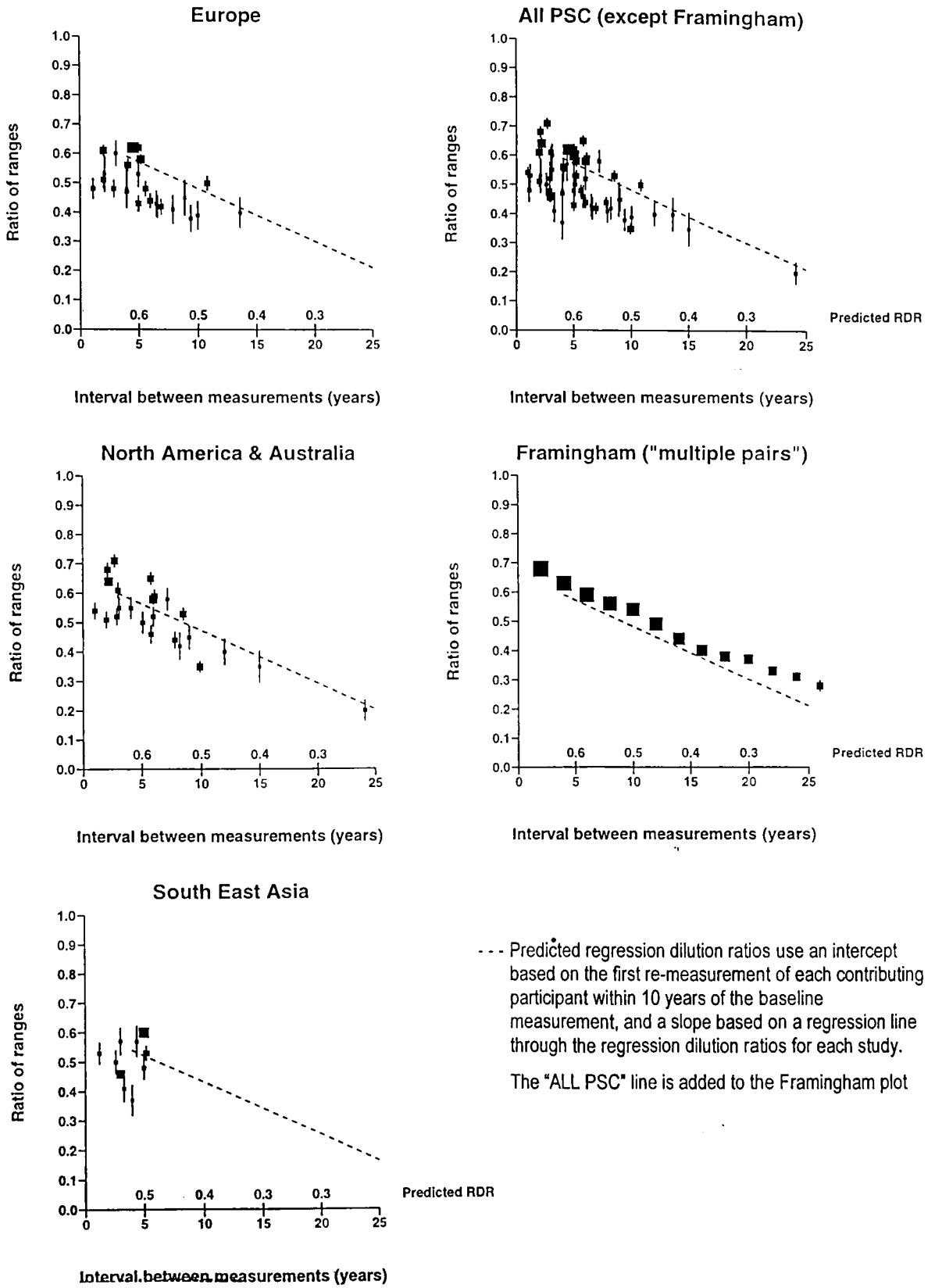


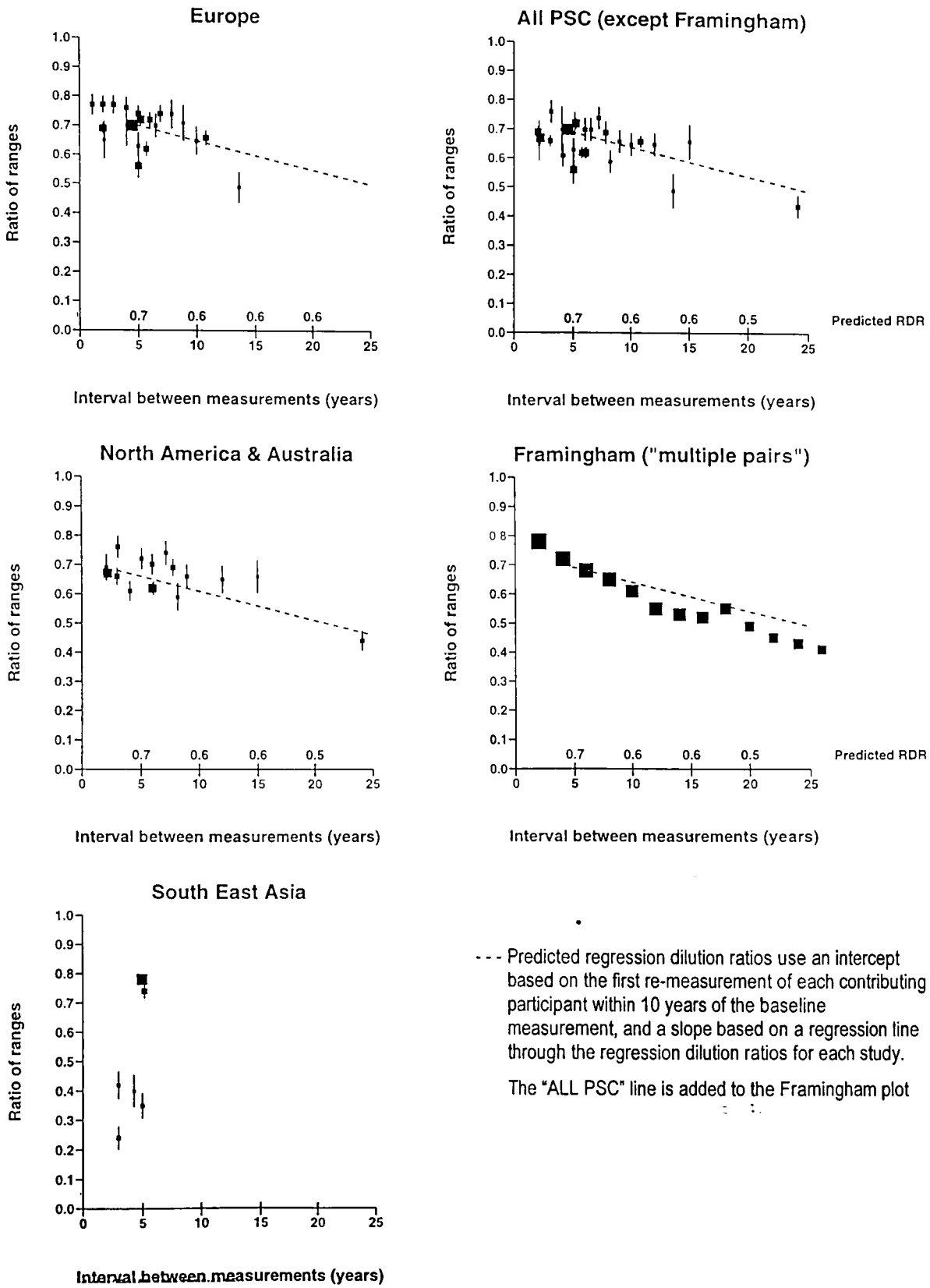
Figure 2: Observed and predicted regression dilution ratios based on ratio of ranges (b) Diastolic BP



- - - Predicted regression dilution ratios use an intercept based on the first re-measurement of each contributing participant within 10 years of the baseline measurement, and a slope based on a regression line through the regression dilution ratios for each study.

The "ALL PSC" line is added to the Framingham plot

Figure 2: Observed and predicted regression dilution ratios based on ratio of ranges (c) Total cholesterol

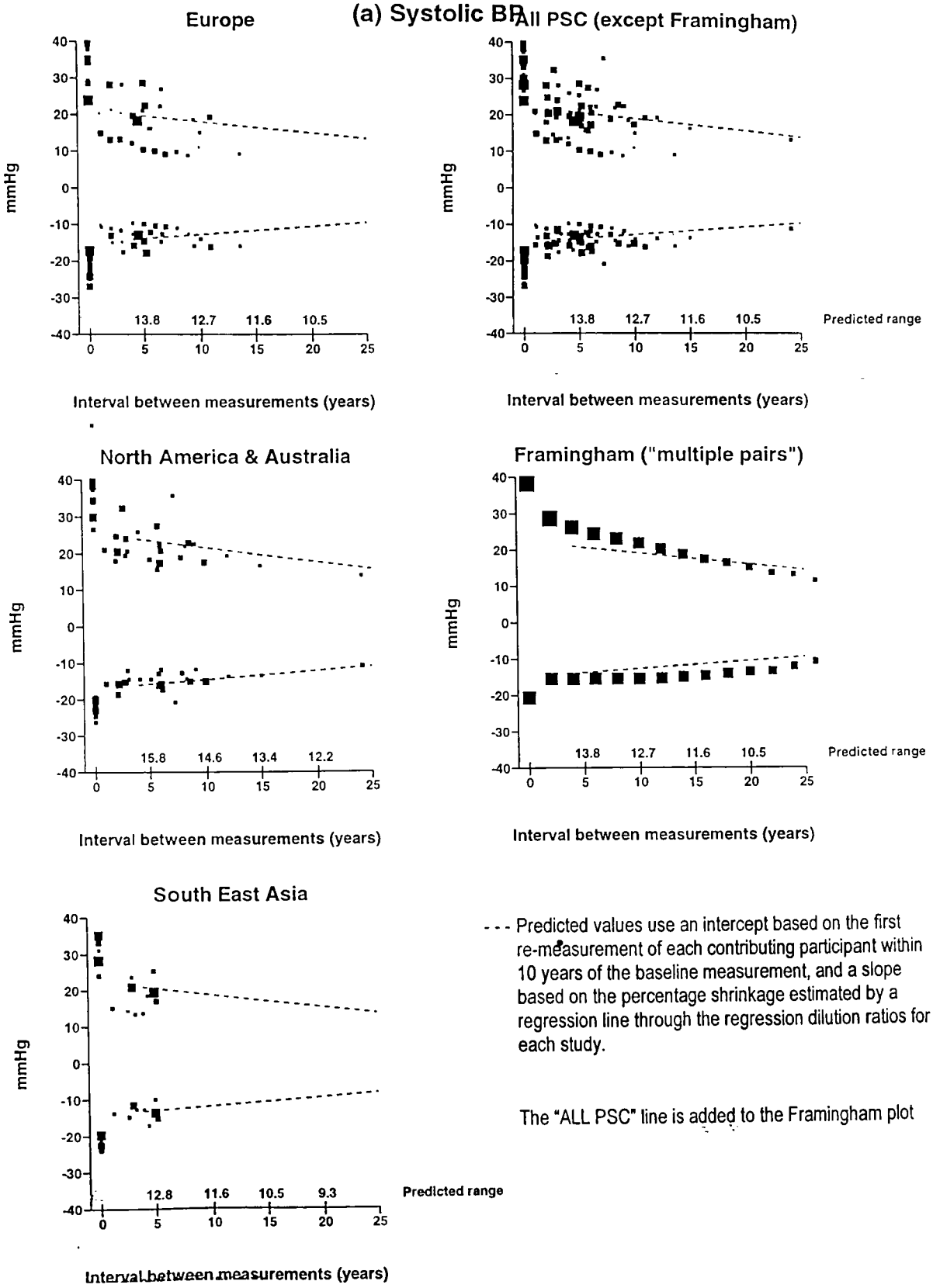


- - - Predicted regression dilution ratios use an intercept based on the first re-measurement of each contributing participant within 10 years of the baseline measurement, and a slope based on a regression line through the regression dilution ratios for each study.

The "ALL PSC" line is added to the Framingham plot

Figure 3: Mean standardised values in the top and bottom fifths defined by baseline measurements, at baseline and subsequent re-surveys

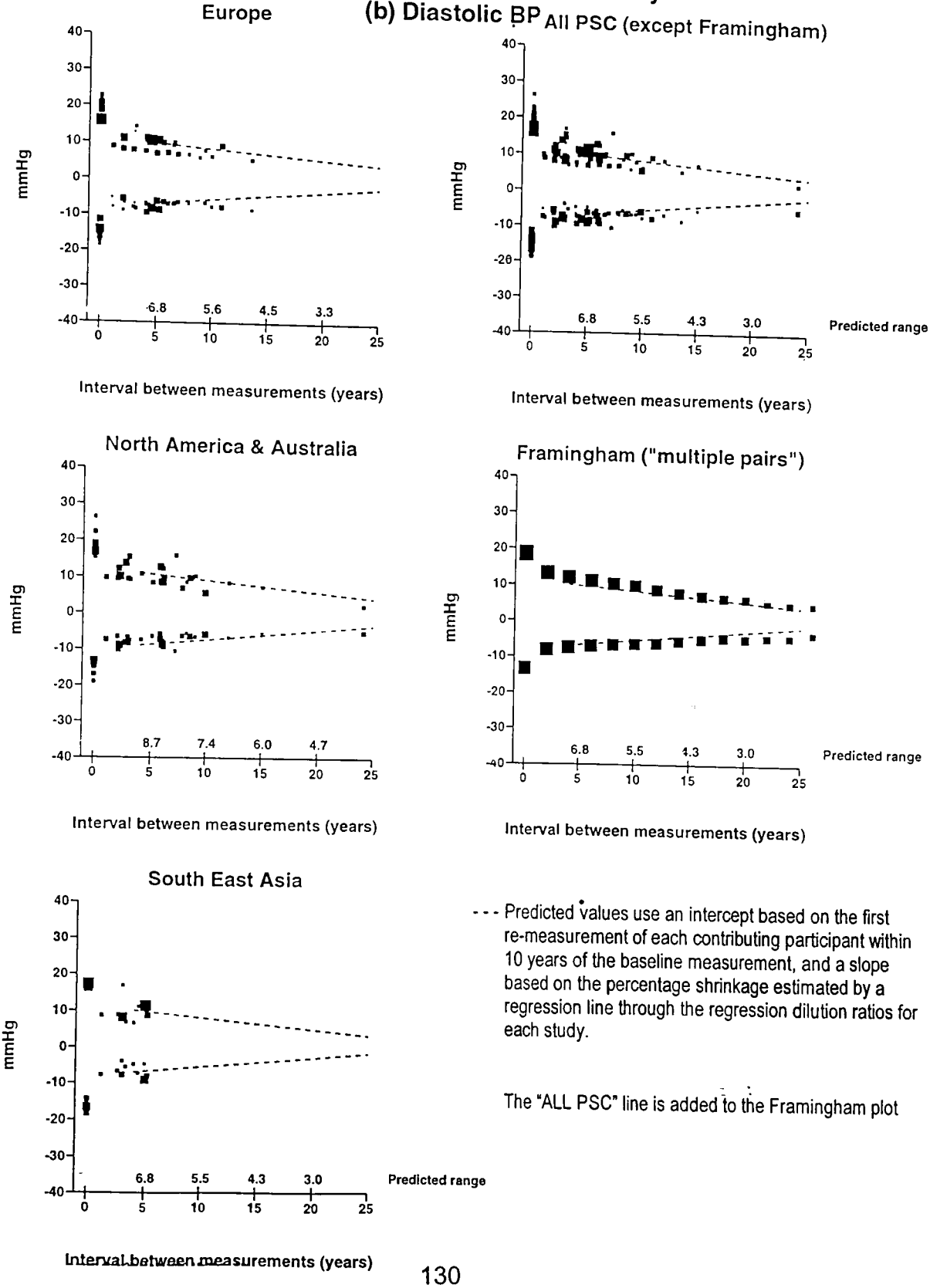
(a) Systolic BP



- - - Predicted values use an intercept based on the first re-measurement of each contributing participant within 10 years of the baseline measurement, and a slope based on the percentage shrinkage estimated by a regression line through the regression dilution ratios for each study.

The "ALL PSC" line is added to the Framingham plot

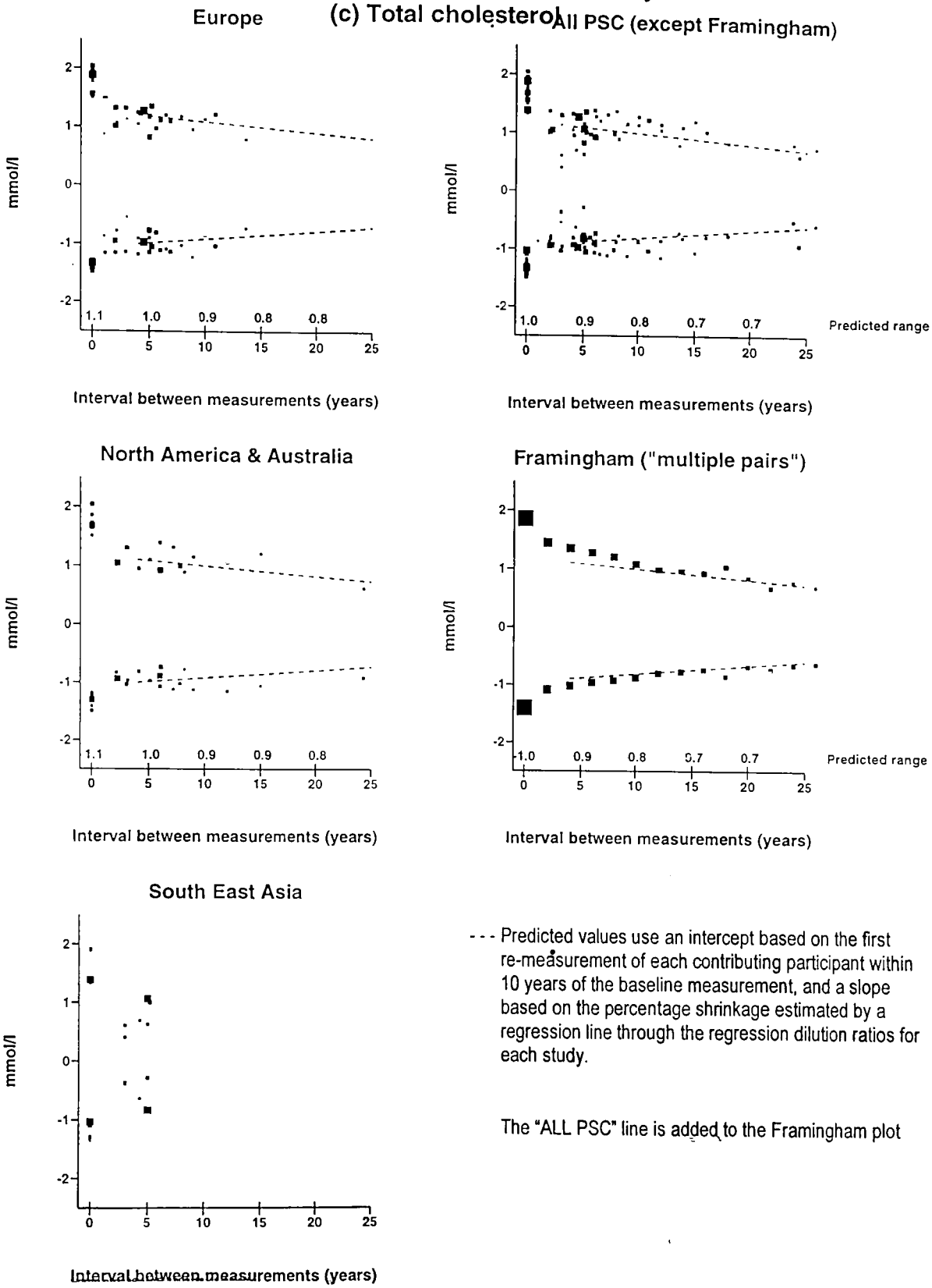
Figure 3: Mean standardised values in the top and bottom fifths defined by baseline measurements, at baseline and subsequent re-surveys



--- Predicted values use an intercept based on the first re-measurement of each contributing participant within 10 years of the baseline measurement, and a slope based on the percentage shrinkage estimated by a regression line through the regression dilution ratios for each study.

The "ALL PSC" line is added to the Framingham plot

Figure 3: Mean standardised values in the top and bottom fifths defined by baseline measurements, at baseline and subsequent re-surveys



--- Predicted values use an intercept based on the first re-measurement of each contributing participant within 10 years of the baseline measurement, and a slope based on the percentage shrinkage estimated by a regression line through the regression dilution ratios for each study.

The "ALL PSC" line is added to the Framingham plot

Figure 4: Observed and predicted regression dilution ratios for systolic BP, diastolic BP and total cholesterol in Caerphilly and Busselton

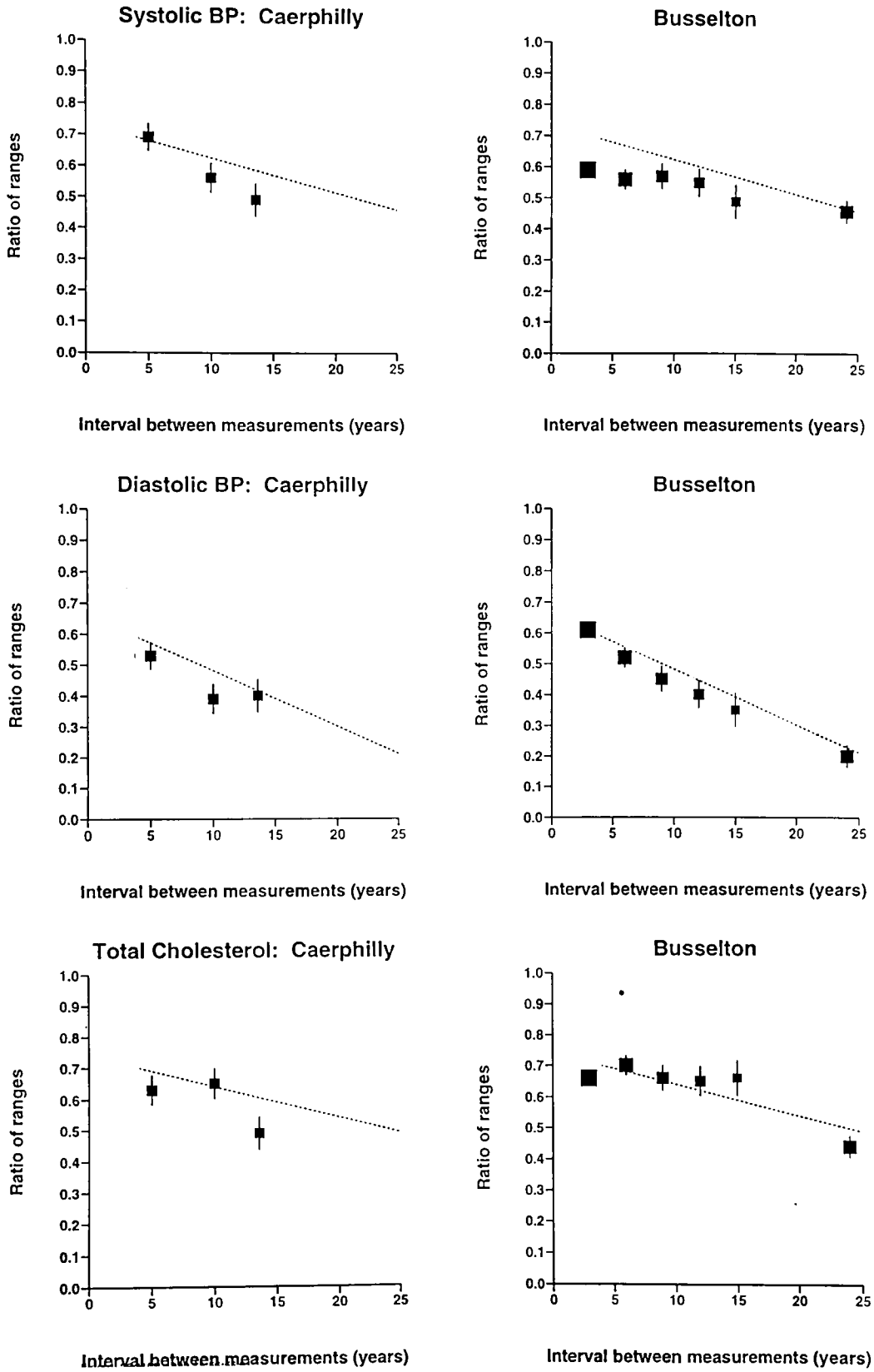


Figure 5: Observed and predicted "usual" values in the top & bottom fifths for systolic BP, diastolic BP and total cholesterol in Caerphilly and Busselton

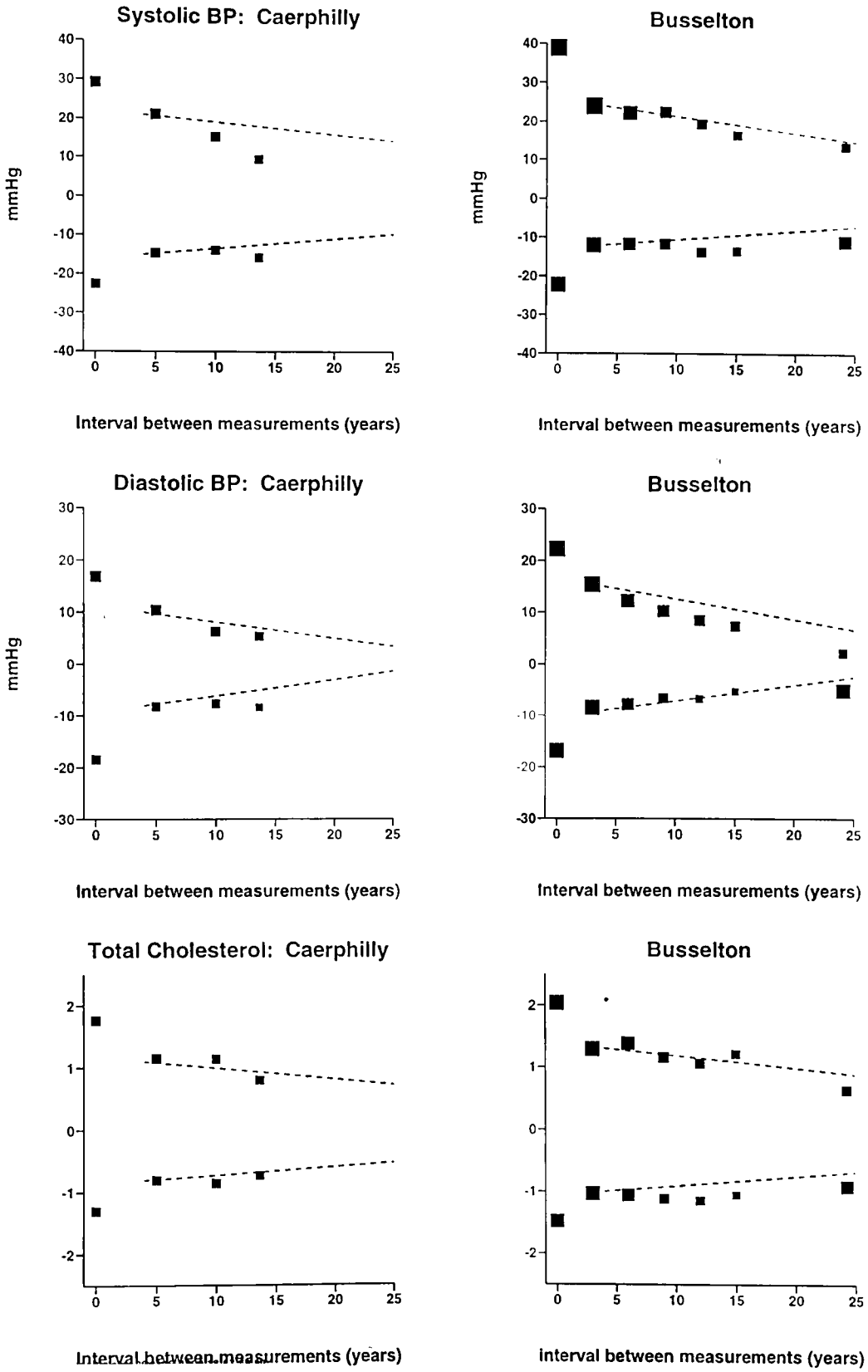


Table 1: Observed regression dilution ratios in each study, calculated using the first re-survey in each participant within 10 years

Region and study name	Mean interval between first re-survey and baseline measurements	Number re-surveyed (000)[1,2]	Systolic BP (mmHg)	Diastolic BP (mmHg)	Total Cholesterol (mmol/litre)
Europe	**4**	**99.7**	**0.69**	**0.59**	**0.71**
Caerphilly	5	1.8	0.69	0.53	0.62
Copenhagen	5	13.8	0.67	0.57	0.71
Finrisk	6	2.5	0.73	0.59	0.71
Israeli IHD	2	9.9	0.75	0.63	0.69
IPC-Paris	5	28.2	0.57	0.46	0.76
Norwegian	5	24.8	0.76	0.63	0.70
NPHS	7	2.3	0.64	0.43	0.69
Paisley/Renfrew	4	8.1	0.53	0.56	0.57
Speedwell	3	2.1	0.71	0.59	[1]
UK HDPP	5	6.2	0.59	0.49	0.62
N. America & Australia	**3**	**57.4**	**0.69**	**0.60**	**0.68**
ARIC	2	14.3	0.71	0.63	0.67
Busselton	3	5.2	0.59	0.60	0.66
CHS[2]	1	5.0	0.60	0.53	[1]
Evans County	7	2.5	0.67	0.58	0.73
Honolulu	2	7.6	0.75	0.68	0.68
MHHP[2]	4	5.4	0.73	0.52	0.74
NHEFS[2]	9	5.4	0.51	0.35	[1]
Puerto-Rico HHP[2]	3	9.2	0.79	0.70	[1]
Tecumseh	4	2.8	0.66	0.52	0.61
East Asia	**4**	**45.0**	**0.64**	**0.54**	**0.67**
Akita	5	6.2	0.56	0.52	0.74
Japanese Railway	5	22.5	0.69	0.61	0.78
Ohasama[2]	1	2.3	0.64	0.56	[1]
Seven cities: China	3	8.4	0.54	0.48	0.23
Shanghai	4	4.2	0.65	0.49	0.39
Shibata	4	1.4	0.61	0.57	0.40
All PSC (excl. Framingham)	**4**	**254.8**	**0.69**	**0.59**	**0.70**
Framingham (multiple pairs)	2	53.2	0.77	0.70	0.78

1 Not all individuals provided re-measurements of blood cholesterol

2 In some studies, the numbers differ from those presented in Chapter 3, Table 2. This table only gives the number of participants with a re-survey **within 10 years** of the baseline visit

Table 2: Predicted regression dilution ratios and interquintile ranges for standardised values within each region for studies with re-survey data: systolic BP, diastolic BP and total cholesterol

Interval between re- survey and baseline measurements	Region	Systolic BP (mmHg)		Diastolic BP (mmHg)		Total Cholesterol mmol) litre)	
		RDR	Interquintile range	*RDR*	Interquintile range	*RDR*	Interquintile range
Baseline	Europe	1	48	1	28	1	3.1
	N. America/Australia	1	57	1	33	1	3.0
	East Asia	1	53	1	32	1	2.6
	All PSC**	**1**	**51**	**1**	**30**	**1**	**3.1**
	Framingham (multiple pairs) ***	1	58	1	32	1	3.2
5 years	Europe	0.68	33	0.58	16	0.70	2.2
	N America/Australia	0.67	38	0.57	19	0.67	2.0
	East Asia	0.63	33	0.55	18	*	*
	All PSC**	**0.68**	**35**	**0.59**	**18**	**0.69**	**2.1**
	Framingham (multiple pairs) ***	0.71	41	0.61	20	0.70	2.2
10 years	Europe	0.63	30	0.49	14	0.65	2.0
	N America/Australia	0.62	35	0.48	16	0.62	1.9
	East Asia	0.58	31	0.46	15	*	*
	All PSC**	**0.63**	**32**	**0.50**	**15**	**0.64**	**2.0**
	Framingham (multiple pairs) ***	0.65	38	0.54	17	0.61	2.0
15 years	Europe	0.57	27	0.40	11	0.60	1.9
	N America/Australia	0.56	32	0.39	13	0.57	1.7
	East Asia	0.52	28	0.37	12	*	*
	All PSC**	**0.57**	**29**	**0.41**	**12**	**0.59**	**1.8**
	Framingham (multiple pairs) ***	0.59	34	0.42	13	0.53	1.7
20 years	Europe	0.52	25	0.31	9	0.55	1.7
	N America/Australia	0.51	29	0.30	10	0.52	1.6
	East Asia	0.47	25	0.28	9	*	*
	All PSC**	**0.52**	**27**	**0.32**	**10**	**0.54**	**1.7**
	Framingham (multiple pairs) ***	0.52	30	0.37	12	0.49	1.6

* There were only six re-surveys from 5 studies, so predicting regression dilution ratios was not possible.

** excluding Framingham

*** Framingham values are observed values. Years 5 and 15 are interpolated from years 4 & 6 and years 14 & 16 respectively.

Table 3: Means[1] of the standardised values at baseline and at the first re-survey within 10 years in each of five groups defined by baseline measurements: systolic BP, diastolic BP and cholesterol

Mean interval between baseline and re- survey	Region	I	II	III	IV	V	Difference (V-I)
Systolic BP							
Baseline	Europe	-20	-9	0	9	28	48
	USA/Australia	-22	-9	0	11	36	58
	East Asia	-21	-9	0	10	32	53
	All PSC	**-21**	**-9**	**0**	**9**	**30**	**51**
	Framingham (multiple pairs)	-21	-9	0	12	38	59
3-4 years later	Europe	-14	-6	0	6	20	34
	USA/Australia	-16	-7	0	9	24	40
	East Asia	-13	-6	0	8	21	34
	All PSC	**-14**	**-5**	**0**	**7**	**21**	**35**
2 years later	Framingham (multiple pairs)	-16	-7	0	10	29	45
Diastolic BP							
Baseline	Europe	-13	-6	0	5	15	28
	USA/Australia	-15	-5	0	6	18	33
	East Asia	-14	-6	0	7	18	32
	All PSC	**-14**	**-6**	**0**	**6**	**16**	**30**
	Framingham (multiple pairs)	-13	-5	0	6	18	31
3-4 years later	Europe	-7	-3	0	4	10	17
	USA/Australia	-9	-4	0	4	11	20
	East Asia	-7	-4	0	4	10	17
	All PSC	**-7**	**-3**	**0**	**4**	**10**	**17**
2 years later	Framingham (multiple pairs)	-8	-3	0	4	13	21
Cholesterol							
Baseline	Europe	-1.4	-0. 6	0	0.6	1.8	3.2
	USA/Australia	-1.3	-0.5	0	0.6	1.7	3.0
	East Asia	-1.1	-0.4	0	0.5	1.5	2.6
	All PSC	**-1.4**	**-0.6**	**0**	**0.6**	**1.8**	**3.2**
	Framingham (multiple pairs)	-1.4	-0.6	0	0.6	1.9	3.3
3-4 years later	Europe	-1.0	-0.4	0	0.4	1.2	2.2
	USA/Australia	-0.9	-0.4	0	0.4	1.2	2.1
	East Asia	-0.8	-0.3	0	0.3	1.0	1.8
	All PSC	**-1.0**	**-0.4**	**0**	**0.4**	**1.2**	**2.2**
2 years later	Framingham (multiple pairs)	-1.1	-0.5	0	0.5	1.4	2.5

1 Blood pressure values are rounded to nearest integer and cholesterol values to one decimal place.

Table 4: Predicted mean standardised values in the top and bottom groups within each region for studies with re-survey data: systolic BP, diastolic BP and total cholesterol

Interval between re- survey and baseline measurements	Region	Systolic BP (mmHg)		Diastolic BP (mmHg)		Total Cholesterol (mmol/ litre)	
		Bottom	Top	Bottom	Top	Bottom	Top
Baseline	Europe	-21	28	-13	15	-1.4	1.8
	N. America/Australia[1]	-22	36	-15	18	-1.3	1.7
	East Asia	-21	32	-14	18	-1.1	1.5
	All PSC[1]	**-21**	**30**	**-14**	**16**	**-1.4**	**1.8**
	Framingham (multiple pairs)[2]	-21	38	-13	18	-1.4	1.9
5 years	Europe	-13	19	-7	10	-1.0	1.2
	N America/Australia[1]	-15	24	-8	11	-0.9	1.2
	East Asia	-13	20	-7	10	-0.8	1.0
	All PSC[1]	**-14**	**21**	**-7**	**10**	**-1.0**	**1.2**
	Framingham (multiple pairs)[2]	-16	25	-7	12	-1.0	1.3
10 years	Europe	-12	18	-6	8	-0.9	1.1
	N America/Australia[1]	-14	22	-7	9	-0.9	1.0
	East Asia	-12	19	-6	8	-0.7	0.9
	All PSC[1]	**-13**	**19**	**-6**	**9**	**-0.9**	**1.0**
	Framingham (multiple pairs)[2]	-15	22	-7	10	-0.9	1.1
15 years	Europe	-11	16	-5	7	-0.9	1.0
	N America/Australia[1]	-13	20	-6	7	-0.8	0.9
	East Asia	-11	17	-4	7	-0.6	0.8
	All PSC[1]	**-12**	**18**	**-5**	**7**	**-0.8**	**1.0**
	Framingham (multiple pairs)[2]	-15	18	-5	7	-0.8	1.0
20 years	Europe	-10	15	-3	6	-0.8	1.0
	N America/Australia[1]	-12	18	-4	5	-0.7	0.8
	East Asia	-9	15	-3	5	-0.6	0.7
	All PSC[1]	**-10**	**16**	**-4**	**6**	**-0.7**	**0.9**
	Framingham (multiple pairs)[2]	-14	15	-5	6	-0.7	0.9

1 excluding Framingham

2 Framingham values are observed values. Years 5 and 15 are interpolated from years 4 & 6 and years 14 & 16 respectively.

Chapter 7: Blood pressure and premature death

Establishing the real relationships by appropriate analyses of large-scale epidemiological evidence

Chapter summary

The associations of blood pressure and cholesterol with various causes of death have been investigated in many prospective observational studies. Yet the importance of these two risk factors has generally been seriously under-estimated, due to insufficient deaths from particular causes within individual studies and, importantly, to inappropriate analyses of most studies. In particular, previous analyses have not (a) allowed for regression dilution; (b) considered the full range of blood pressure or cholesterol; (c) been appropriately age-specific; and (d) been able to determine reliably associations with less common causes of death (due to insufficient deaths). As a consequence, the extent to which a moderate reduction in blood pressure or cholesterol can influence the risk of vascular diseases has been under-estimated. With data on 136,000 deaths in the Prospective Studies Collaboration, the true age-specific relationships of the usual blood pressure and cholesterol with various causes of death can be established by appropriate statistical analysis. This chapter illustrates the importance of using appropriate statistical analyses to asses the true relationships, in particular, describing the cumulative effect of correcting for each of these errors when assessing the age-specific role of blood pressure in ischaemic heart disease and stroke risk. Furthermore, because of the large dataset available, appropriate analyses allow relationships with less common causes of death to be explored reliably.

1 Results: the relevance of blood pressure in premature mortality

Although the blood pressure-risk relationships are discussed largely with regard to systolic blood pressure (SBP), the risk-relationships with systolic and diastolic blood pressure (DBP) are given in the figures and tables. The strength of the risk relationships with average blood pressure (ABP: see later) have also been tabulated.

1.1 Blood pressure: the link with ischaemic heart disease and stroke

Blood pressure and ischaemic heart disease

Even using uncorrected baseline measurements of blood pressure, the PSC data confirmed findings from individual studies of a strong association between SBP and ischaemic heart disease (IHD) death over all ages, suggesting the real relationship was likely to be somewhat stronger. This association persisted after adjustment for total cholesterol, smoking and (relative) weight, so that overall, a 25% lower risk of IHD death was associated with each 21 mmHg lower SBP (Figure 1(a)), although there was significant heterogeneity between the results from individual studies (see later). Using baseline measurements, the relationship appeared to be log-linear, but with some suggestion of attenuation in the strength of the relationship (i.e. plateauing of risk) at both extremes of the blood pressure distribution.

Correcting for the regression dilution bias

As discussed in detail earlier in this thesis, the use of baseline measurements tends to underestimate the real strength of any relationship between a risk factor and the disease of interest because of regression dilution. With an average time to death of 11 years, the relationship shown in Figure 1(a) underestimated the true strength of

the SBP-IHD relationship by 35-40%. After correction for regression dilution, a 25% lower risk of IHD was associated with only 13 mmHg lower usual SBP, rather than 21 mmHg (Figure 1(b)), and the apparent levelling of the association at the top of the blood pressure distribution disappeared. Indeed, if anything, the relationship steepened.

Considering the full range of blood pressure

With more than 3,000 IHD deaths among individuals in the bottom fifth of their cohort's distribution and 16,000 among those in the top fifth, it was possible to sub-divide the population more finely, with sufficient data even in the extremes of the risk factor distributions to assess reliably the risk relationships. After dividing each cohort into 15 groups (according to increasing values of SBP), there were 1,000 deaths among those with the lowest blood pressure (whose usual SBPs were, on average, 15 mmHg below the mean SBP for their cohort) and 6,000 deaths among those with the highest blood pressure (whose usual SBPs were, on average, 25 mmHg higher than the mean SBP for their cohort). The relationship remained log-linear throughout the entire range, with no evidence of thresholds above which higher SBP was not associated with increased risk, *nor* below which lower SBP was no longer associated with lower risk (Figure 2(a)). When only acute myocardial infarctions were considered, the relationships were about 25% steeper (Table 1(a)). Therefore, whatever the initial level of usual SBP (within the range studied), each 10 mmHg decrease was associated with 21% lower risk of IHD. If the strength of this relationship is similar in different populations, this transforms into a log-linear relationship from around 105 mmHg in populations where blood pressure levels are typically low (such as the ARIC study with a mean SBP of 121 mmHg) to around 175

mmHg in populations where blood pressure levels are typically high (such as the Evans County study with a mean SBP of 149 mmHg).

Appropriate age-specific analyses

Most individual studies have only a limited number of deaths among those aged under 60, thus preventing any reliable estimation of the risk relationship within this age group. However, when data from these studies were combined, there were 8,000 individuals (7,000 men and 800 women) who had died from IHD before they were 60. A further 12,000 (10,000 men and 2,000 women) had done so in their sixties, and 12,000 (8,000 men and 4,000 women) in their seventies. To determine any differences between men and women in the magnitude of the relationships between blood pressure and vascular death, they were initially analysed separately *within* each age group to avoid confounding by age at death. Figure 3 shows how the importance of SBP in IHD risk varied substantially with age at death, being much stronger in the younger age groups, and this trend was similar for men and women. Therefore, when determining the shapes of the risk relationships separately within different age groups, men and women have been analysed together (in sex-standardised analyses) for statistical stability. Furthermore, the age-specific analyses were restricted to fifths of SBP because, even with data from over one million people, finer divisions led to unstable estimates of risk within age-specific sub-groups (particularly at younger ages).

Deaths in early middle age tended to occur a few years closer to the baseline measurements than those in old age, and thus, the extent to which regression dilution had the potential to bias the risk relationships varied slightly by age.

Therefore, different (appropriately time-dependent) corrections for regression dilution were used *for each age group*. After correction for regression dilution, a 25% lower risk of IHD death was associated with only 9 mmHg lower usual SBP among those dying before age 60 (Figure 3(a)). This shows a relationship that is more than twice as steep as that suggested by the initial inappropriate analysis which took all ages together and made no correction for the regression dilution bias. The table below shows how the importance of blood pressure in premature deaths caused by IHD was under-estimated by inappropriate analyses: -

Difference in blood pressure (mmHg) associated with 25% lower risk of premature death caused by IHD (N~40,000), estimated from each analysis

	mmHg ↓	
	SBP	DBP
Baseline BP, all ages	21	13
Usual BP, all ages	13	6
Usual BP, **age 40-59 (N~8,000)**	9	5

Blood pressure and stroke

The relationship between blood pressure (systolic and diastolic) and stroke was substantially steeper (on a logarithmic scale) than that between blood pressure and IHD, although there was significant heterogeneity between study results (see later). Uncorrected baseline measurements suggested that risk of stroke (of all types) was 25% lower for every 16 mmHg lower SBP (Figure 4(a)). After correction for regression dilution, only 10 mmHg lower usual SBP resulted in a 25% lower stroke risk (Figure 4(b)), and this relationship remained log-linear throughout the entire range (Figure 5(a)).

There were 1,400 deaths from stroke among those dying in their forties and fifties, 2,400 among those dying in their sixties and 2,900 among those dying in their seventies. When the SBP-stroke relationship was investigated separately within age- and sex-specific groups, the relationship was about 70% steeper in early middle age than in old age. Within each age group, however, the risk relationship was similar for men and women (Figure 6(a)). In early middle age, there was a 25% lower stroke rate for every 6 mmHg lower usual SBP (Figure 6(a)) or 3mmHg lower DBP (Figure 6(b)). The table below shows how the proportional importance of blood pressure in premature deaths due to stroke (of any type) was under-estimated by inappropriate analyses: -

Difference in blood pressure (mmHg) associated with 25% lower risk of premature death caused by stroke (N~13,000), estimated from each analysis

	mmHg ↓	
	SBP	DBP
Baseline BP, all ages	16	9
Usual BP, all ages	10	4
Usual BP, **age 40-59 (N~2,000)**	6	3

Hence, as with IHD, inappropriate correction for, or total neglect of, the regression dilution bias and lack of age-specific analyses has meant that people have seriously underestimated the strength of the association between usual blood pressure and stroke in middle age. Consequently, the extent to which even small reductions in blood pressure might substantially reduce the subsequent risk of stroke in middle age has been grossly underestimated.

Stroke sub-types

Several studies and one overview (Eastern Stroke and Coronary Heart Disease Collaborative Research Group, 1998) have observed stronger associations between blood pressure and haemorrhagic stroke than between blood pressure and ischaemic stroke. However, at younger ages a greater proportion of strokes may be due to haemorrhage (Bamford et al., 1990), and within these data, the distribution of age at death was markedly different for the two types of strokes. About 25% of the haemorrhagic strokes occurred in those aged 40-59 compared with only 15% of the ischaemic strokes, and therefore it was fundamentally important to consider the age at which an individual died from stroke. Otherwise, any differences observed in the risk associations may be artefacts of age at death. Of the 13,000 stroke deaths in the PSC database, 3,000 were classified as haemorrhagic, 2,500 were classified as ischaemic and 1,000 were due to subarachnoid haemorrhage. The remaining 6,500 could not be classified (see Chapter 4: Study Characteristics).

If all ages were (inappropriately) analysed together, the relative hazards for subarachnoid and primary intra-cerebral haemorrhages appeared about 25% stronger than that for ischaemic stroke. However, when analysed by age at death, the relative hazards for ischaemic and haemorrhagic stroke were the same within each age band:

Difference in *usual* SBP[1] (mmHg) associated with 25% lower risk of haemorrhagic and ischaemic stroke deaths, by age at death

Age at death (years)	Haemorrhagic		Ischaemic	
	N	mmHg ↓	N	mmHg ↓
All ages[2]	2,896	9	2,545	11
40 – 59	648	6	245	6
60 – 69	801	8	622	8
70 – 79	954	12	999	10

[1] Differences in usual DBP are approximately half the differences in SBP.
[2] Includes deaths before age 40 and at 80 or above.

Thus, in early middle age, every 6 mmHg lower usual SBP was associated with 25% lower risk of either haemorrhagic or ischaemic stroke.

After accounting for the somewhat stronger effects of regression dilution, each 5 mmHg difference in DBP was associated with a similar change in the risk of these vascular deaths as 10 mmHg difference in SBP throughout the entire range, and in all ages (Table 1-2(b) and Figure 1(c) and (d) and Figure 2-8(b)).

1.2 Blood pressure and other causes of death

Other vascular causes of death

There was little effect of blood pressure on the risk of death from rheumatic heart disease, but apart from that blood pressure (both systolic and diastolic) was positively related to the other vascular causes of death. For each 10 mmHg lower systolic BP there were 20% fewer deaths due to "other vascular" causes (that is, non-IHD, non-stroke vascular causes). In particular, blood pressure was strongly associated with heart failure and aortic aneurysm. It was also strongly and positively associated with

renal and diabetic deaths, most of which will have had a large vascular component (Table 1).

Non-vascular causes of death

Blood pressure was quite strongly associated with upper aerodigestive cancers (i.e. mouth, oesophagus and larynx) – each 10 mmHg lower SBP being associated with 17% fewer deaths, and with deaths from hepatic diseases - each 10 mmHg lower SBP was associated with 24% fewer deaths. These associations did not attenuate when deaths within 5 years of the blood pressure measurement were removed to explore the possibility of reverse causality (Table 2). However, alcohol consumption both increases blood pressure and causes liver cirrhosis and upper aerodigestive cancers. Consequently, much of these apparent associations may have been due to confounding. To investigate these potential confounders, the subset of individuals for whom smoking and alcohol data were available was re-analysed. The unadjusted results were comparable to those from the whole dataset, suggesting this was an unbiased subset, but although adjusting for the crude variables of current/not current smoker and current/not current drinker attenuated the relationships with liver disease, it did not attenuate the relationships with upper aerodigestive cancers (Table 3). More work is needed to investigate this surprisingly strong association.

Blood pressure was negatively related to three major lung diseases: chronic obstructive pulmonary disease (COPD), pulmonary tuberculosis (TB) and lung cancer.

2 Implications for further research

Different indices of blood pressure

There are several different indices of blood pressure that might provide additional information about risk, or be better predictors of risk, over and above those provided by SBP and DBP. Indices such as pulse pressure (the difference between SBP and DBP) and the mean arterial pressure ($^1/_3$SBP+$^2/_3$DBP) have previously been considered. Inclusion of SBP and DBP in the regression model simultaneously suggested that a simple average of SBP and DBP might provide the best indicator, and for each disease the log hazard ratios for a 10 mmHg increase in average blood pressure (ABP) and its standard error together with the χ^2 for each model are shown in Tables 1-3(c). Comparison of the χ^2 suggest that, at least for vascular diseases, systolic blood pressure was a better predictor of risk than diastolic blood pressure, but that this simple average provided a more reliable predictor of risk than either pressures alone (or, mean arterial pressure or pulse pressure). The divergence between the systolic and diastolic pressures with age suggests that, as for SBP and DBP, the relevance of all these indices might be importantly different at different ages.

Controlling for confounders

Although some preliminary investigation of the effects of confounders on the relationships between blood pressure and cause-specific mortality risks was performed within this thesis (Table 3), a more thorough investigation awaits analyses from the wider scope of the PSC. In particular, analyses will be repeated excluding participants with any history of vascular or diabetic disease, and any apparent

associations of blood pressure with tobacco-related or alcohol-related illnesses will be assessed in non-smokers and non-drinkers, respectively. Although it will not be possible to control fully for any confounding effects of alcohol using the available data, if the relationships between blood pressure and alcohol-related diseases are attenuated after controlling for the crude measures of alcohol consumption available, then a more reliable measure could presumably attenuate the relationships still further, perhaps leaving no independent associations.

Heterogeneity

There was significant heterogeneity between the studies for the effects of BP on both CHD (Figure 7) and stroke (Figure 8) that could not be fully explained by different years of baseline measurements, mean values at baseline, the proportion of females or measurement techniques. However, a major contributing factor was the attenuation of the proportional effect of SBP with increasing age. More work is needed to explore the heterogeneity, in particular, looking at the age-specific strength of the associations between blood pressure and cause-specific vascular mortality risks separately in individuals with and without prior history of vascular disease, in diabetics and non-diabetics and in different levels of obesity.

Preliminary age- and region-specific analyses suggested that the strength of the blood pressure-stroke relationships might differ between the different regions studied (Figure 9), although within each region the relative hazards were log-linear and attenuated with age. However, there were only a few hundred stroke deaths and even fewer IHD deaths in each age-group and limited data to control for differentials in important confounders, such as smoking, alcohol and relative weight. More

detailed investigation of regional differences awaits the second phase of the Eastern Stroke and Coronary Heart Disease Collaboration (Eastern Stroke and Coronary Heart Disease Collaborative Research Group, 1998).

There were concerns that studies using self-reported measurements (US Health Professionals (Rimm et al., 1993), US Nurses Health Study (Belanger et al., 1978) and US Physicians Health Study (Buring et al., 1995)) would show qualitatively different relationships to conventional observational studies. Sensitivity analyses comparing age-specific hazard ratios for IHD and stroke death per 10 mmHg increase in SBP for the North America & Australia studies before and after excluding these studies suggested that they can usefully contribute to the combined analyses:

Comparison of log hazard ratios for North America & Australia with and without studies with self-reported blood pressure measurements

Number of deaths HR (99% CI)		Age at risk 40-59	 60-69	 70-79
IHD	With	791 1.48 (1.40-1.57)	1,649 1.38 (1.32-1.43)	2,193 1.23 (1.19-1.28)
	Without	691 1.49 (1.40-1.58)	1,376 1.35 (1.29-1.40)	1,995 1.23 (1.18-1.28)
All stroke	With	204 1.74 (1.58-1.90)	430 1.50 (1.40-1.61)	808 1.35 (1.28-1.43)
	Without	182 1.70 (1.54-1.87)	374 1.49 (1.39-1.61)	767 1.35 (1.27-1.43)

3 Discussion

The primary aim of this chapter was to illustrate the extent to which inappropriate statistical analysis can lead to misleading results concerning the role of blood pressure in premature death due to IHD and stroke, and in other less common causes of death. If such inappropriate analyses are also based on insufficient data, the scope for random error further limits any interpretation that might be made. Only

by taking proper account of (a) the regression dilution bias; (b) the full range of blood pressure; (c) age at death; and (d) by studying adequate numbers of deaths, does it become possible to establish reliably the real relationship between blood pressure and IHD and stroke, and between blood pressure and other, less common, causes of death. Thus, by taking proper account of each of the above, these analyses demonstrate that as a cause of premature death due to IHD or stroke, blood pressure is more than twice as important as was previously supposed. The associations were approximately log-linear throughout the entire range of blood pressures studied (that is, there was no threshold above or below which blood pressure was not positively associated with IHD risk). Although similar in men and women, the sizes of the associations were highly dependent on age at death for both IHD and stroke, with the proportional relationships being nearly twice as strong among younger people than among those who were older. The risk associations were about 30% steeper for stroke than for IHD at all ages, but after appropriate stratification by age at death, were similar for haemorrhagic and ischaemic strokes.

With regard to other causes of death, blood pressure was positively associated with renal and diabetic disease, and although there were too few deaths from these diseases to determine reliably the age- and sex-specific relationships with blood pressure independently, many of the deaths would have been vascular in nature. Thus, the associations with blood pressure are primarily concerned with the vascular consequences of these diseases, and by applying the principles set out in this thesis, the importance of blood pressure as a determinant of risk can be realised.

4 Conclusion

In conclusion, blood pressure is a far more important determinant of premature death from IHD and stroke than has previously been realised, in particular among those who are younger. Furthermore, the extent to which risk is diminished by a given prolonged lower blood pressure was independent of the blood pressure range studied in the PSC.

Figures and tables

Table 1: Strength of the associations of *usual* blood pressure with various causes of death:
log hazard ratios per (a) 10 mmHg higher SBP; (b) 5 mmHg higher DBP; and (c) 10 mmHg higher ABP

Table 2: Strength of the associations of *usual* blood pressure with various causes of death, excluding deaths within the first 5 years:
log hazard ratios per (a) 10 mmHg higher SBP; (b) 5 mmHg higher DBP; and (c) 10 mmHg higher ABP

Table 3: Strength of the associations of *usual* blood pressure with various causes of death, before and after controlling for smoking (yes/no), drinking (yes/no) and relative weight at baseline, and excluding deaths within the first 5 years:
log hazard ratios per (a) 10 mmHg higher SBP; (b) 5 mmHg higher DBP; and (c) 10 mmHg higher ABP

Figure 1: Proportional IHD risks by
(a) baseline SBP, (b) usual SBP, (c) baseline DBP and (d) usual DBP
Ischaemic heart disease
(Total: 39287 deaths)

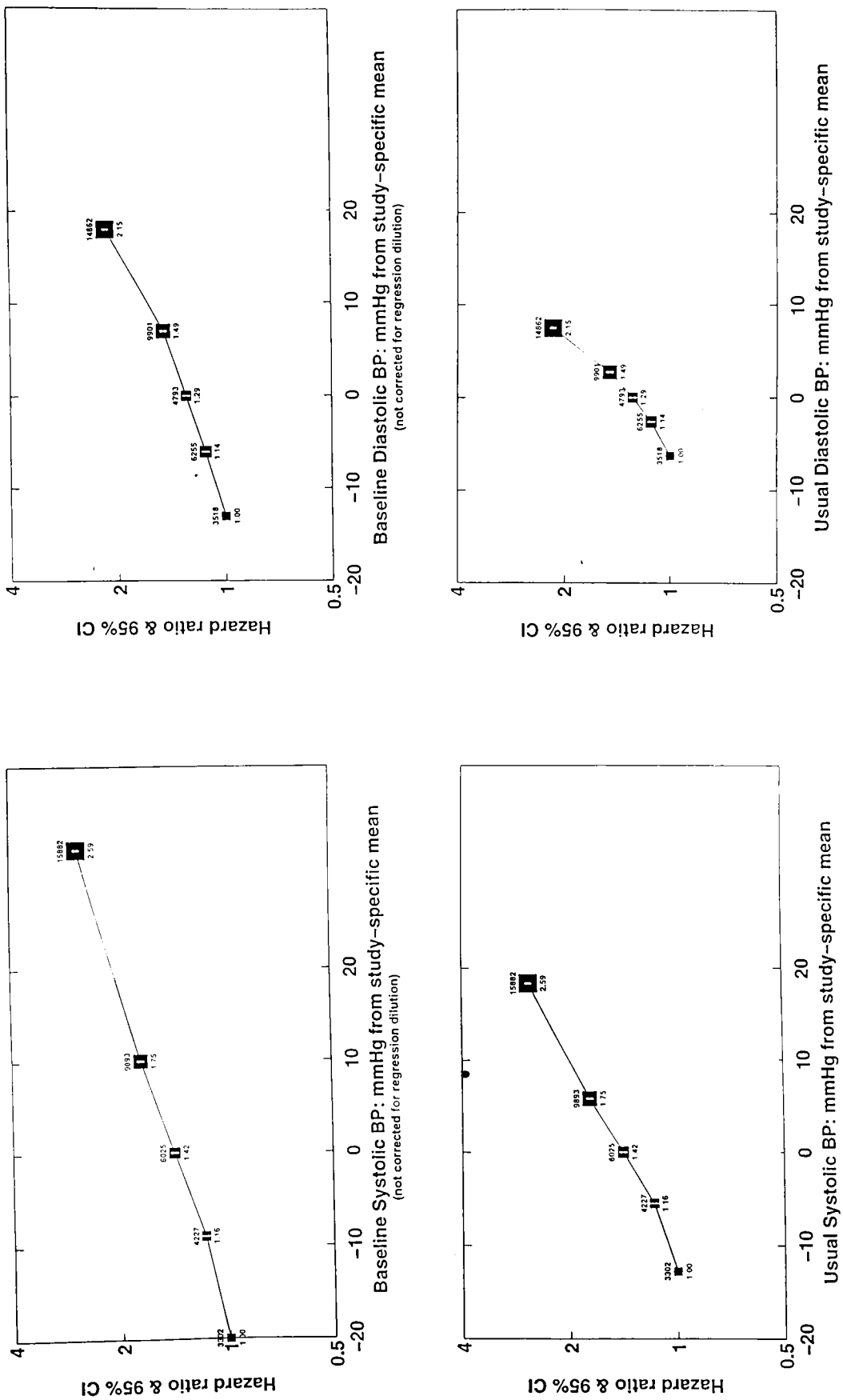
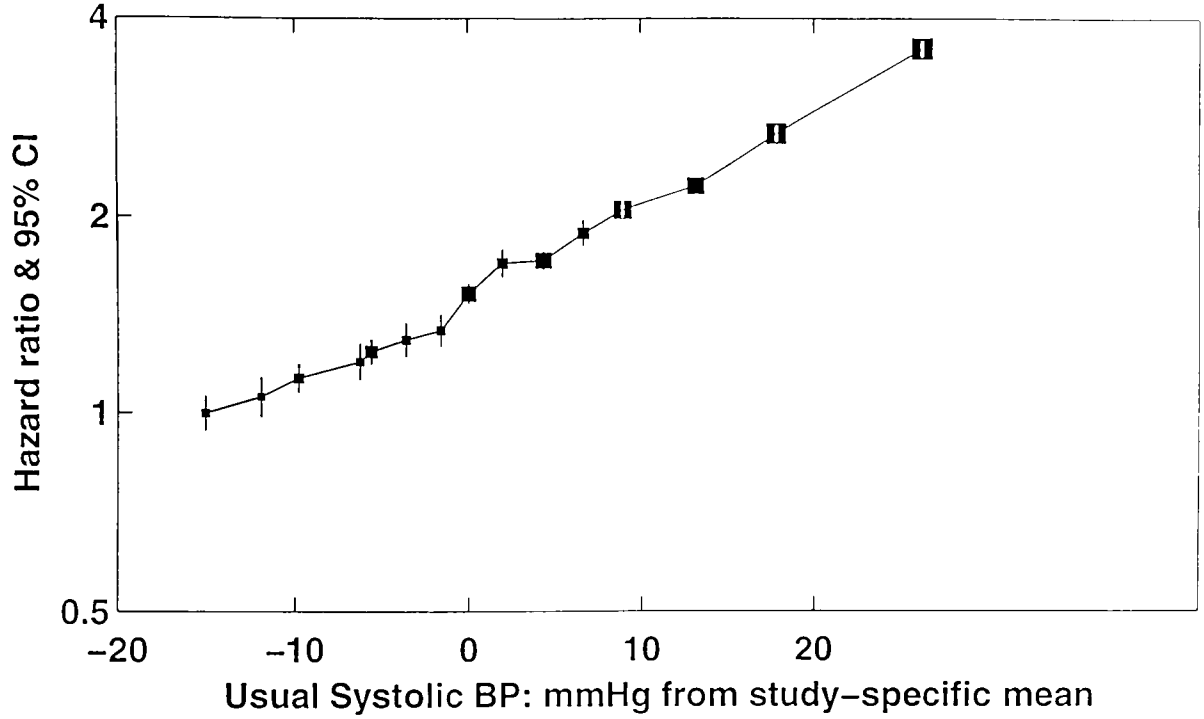


Figure 2: Proportional IHD risks throughout the entire range of (a) usual SBP and (b) usual DBP

Ischaemic heart disease

(Total: 39287 deaths)



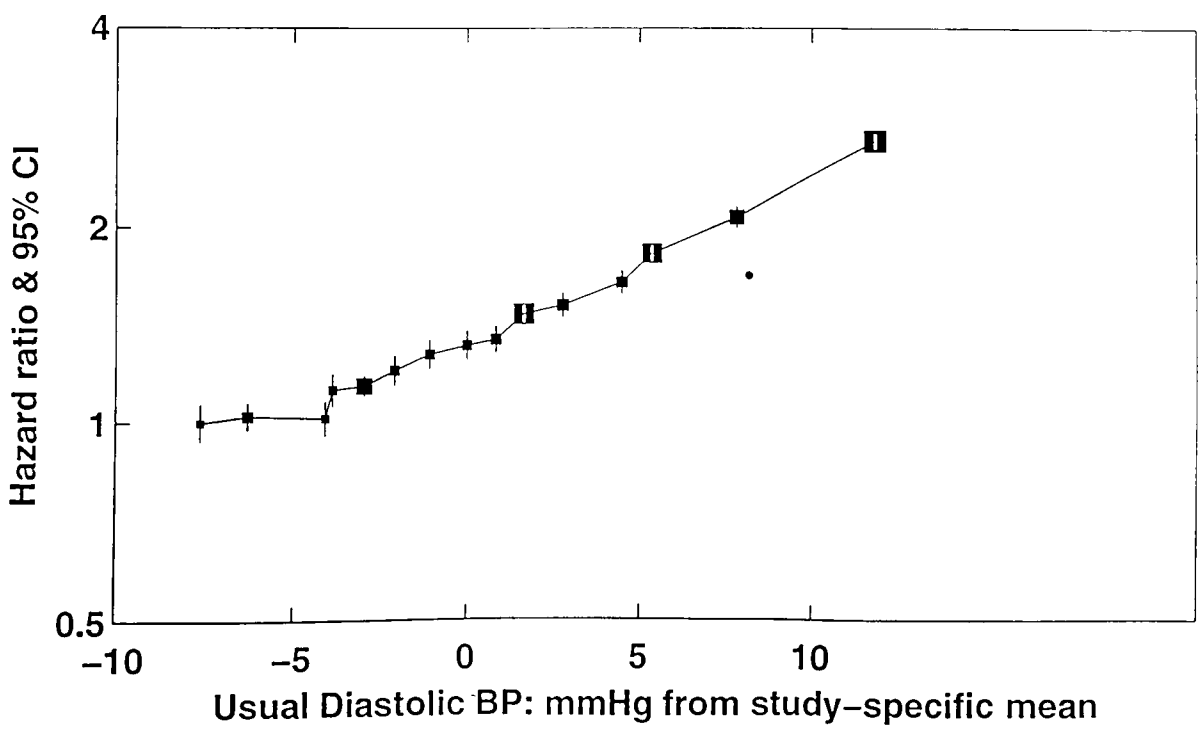
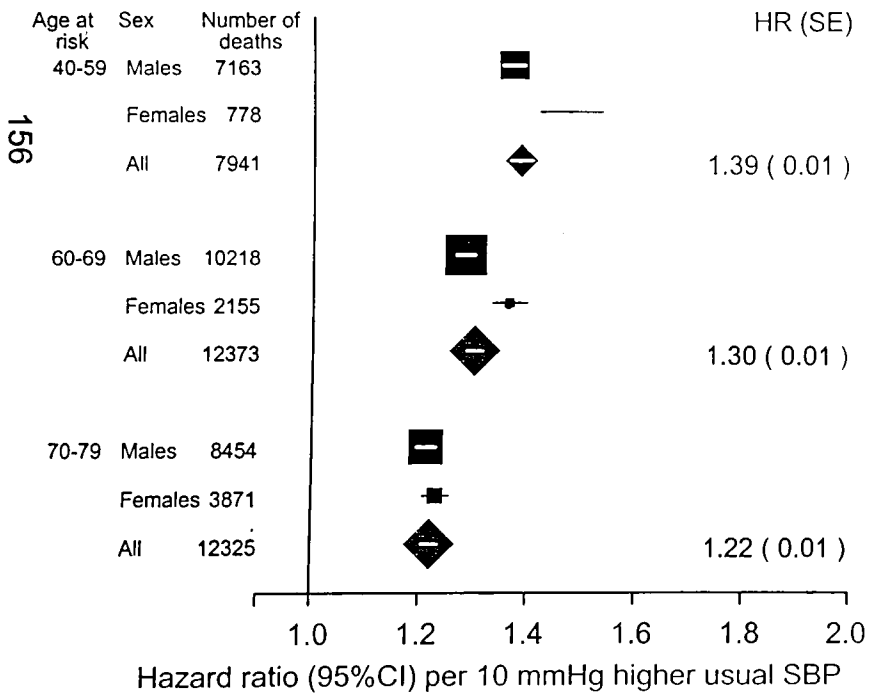


Figure 3(a): Strength and shape of the age-and sex-specific associations between SBP and IHD risk

Ischaemic heart disease

(Total: 32639 deaths)



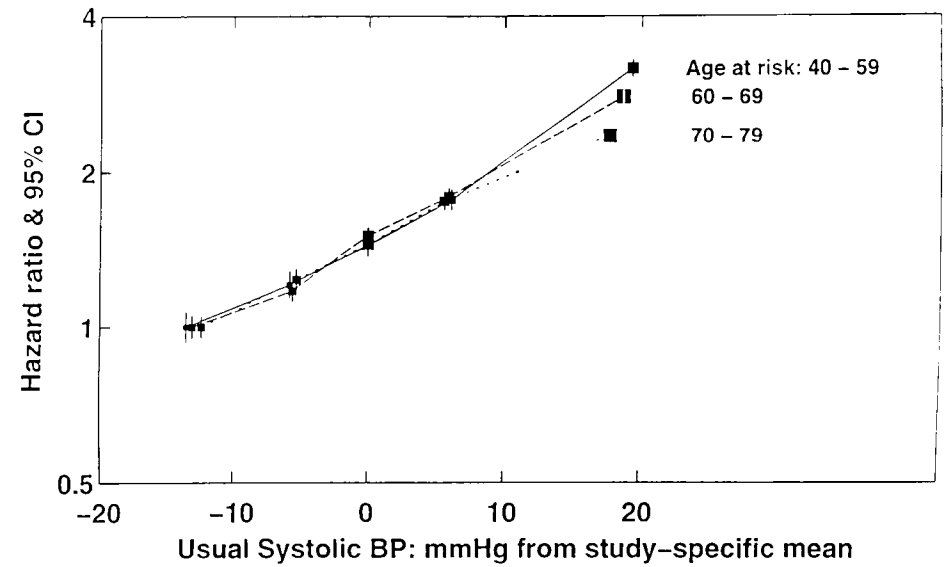
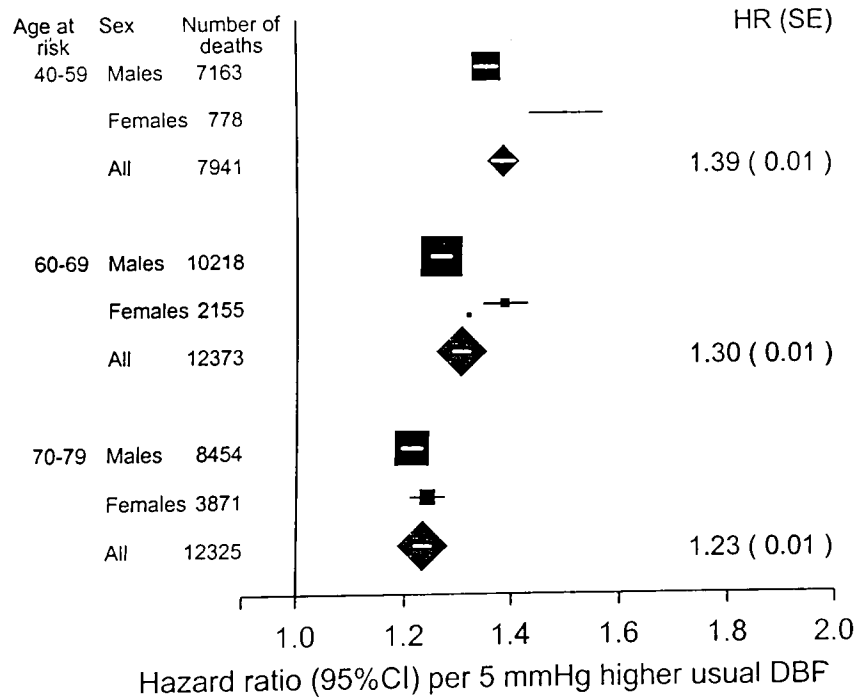


Figure 3(b): Strength and shape of the age-and sex-specific associations between DBP and IHD risk

Ischaemic heart disease

(Total: 32639 deaths)



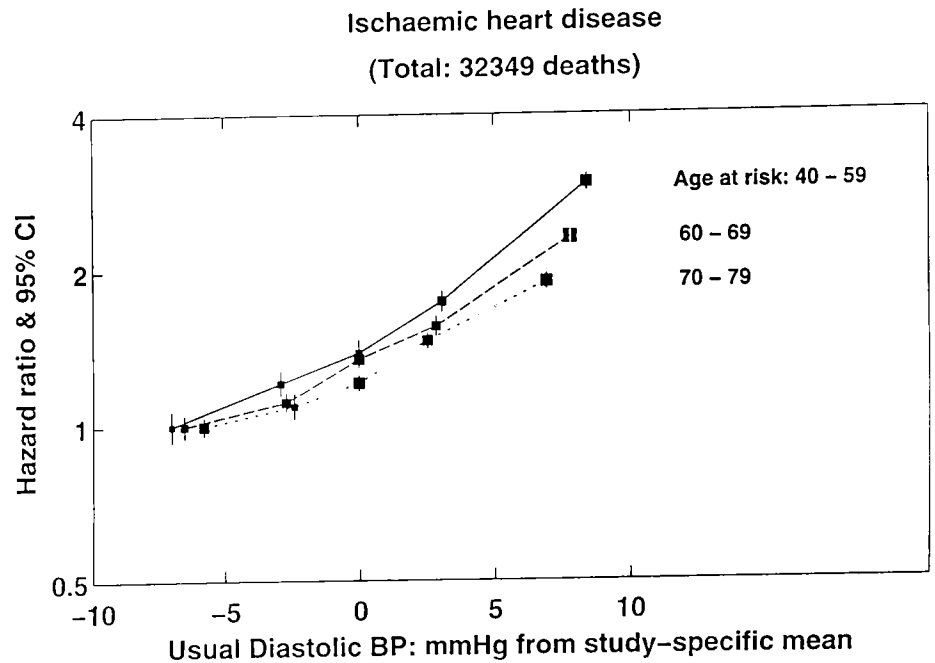


Figure 4: Proportional stroke risks by
(a) baseline SBP, (b) usual SBP, (c) baseline DBP and (d) usual DBP
All cerebrovascular disease
(Total: 13229 deaths)

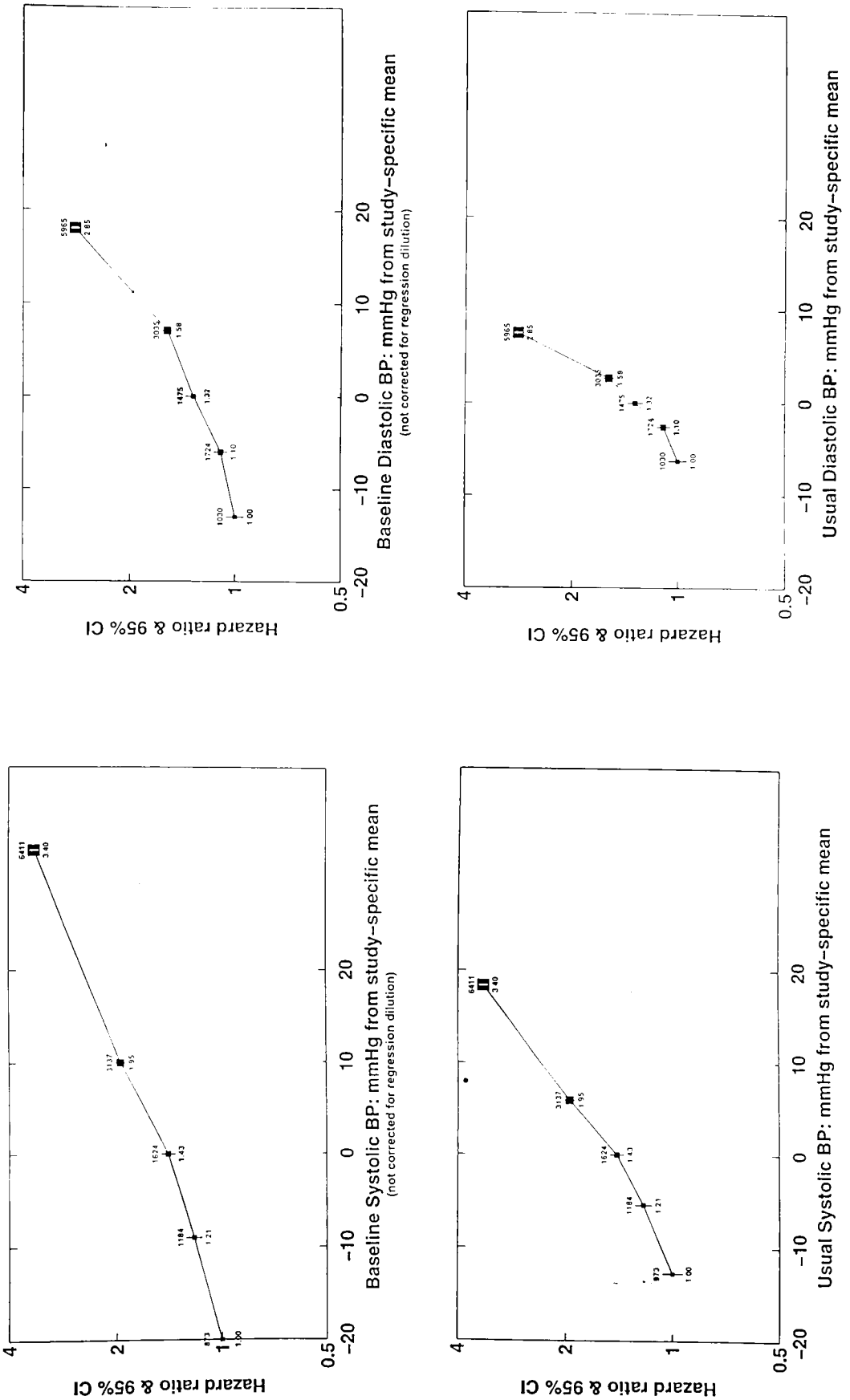
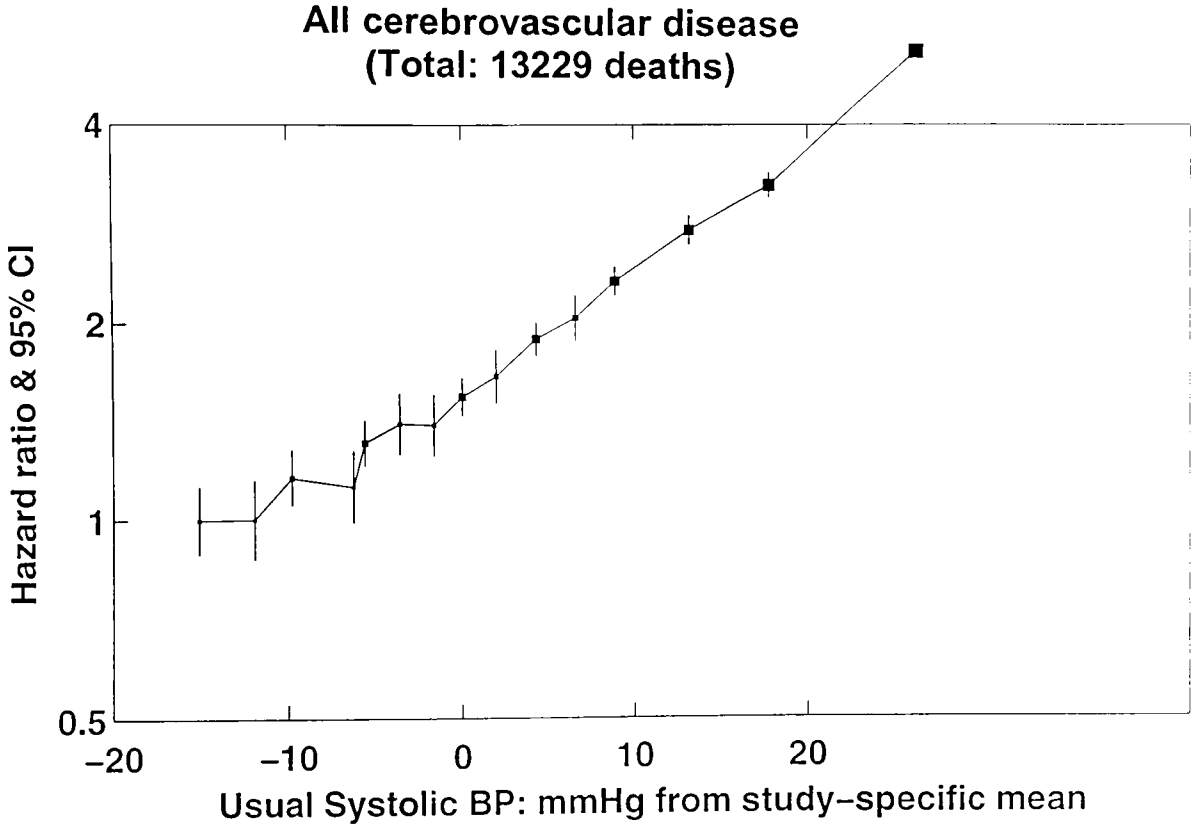


Figure 5: Proportional stroke risk throughout the entire range of (a) usual SBP and (b) usual DBP



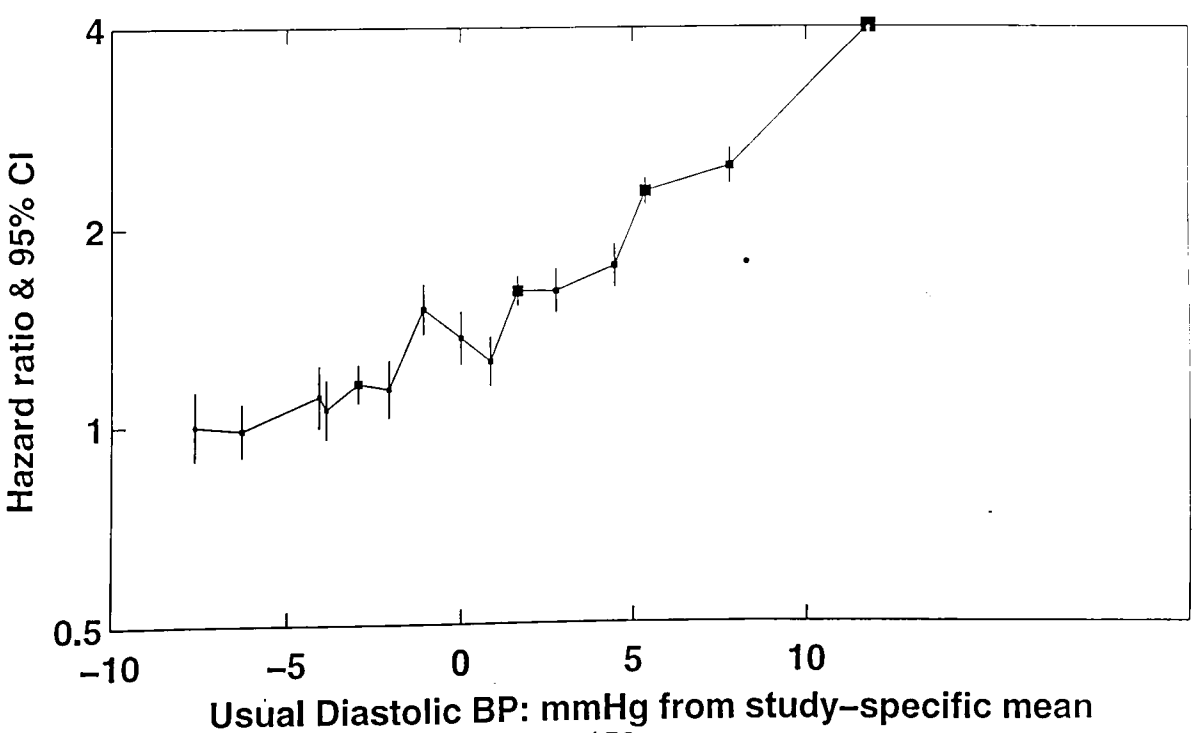
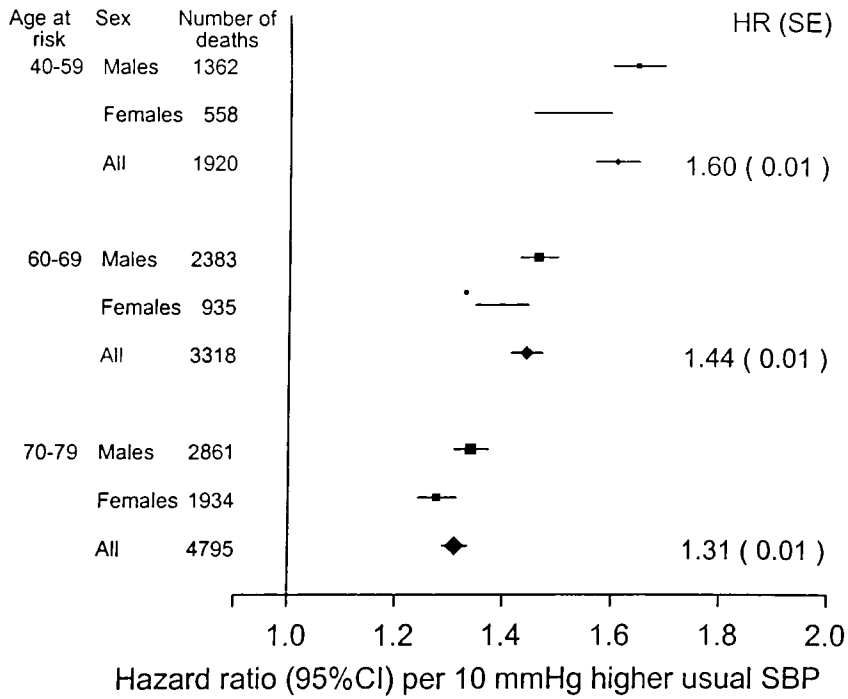


Figure 6(a): Strength and shape of the age-and sex-specific associations between SBP and stroke risk

All cerebrovascular disease
(Total: 10033 deaths)



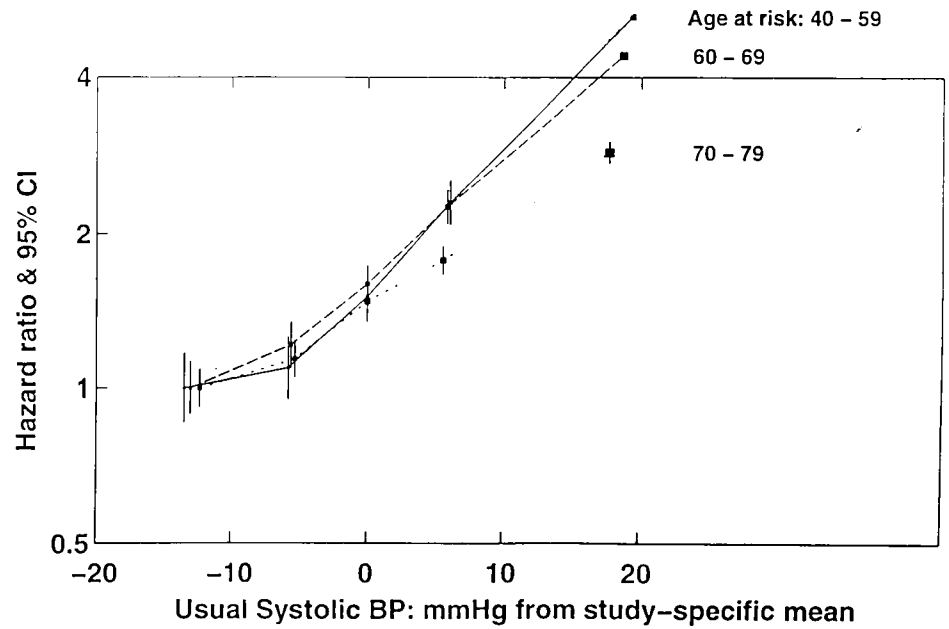
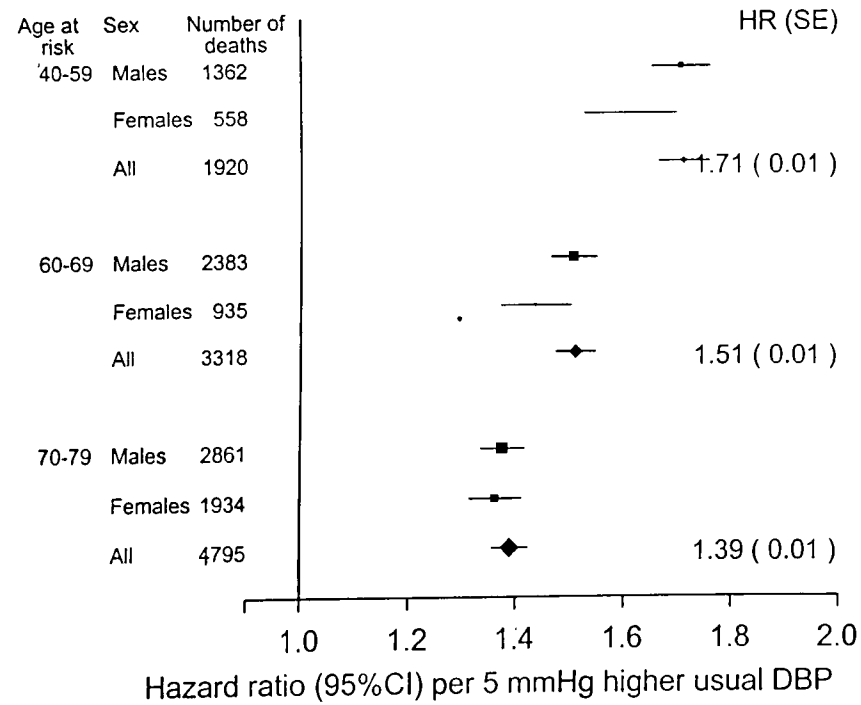


Figure 6(b): Strength and shape of the age-and sex-specific associations between DBP and stroke risk

All cerebrovascular disease
(Total: 10033 deaths)



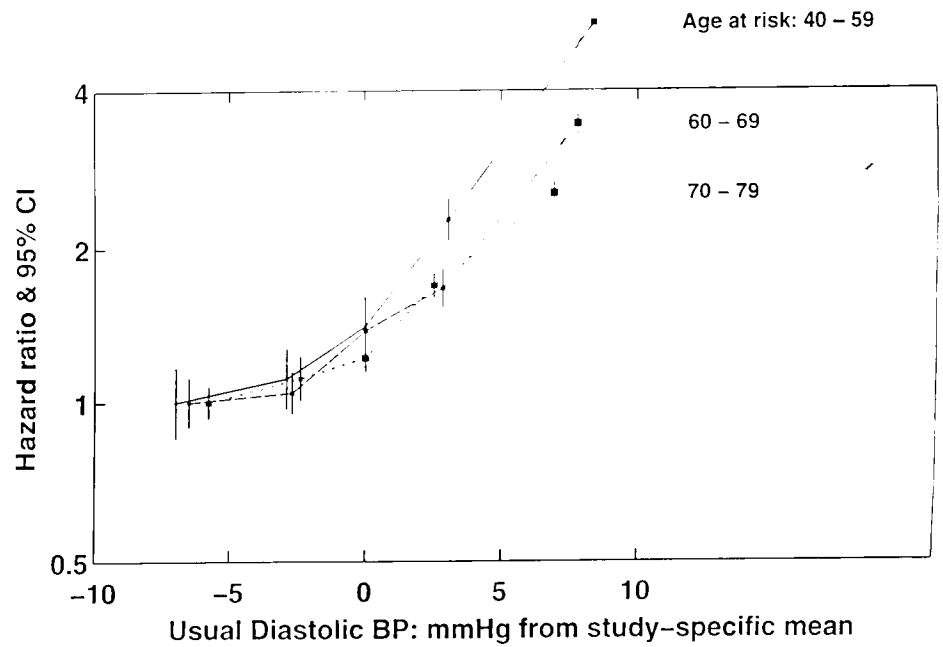


Figure 7(a): Strength of the study-specific associations between *usual* SBP and IHD risk, with study- and duration-specific corrections for regression dilution

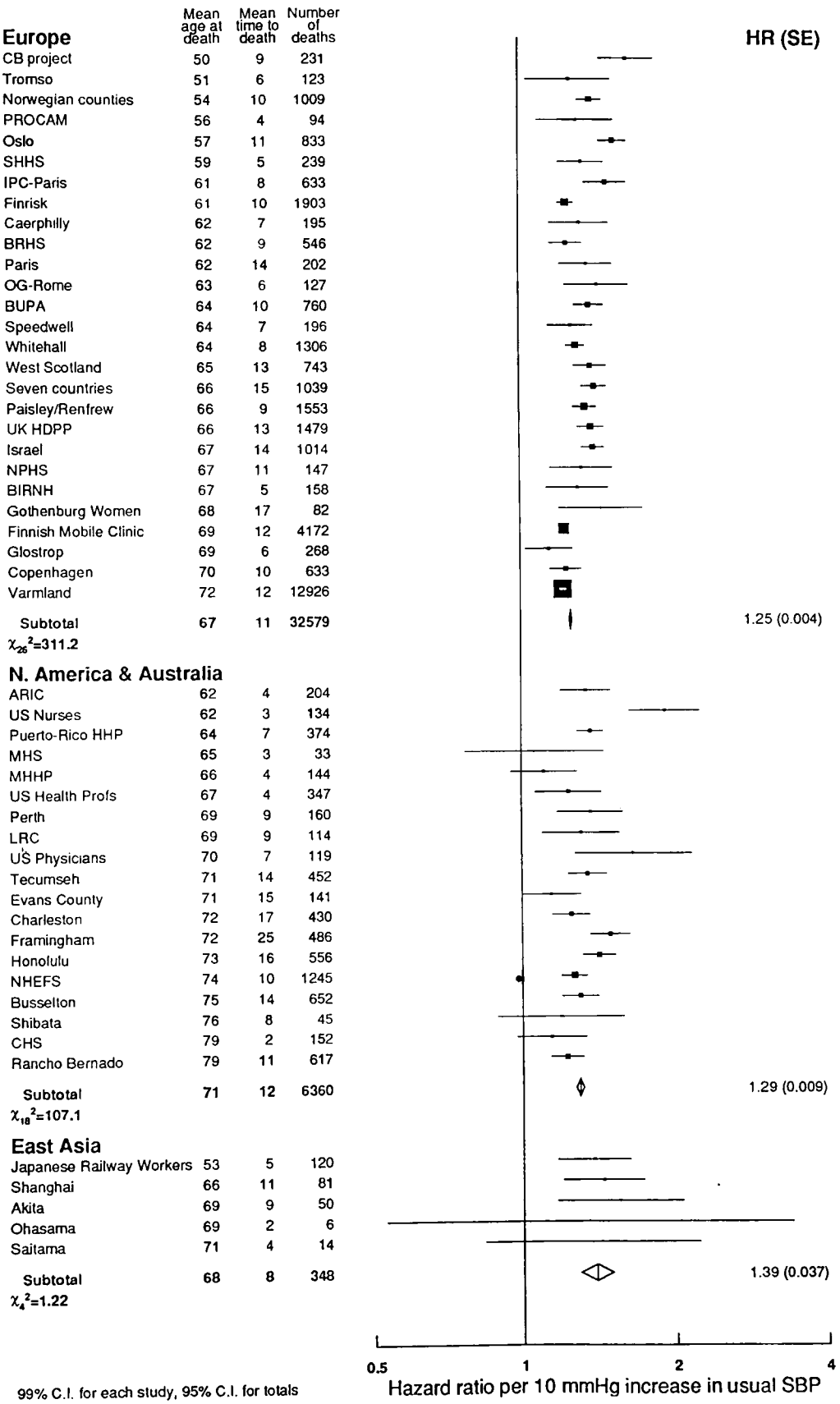


Figure 7(b): Strength of the study-specific associations between *usual* DBP and IHD risk, with study- and duration-specific corrections for regression dilution

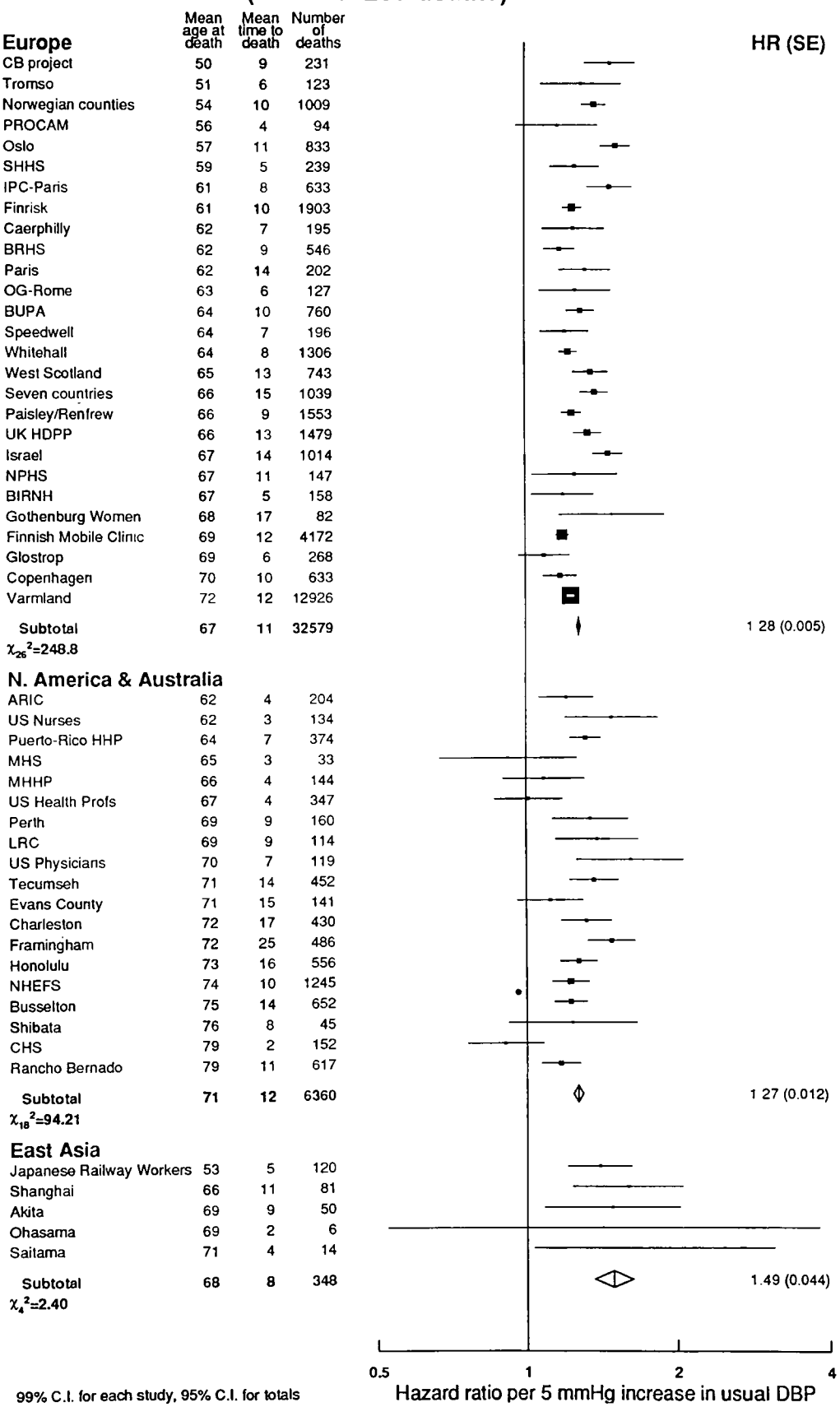


Figure 8(a): Strength of the study-specific associations between *usual* SBP and stroke risk, with study- and duration-specific corrections for regression dilution

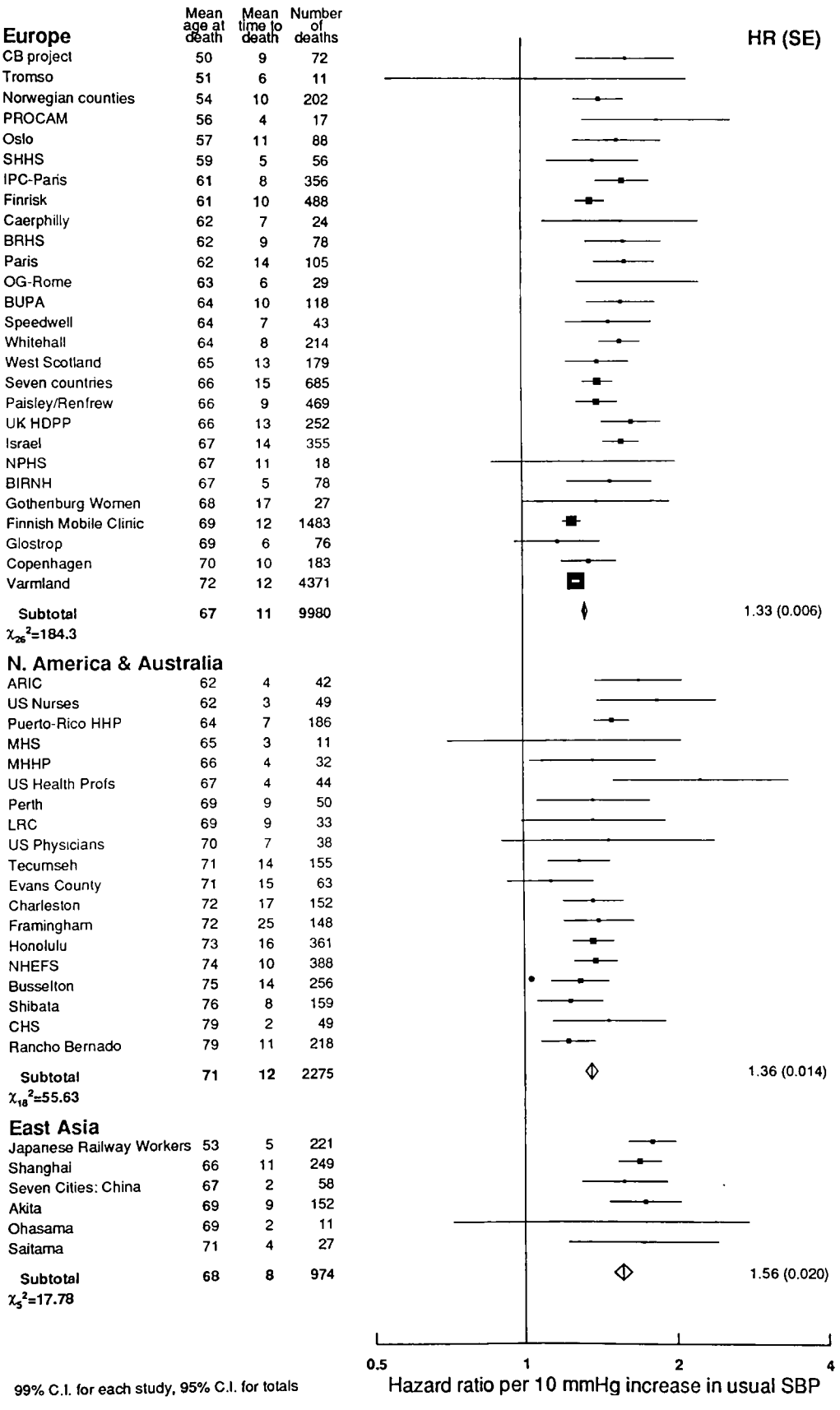


Figure 8(b): Strength of the study-specific associations between *usual* DBP and stroke risk, with study- and duration-specific corrections for regression dilution

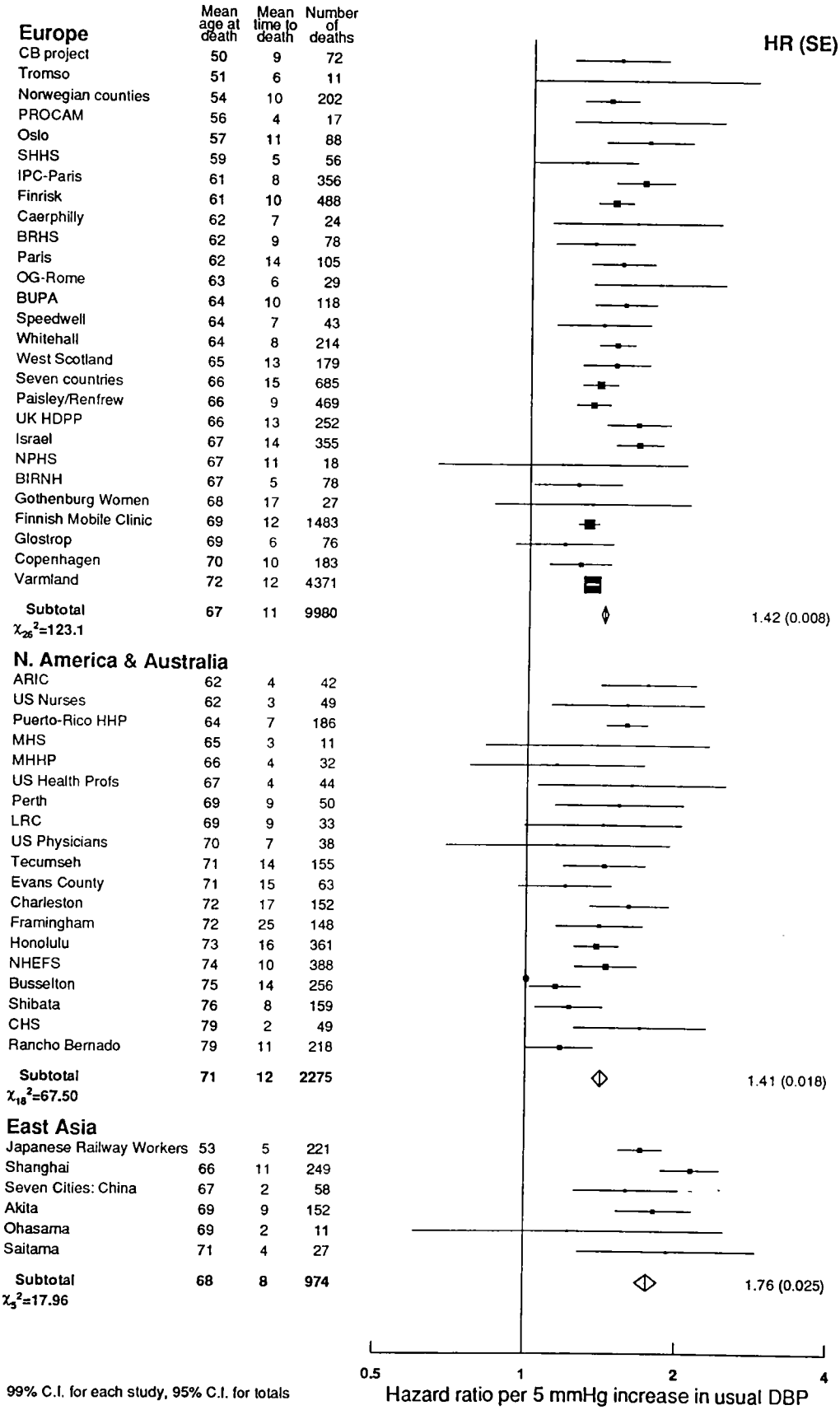


Figure 9: Strength of the age-and region-specific associations between usual SBP and stroke risk

All cerebrovascular disease (Total: 10033 deaths)

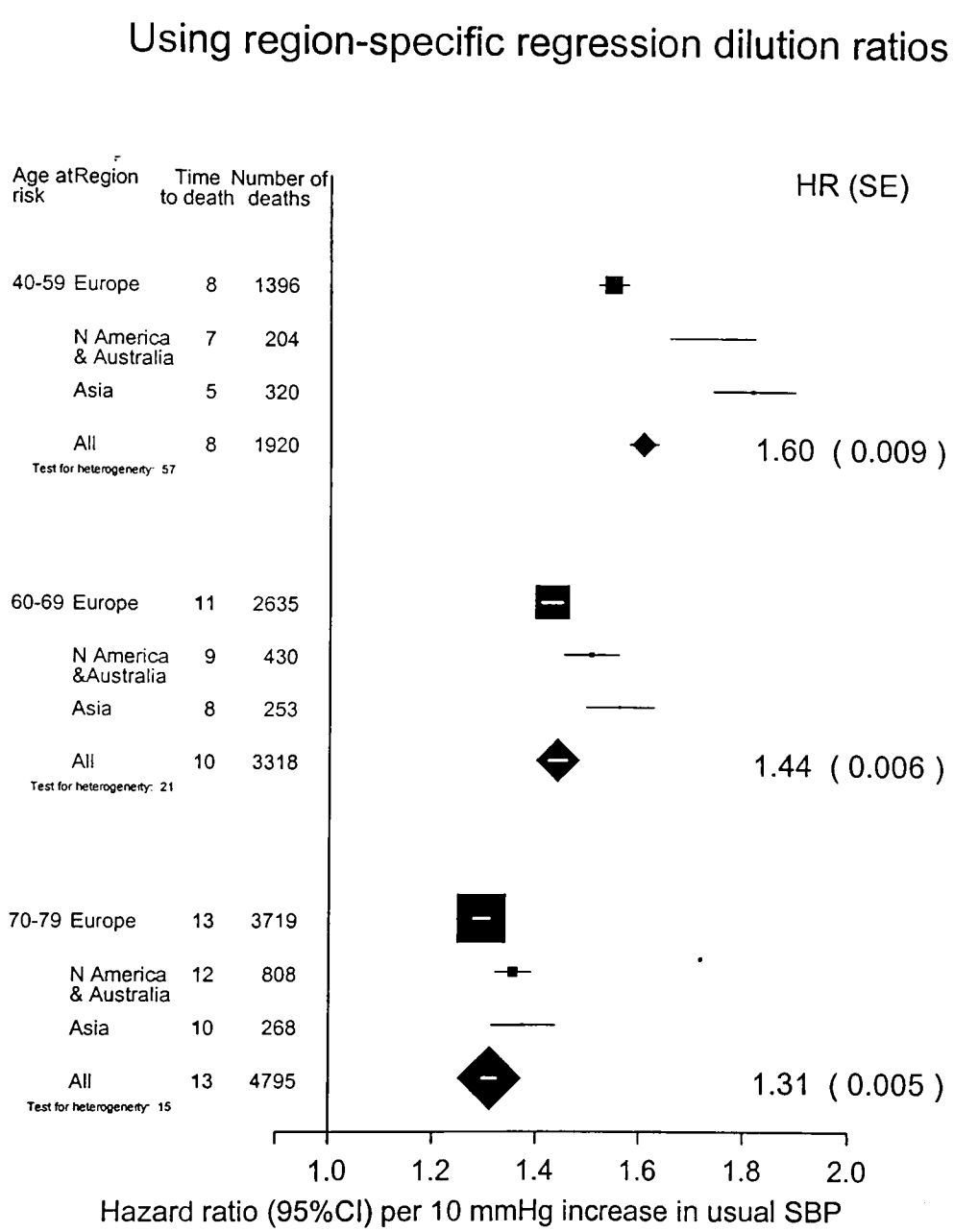


Table 1(a): Strength of the associations of *usual* SBP with various causes of death

log hazard ratios per 10 mmHg higher *usual** SBP

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**64,002**	**0.243**	**0.003**	**8770.5**
IHD	**39,287**	**0.230**	**0.003**	**4736.5**
Acute MI	**14,421**	**0.285**	**0.006**	**2365.2**
Old MI	**6,479**	**0.247**	**0.009**	**753.9**
All CVA	**13,229**	**0.297**	**0.005**	**3015.3**
Subarachnoid haem	**974**	**0.326**	**0.022**	**227.8**
Haem stroke	**2,896**	**0.337**	**0.011**	**907.0**
Isch stroke	**2,545**	**0.273**	**0.012**	**498.2**
Unclassified stroke	**6,770**	**0.285**	**0.008**	**1407.0**
Other vascular	**11,486**	**0.220**	**0.006**	**1249.3**
Atherosclerosis	**1,657**	**0.163**	**0.015**	**114.0**
Heart Failure	**931**	**0.266**	**0.024**	**126.5**
Aortic aneurysm	**813**	**0.266**	**0.025**	**110.4**
Chronic rheum. heart disease	**368**	**0.106**	**0.040**	**7.0**
All cancer	**29,227**	**0.033**	**0.005**	**47.9**
Lung	6,473	-0.012	0.011	1.3
Colorectal	**2,883**	**0.057**	**0.015**	**13.8**
Stomach	2,163	0.014	0.018	0.6
Prostate	1,383	0.043	0.022	3.8
Breast	1,341	0.018	0.024	0.6
Pancreatic	**1,297**	**0.066**	**0.023**	**8.1**
Leukaemia	779	-0.009	0.031	0.1
Brain	688	-0.060	0.036	2.8
Bladder	678	0.045	0.031	2.0
Renal	**606**	**0.123**	**0.033**	**13.9**
Ovarian	495	0.011	0.038	0.1
Melanoma	312	0.004	0.054	0.0
"Alcohol related" cancers				
Upper aerodigestive	**1,503**	**0.191**	**0.020**	**92.7**
Oesophagus	**776**	**0.179**	**0.028**	**42.1**
Oral (inc. larynx)	**472**	**0.243**	**0.035**	**48.8**
Larynx	255	0.129	0.050	6.6
Liver	**660**	**0.140**	**0.031**	**20.4**
COPD	**2,581**	**-0.055**	**0.016**	**12.0**
Suicide & trauma	2,057	0.025	0.018	1.8
Pneumonia	**1,502**	**0.129**	**0.019**	**45.4**
Liver	**1,467**	**0.274**	**0.019**	**203.1**
Renal	**914**	**0.357**	**0.021**	**289.8**
Diabetes	**747**	**0.372**	**0.024**	**241.9**
Peptic ulcer	**376**	**0.184**	**0.038**	**22.9**
Dementia	271	0.005	0.049	0.0
TB	**236**	**-0.204**	**0.057**	**12.8**
Septicaemia	**224**	**0.194**	**0.049**	**15.6**
Parkinsons	200	-0.003	0.057	0.0
All except vascular, renal, liver	**68,757**	**0.061**	**0.003**	**435.9**
All causes	**136,466**	**0.161**	**0.002**	**7038.9**

* Mean time to death=11 years. Therefore, RDR=0 61

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 1(b): Strength of the associations of *usual* DBP with various causes of death

log hazard ratios per 5 mmHg higher *usual** DBP

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**64,002**	**0.269**	**0.003**	**6943.3**
IHD	**39,329**	**0.246**	**0.004**	**3510.9**
Acute MI	**14,421**	**0.288**	**0.007**	**1743.2**
Old MI	**6,479**	**0.229**	**0.011**	**436.2**
All CVA	**13,229**	**0.360**	**0.007**	**2814.1**
Subarachnoid haem	**974**	**0.382**	**0.026**	**213.7**
Haem stroke	**2,896**	**0.416**	**0.014**	**887.5**
Isch stroke	**2,545**	**0.324**	**0.016**	**425.8**
Unclassified stroke	**6,770**	**0.343**	**0.010**	**1301.8**
Other vascular	**11,486**	**0.242**	**0.008**	**976.2**
Atherosclerosis	**1,657**	**0.171**	**0.020**	**73.8**
Heart Failure	**931**	**0.232**	**0.029**	**65.1**
Aortic aneurysm	**813**	**0.402**	**0.028**	**211.1**
Chronic rheum. heart disease	368	**-0.154**	0.050	9.5
All cancer	**29,227**	**0.010**	0.005	3.7
Lung ca	**6,473**	**-0.080**	**0.012**	**45.5**
Colorectal ca	**2,883**	**0.067**	**0.017**	**14.9**
Stomach ca	2,163	-0.012	0.020	0.4
Prostate ca	1,383	0.044	0.025	3.0
Breast ca	1,341	0.039	0.027	2.1
Pancreatic ca	**1,297**	**0.078**	**0.026**	**8.9**
Leukaemia	779	-0.041	0.035	1.4
Brain ca	688	-0.008	0.037	0.0
Bladder ca	678	-0.047	0.036	1.7
Renal ca	**606**	**0.126**	**0.037**	**11.5**
Ovarian ca	495	-0.010	0.045	0.0
Melanoma	312	0.042	0.057	0.5
"Alcohol related" cancers				
Upper aerodigestive	**1,503**	**0.135**	**0.023**	**34.7**
Oesophagus ca.	**776**	**0.125**	**0.032**	**15.5**
Oral ca	**472**	**0.182**	**0.041**	**19.8**
Larynx ca	255	0.081	0.056	2.1
Liver ca	660	0.091	0.036	6.5
COPD	2,581	-0.039	0.018	4.5
Suicide & trauma	2,057	0.017	0.020	0.7
Pneumonia	**1,502**	**0.065**	**0.023**	**8.0**
Liver	**1,467**	**0.265**	**0.022**	**140.8**
Renal	**914**	**0.415**	**0.026**	**257.9**
Diabetes	**747**	**0.306**	**0.031**	**99.2**
Peptic ulcer	**376**	**0.141**	**0.046**	**9.4**
Dementia	271	0.044	0.058	0.6
TB	**236**	**-0.256**	**0.063**	**16.6**
Septicaemia	**224**	**0.207**	**0.061**	**11.4**
Parkinsons	200	0.088	0.066	1.8
All except vascular, renal, liver	**68,757**	**0.053**	**0.003**	**238.7**
All causes	**136,466**	**0.167**	**0.002**	**5125.0**

* Mean time to death=11 years. Therefore, RDR=0.46

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 1(c): Strength of the associations of *usual* ABP with various causes of death

log hazard ratios per 10 mmHg higher *usual** ABP

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**64,002**	**0.375**	**0.004**	**9584.6**
IHD	**39,287**	**0.352**	**0.005**	**5067.2**
Acute MI	**14,421**	**0.421**	**0.008**	**2494.2**
Old MI	**6,479**	**0.359**	**0.013**	**741.4**
All CVA	**13,229**	**0.471**	**0.008**	**3471.9**
Subarachnoid haem	**974**	**0.512**	**0.031**	**265.1**
Haem stroke	**2,896**	**0.537**	**0.017**	**1053.8**
Isch stroke	**2,545**	**0.440**	**0.018**	**569.0**
Unclassified stroke	**6,770**	**0.448**	**0.011**	**1604.3**
Other vascular	**11,486**	**0.339**	**0.009**	**1361.0**
Atherosclerosis	**1,657**	**0.262**	**0.024**	**123.9**
Heart Failure	**931**	**0.379**	**0.035**	**119.3**
Aortic aneurysm	**813**	**0.459**	**0.035**	**167.4**
Chronic rheum. heart disease	368	0.035	0.060	0.3
All cancer	**29,227**	**0.038**	**0.007**	**30.1**
Lung	**6,473**	**-0.053**	**0.015**	**12.3**
Colorectal	**2,883**	**0.089**	**0.022**	**16.6**
Stomach	2,163	0.007	0.025	0.1
Prostate	1,383	0.064	0.032	4.0
Breast	1,341	0.037	0.034	1.2
Pancreatic	**1,297**	**0.103**	**0.033**	**9.8**
Leukaemia	779	-0.030	0.044	0.4
Brain	688	-0.060	0.050	1.4
Bladder	678	0.020	0.045	0.2
Renal	**606**	**0.183**	**0.047**	**15.1**
Ovarian	495	0.006	0.056	0.0
Melanoma	312	0.027	0.075	0.1
"Alcohol related" cancers				
Upper aerodigestive	**1,503**	**0.249**	**0.028**	**77.8**
Oesophagus	**776**	**0.233**	**0.039**	**35.3**
Oral (inc. larynx)	**472**	**0.321**	**0.050**	**41.9**
Larynx	255	0.161	0.070	5.3
Liver	**660**	**0.179**	**0.044**	**16.4**
COPD	**2,581**	**-0.075**	**0.023**	**10.4**
Suicide & trauma	2,057	0.035	0.027	1.7
Pneumonia	**1,502**	**0.162**	**0.028**	**33.3**
Liver	**1,467**	**0.394**	**0.027**	**207.0**
Renal	**914**	**0.557**	**0.031**	**329.5**
Diabetes	**747**	**0.518**	**0.035**	**215.3**
Peptic ulcer	**376**	**0.251**	**0.056**	**20.2**
Dementia	271	0.026	0.071	0.1
TB	**236**	**-0.329**	**0.082**	**16.2**
Septicaemia	**224**	**0.293**	**0.072**	**16.4**
Parkinsons	200	0.039	0.082	0.2
All except vascular, renal, liver	**68,757**	**0.087**	**0.004**	**421.7**
All causes	**136,466**	**0.243**	**0.003**	**7469.2**

* Mean time to death=11 years. Therefore, RDR=0.58

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 2(a): Strength of the associations of usual SBP with various causes of death, after removing deaths within 5 years of the BPmeasurement

log hazard ratios per 10 mmHg higher *usual** SBP

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**52,443**	**0.242**	**0.003**	**7075.4**
IHD	**32,241**	**0.235**	**0.004**	**4036.5**
Acute MI	**11,207**	**0.288**	**0.007**	**1842.5**
Old MI	**5,123**	**0.268**	**0.010**	**724.7**
All CVA	**10,947**	**0.280**	**0.006**	**2155.5**
Subarachnoid haem	**710**	**0.281**	**0.026**	**117.5**
Haem stroke	**2,124**	**0.312**	**0.013**	**540.4**
Isch stroke	**2,085**	**0.255**	**0.014**	**345.9**
Unclassified stroke	**6,028**	**0.277**	**0.008**	**1167.7**
Other vascular	**9,255**	**0.216**	**0.007**	**955.4**
Atherosclerosis	**1,389**	**0.171**	**0.017**	**102.9**
Heart Failure	**746**	**0.258**	**0.026**	**96.3**
Aortic aneurysm	**687**	**0.257**	**0.028**	**85.4**
Chronic rheum. heart disease	242	0.058	0.051	1.3
All cancer	**22,944**	**0.042**	**0.005**	**60.9**
Lung	5,073	-0.013	0.012	1.1
Colorectal	**2,246**	**0.074**	**0.017**	**18.3**
Stomach	1,657	0.041	0.020	4.2
Prostate	1,211	0.046	0.024	3.7
Breast	906	0.070	0.027	6.6
Pancreatic	972	0.042	0.027	2.4
Leukaemia	600	0.026	0.035	0.5
Brain	484	-0.067	0.043	2.4
Bladder	566	0.037	0.035	1.2
Renal	**460**	**0.113**	**0.038**	**8.7**
Ovarian	343	0.033	0.045	0.5
Melanoma	239	0.023	0.060	0.1
"Alcohol related" cancers				
Upper aerodigestive	**1,204**	**0.212**	**0.022**	**93.9**
Oesophagus	**620**	**0.215**	**0.030**	**51.1**
Oral (inc. larynx)	**375**	**0.241**	**0.039**	**38.0**
Larynx	**209**	**0.147**	**0.055**	**7.2**
Liver	**515**	**0.141**	**0.036**	**15.7**
COPD	**2,171**	**-0.058**	**0.018**	**10.7**
Suicide & trauma	1,510	0.026	0.022	1.4
Pneumonia	**1,313**	**0.120**	**0.021**	**34.0**
Liver	**1,132**	**0.282**	**0.022**	**163.0**
Renal	**729**	**0.301**	**0.024**	**151.5**
Diabetes	**598**	**0.377**	**0.027**	**199.0**
Peptic ulcer	**288**	**0.214**	**0.044**	**24.2**
Dementia	255	0.019	0.050	0.2
TB	168	-0.160	0.068	5.6
Septicaemia	**167**	**0.214**	**0.055**	**15.2**
Parkinsons	186	0.025	0.059	0.2
All except vascular, renal, liver	**55,394**	**0.067**	**0.003**	**426.0**
All causes	**110,655**	**0.163**	**0.002**	**5820.1**

* Mean time to death=11 years. Therefore, RDR=0.61

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 2(b): Strength of the associations of usual DBP with various causes of death, after removing deaths within 5 years of the BPmeasurement

log hazard ratios per 5 mmHg higher usual* DBP

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**52,443**	**0.267**	**0.004**	**5540.5**
IHD	**32,241**	**0.250**	**0.005**	**2957.4**
Acute MI	**11,207**	**0.289**	**0.008**	**1333.8**
Old MI	**5,123**	**0.253**	**0.012**	**434.9**
All CVA	**10,947**	**0.333**	**0.008**	**1922.5**
Subarachnoid haem	**710**	**0.330**	**0.031**	**111.6**
Haem stroke	**2,124**	**0.376**	**0.017**	**500.5**
Isch stroke	**2,085**	**0.299**	**0.018**	**285.8**
Unclassified stroke	**6,028**	**0.328**	**0.010**	**1039.8**
Other vascular	**9,255**	**0.242**	**0.009**	**779.8**
Atherosclerosis	**1,389**	**0.188**	**0.022**	**73.3**
Heart Failure	**746**	**0.245**	**0.032**	**58.2**
Aortic aneurysm	**687**	**0.368**	**0.031**	**144.6**
Chronic rheum. heart disease	242	-0.141	0.062	5.2
All cancer	**22,944**	**0.020**	**0.006**	**11.0**
Lung ca	**5,073**	**-0.084**	**0.013**	**39.4**
Colorectal ca	**2,246**	**0.090**	**0.019**	**21.4**
Stomach ca	1,657	0.017	0.023	0.5
Prostate ca	1,211	0.047	0.027	3.0
Breast ca	906	0.074	0.032	5.4
Pancreatic ca	972	0.069	0.030	5.3
Leukaemia	600	0.002	0.039	0.0
Brain ca	484	-0.032	0.045	0.5
Bladder ca	566	-0.038	0.040	0.9
Renal ca	**460**	**0.135**	**0.042**	**10.0**
Ovarian ca	343	0.022	0.052	0.2
Melanoma	239	0.031	0.064	0.2
"Alcohol related" cancers				
Upper aerodigestive	**1,204**	**0.147**	**0.025**	**33.3**
Oesophagus ca.	**620**	**0.150**	**0.035**	**18.4**
Oral ca	**375**	**0.171**	**0.046**	**13.8**
Larynx ca	209	0.093	0.062	2.2
Liver ca	**515**	**0.117**	**0.040**	**8.4**
COPD	2,171	-0.051	0.020	6.4
Suicide & trauma	1,510	0.013	0.023	0.3
Pneumonia	1,313	0.051	0.025	4.2
Liver	**1,132**	**0.304**	**0.025**	**146.4**
Renal	**729**	**0.369**	**0.030**	**153.2**
Diabetes	**598**	**0.340**	**0.034**	**101.0**
Peptic ulcer	**288**	**0.181**	**0.052**	**12.0**
Dementia	255	0.062	0.059	1.1
TB	**168**	**-0.285**	**0.078**	**13.4**
Septicaemia	167	0.157	0.072	4.8
Parkinsons	186	0.106	0.068	2.4
All except vascular, renal, liver	**55,394**	**0.060**	**0.004**	**245.2**
All causes	**110,655**	**0.169**	**0.003**	**4236.4**

* Mean time to death=11 years. Therefore, RDR=0.46

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 2(c): Strength of the associations of usual ABP with various causes of death, after removing deaths within 5 years of the BPmeasurement

log hazard ratios per 10 mmHg higher usual* ABP

Cause of death	No of deaths	RDR*	log(HR)	SE(log(HR))	chi-square
All vascular	**52,443**	**0.58**	**0.372**	**0.004**	**7694.6**
IHD	**32,241**	**0.58**	**0.358**	**0.005**	**4298.9**
Acute MI	**11,207**	**0.58**	**0.424**	**0.010**	**1929.9**
Old MI	**5,123**	**0.58**	**0.389**	**0.015**	**717.6**
All CVA	**10,947**	**0.58**	**0.441**	**0.009**	**2441.1**
Subarachnoid haem	**710**	**0.58**	**0.443**	**0.038**	**137.8**
Haem stroke	**2,124**	**0.58**	**0.492**	**0.020**	**616.7**
Isch stroke	**2,085**	**0.58**	**0.409**	**0.021**	**390.3**
Unclassified stroke	**6,028**	**0.58**	**0.433**	**0.012**	**1311.6**
Other vascular	**9,255**	**0.58**	**0.334**	**0.010**	**1053.2**
Atherosclerosis	**1,389**	**0.58**	**0.277**	**0.026**	**114.6**
Heart Failure	**746**	**0.58**	**0.377**	0.039	**95.4**
Aortic aneurysm	**687**	**0.58**	**0.431**	0.039	**122.8**
Chronic rheum. heart disease	242	0.58	-0.009	0.075	0.0
All cancer	**22,944**	**0.58**	**0.052**	**0.008**	**44.1**
Lung ca	**5,073**	**0.58**	**-0.056**	**0.017**	**10.7**
Colorectal ca	**2,246**	**0.58**	**0.117**	**0.025**	**22.6**
Stomach ca	1,657	0.58	0.048	0.029	2.8
Prostate ca	1,211	0.58	0.068	0.034	4.0
Breast ca	**906**	**0.58**	**0.105**	**0.039**	**7.1**
Pancreatic ca	972	0.58	0.075	0.038	3.9
Leukaemia	600	0.58	0.025	0.050	0.3
Brain ca	484	0.58	-0.079	0.060	1.7
Bladder ca	566	0.58	0.018	0.050	0.1
Renal ca	**460**	**0.58**	**0.177**	**0.054**	**10.7**
Ovarian ca	343	0.58	0.044	0.065	0.5
Melanoma	239	0.58	0.039	0.084	0.2
"Alcohol related" cancers					
Upper aerodigestive	**1,204**	**0.58**	**0.274**	**0.031**	**77.5**
Oesophagus ca.	**620**	**0.58**	**0.280**	**0.043**	**42.5**
Oral ca	**375**	**0.58**	**0.314**	**0.056**	**31.6**
Larynx ca	209	0.58	0.185	0.077	5.7
Liver ca	**515**	**0.58**	**0.191**	**0.050**	**14.5**
COPD	**2,171**	**0.58**	**-0.083**	**0.026**	**10.6**
Suicide & trauma	1,510	0.58	0.034	0.032	1.1
Pneumonia	**1,313**	**0.58**	**0.146**	**0.030**	**23.4**
Liver	**1,132**	**0.58**	**0.418**	**0.031**	**181.8**
Renal	**729**	**0.58**	**0.478**	**0.036**	**180.1**
Diabetes	**598**	**0.58**	**0.533**	**0.039**	**186.9**
Peptic ulcer	**288**	**0.58**	**0.301**	**0.063**	**22.5**
Dementia	255	0.58	0.049	0.073	0.5
TB	**168**	**0.58**	**-0.297**	**0.098**	**9.1**
Septicaemia	**167**	**0.58**	**0.291**	**0.082**	**12.7**
Parkinsons	186	0.58	0.075	0.084	0.8
All except vascular, renal, liver	**55,394**	**0.58**	**0.097**	**0.005**	**417.7**
All causes	**110,655**	**0.58**	**0.246**	**0.003**	**6166.3**

* Mean time to death=11 years Therefore, RDR=0.58

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 3(a): Strength of the associations of *usual* SBP with various causes of death, after removing deaths within 5 years of the blood pressure measurement

For those individuals with duration, sex, status, SBP, DBP, total cholesterol AND smoking (Y/N), alcohol (Y/N), height and weight

log hazard ratios per 10mm/Hg higher usual* SBP

		Adjusted for cohort, age and sex			Adjusted for cohort, age, sex, smoking (Y/N), alcohol (Y/N), weight (relative to height)		
Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square	log(HR)	SE(log(HR))	chi-square
All vascular	**11,956**	**0.282**	**0.006**	**1942.3**	**0.275**	**0.007**	**1714.9**
IHD	**7,088**	**0.269**	**0.008**	**1021.0**	**0.263**	**0.009**	**898.0**
Acute MI	**4,153**	**0.277**	**0.011**	**634.4**	**0.269**	**0.011**	**554.4**
Old MI	**2,409**	**0.265**	**0.015**	**331.7**	**0.260**	**0.015**	**296.3**
All CVA	**2,411**	**0.315**	**0.014**	**528.2**	**0.316**	**0.014**	**490.4**
Subarachnoid haem	**163**	**0.312**	**0.057**	**29.7**	**0.338**	**0.059**	**32.6**
Haem stroke	**453**	**0.399**	**0.031**	**160.9**	**0.390**	**0.033**	**137.9**
Isch stroke	**599**	**0.306**	**0.027**	**130.0**	**0.308**	**0.028**	**121.3**
Unclassified stroke	**1,196**	**0.290**	**0.020**	**220.3**	**0.292**	**0.020**	**207.3**
Other vascular	**2,457**	**0.282**	**0.014**	**402.9**	0.269	0 015	339.1
Atherosclerosis	**150**	**0.193**	**0.056**	**11.9**	**0.205**	**0.059**	**12.2**
Heart Failure	**372**	**0.268**	**0.037**	**53.3**	**0.261**	**0.039**	**45.2**
Aortic aneurysm	**314**	**0.243**	**0.041**	**34.7**	**0.234**	**0.043**	**29.5**
Chronic rheum. heart disea	89	0.041	0.082	0.2	0.054	0 086	0.4
All cancer	**7,823**	**0.038**	**0.009**	**17.0**	**0.049**	**0.010**	**26.2**
Lung	1,945	-0.035	0.019	3.2	0.015	0 020	0.6
Colorectal	**917**	**0.091**	**0.026**	**12.0**	**0.076**	**0.027**	**7.6**
Stomach	677	0.030	0.031	0.9	0.036	0.032	1 2
Prostate	558	0 005	0.036	0.0	0.000	0.037	0.0
Breast	316	0.075	0.044	2.9	0.057	0.046	1 5
Pancreatic	383	0.005	0.043	0.0	0.004	0.044	0.0
Leukaemia	218	0.026	0.055	0.2	0.019	0 058	0.1
Brain	169	-0.003	0.070	0.0	-0.002	0.073	0.0
Bladder	221	0.022	0.055	0.2	0.021	0.057	0.1
Renal	187	0.065	0.061	1.1	0.051	0.063	0.6
Ovarian	103	0.134	0.076	3.1	0.143	0 080	3.2
Melanoma	84	0.011	0.098	0.0	-0.007	0.102	0.0
"Alcohol related" cancers							
Upper aerodigestive	**441**	**0.145**	**0.037**	**15.4**	**0.193**	**0.038**	**26.4**
Oesophagus	**261**	**0.157**	**0.048**	**10.6**	**0.194**	**0.049**	**15.6**
Oral	137	0.134	0.065	4.3	**0.183**	**0.066**	**7.7**
Larynx	43	0.106	0.125	0.7	0.223	0.126	3.1
Liver	**172**	**0.178**	**0.060**	**8.8**	0.111	0 063	3.1
COPD	**892**	**-0.083**	0.028	**8.8**	0.022	0.028	0.6
Suicide & trauma	**379**	**0.113**	**0.044**	**6.7**	**0.144**	**0.045**	**10.3**
Pneumonia	**627**	**0.095**	**0.030**	**10.1**	**0.127**	**0.031**	**16.7**
Liver	**400**	**0.253**	**0.038**	**45.2**	**0.216**	**0.039**	**29.9**
Renal	**297**	**0.280**	**0.039**	**51.3**	**0.229**	**0.041**	**31.1**
Diabetes	**301**	**0.349**	**0.038**	**84.0**	**0.267**	**0.041**	**42.5**
Peptic ulcer	101	0.125	0.077	2 7	0.150	0.079	3.6
Dementia	147	0.006	0.064	0.0	0.053	0 066	0.6
TB	56	-0.216	0.115	3.5	-0.022	0.112	0.0
Septicaemia	**86**	**0.215**	**0.075**	**8.3**	**0.205**	**0.079**	**6.7**
Parkinsons	105	-0.046	0.079	0.3	-0.026	0.080	0 1
All except vascular, renal, liver	**16,149**	**0.072**	**0.006**	**129.9**	**0.087**	**0.007**	**178.3**
All causes	**29,725**	**0.172**	**0.004**	**1537.7**	**0.175**	**0.005**	**1492.6**

* Mean time to death=11 years. Therefore, RDR=0 61

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 3(b): Strength of the associations of usual DBP with various causes of death, after removing deaths within 5 years of the blood pressure measurement

For those individuals with duration, sex, status, SBP, DBP, total cholesterol AND smoking (Y/N), alcohol (Y/N), height and weight

log hazard ratios per 5mm/Hg higher usual* DBP

		Adjusted for cohort, age and sex			Adjusted for cohort, age, sex, smoking (Y/N), alcohol (Y/N), weight (relative to height)		
Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square	log(HR)	SE(log(HR))	chi-square
All vascular	**11,956**	**0.279**	**0.008**	**1293.5**	**0.275**	**0.008**	**1132.0**
IHD	**7,088**	**0.253**	**0.010**	**623.1**	**0.246**	**0.011**	**531.4**
Acute MI	**4,153**	**0.272**	**0.013**	**427.1**	**0.266**	**0.014**	**365.7**
Old MI	**2,409**	**0.233**	**0.018**	**169.1**	**0.224**	**0.019**	**142.0**
All CVA	**2,411**	**0.327**	**0.017**	**375.0**	**0.340**	**0.018**	**361.2**
Subarachnoid haem	**163**	**0.382**	**0.067**	**32.6**	**0.436**	**0.071**	**37.7**
Haem stroke	**453**	**0.480**	**0.036**	**175.6**	**0.493**	**0.039**	**159.5**
Isch stroke	**599**	**0.318**	**0.035**	**84.2**	**0.332**	**0.037**	**82.7**
Unclassified stroke	**1,196**	**0.264**	**0.024**	**119.1**	**0.272**	**0.026**	**113.4**
Other vascular	**2,457**	**0.304**	**0.017**	**315.2**	**0.293**	**0.018**	**263.1**
Atherosclerosis	150	0.174	0.068	6.5	0.180	0.071	6.4
Heart Failure	**372**	**0.261**	**0.046**	**32.4**	**0.254**	**0.049**	**27.2**
Aortic aneurysm	**314**	**0.396**	**0.046**	**75.1**	**0.407**	**0.049**	**68.7**
Chronic rheum heart disea	89	-0.255	0.105	5 9	-0.261	0.111	5.6
All cancer	7,823	0.004	0.011	0.2	0.028	0.011	6 1
Lung	**1,945**	**-0.125**	**0.022**	**32.2**	-0.036	0.023	2.5
Colorectal	**917**	**0.096**	**0.031**	**9.8**	**0.085**	**0.032**	**7.0**
Stomach	677	0.010	0.036	0.1	0.028	0.038	0 6
Prostate	558	0.046	0.040	1.3	0.041	0.042	1 0
Breast	316	0.000	0.054	0.0	-0.028	0.056	0.3
Pancreatic	383	0.008	0.049	0.0	0.019	0.050	0.1
Leukaemia	218	0.030	0.063	0 2	0.032	0.067	0 2
Brain	169	0.002	0.076	0.0	-0.002	0.080	0.0
Bladder	221	-0.067	0.065	1.0	-0.069	0.069	1 0
Renal	187	0.094	0.068	1.9	0.084	0.071	1 4
Ovarian	103	0.065	0.093	0.5	0.071	0.096	0.5
Melanoma	84	-0.137	0 110	1.6	-0.185	0.115	2 6
"Alcohol related" cancers							
Upper aerodigestive	441	0.102	0.043	5.5	**0.191**	**0.045**	**18.0**
Oesophagus	261	0.108	0.056	3 8	**0.181**	**0.058**	**9.8**
Oral	137	0.099	0.078	1.6	0.192	0.081	5.6
Larynx	43	0.069	0.144	0 2	0.249	0.155	2.6
Liver	172	0.150	0 071	4 5	0.047	0.075	0.4
COPD	892	-0.061	0.032	3 7	**0.122**	**0.032**	14.2
Suicide & trauma	379	0.095	0 050	3.6	**0.144**	**0.052**	7.6
Pneumonia	627	0.030	0 037	0 6	0.083	0.038	4.6
Liver	**400**	**0.285**	**0.044**	**42.9**	**0.240**	**0.047**	**26.5**
Renal	**297**	**0.345**	**0.047**	**53.0**	**0.290**	**0.051**	**32.0**
Diabetes	**301**	**0.269**	**0.049**	**30.1**	**0.142**	**0.054**	**7.0**
Peptic ulcer	101	0.081	0.094	0 7	0.151	0.098	2.4
Dementia	147	0.076	0.077	1.0	0.133	0.080	2 7
TB	**56**	**-0.368**	**0.127**	**8.4**	-0.121	0.142	0.7
Septicaemia	86	0 206	0.095	4 7	0.185	0.100	3.4
Parkinsons	105	0.019	0.093	0 0	0.050	0.097	0.3
All except vascular, renal, liver	**16,149**	**0.050**	**0.007**	**45.4**	**0.079**	**0.008**	**104.8**
All causes	**29,725**	**0.156**	**0.005**	**895.0**	**0.168**	**0.005**	**940.0**

* Mean time to death=11 years Therefore, RDR=0.46

Results with p<0 01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 3(c): Strength of the associations of usual ABP with various causes of death, after removing deaths within 5 years of the blood pressure measurement

For those individuals with duration, sex, status, SBP, DBP, total cholesteroi AND smoking (Y/N), alcohol (Y/N), height and weight

log hazard ratios per 10mm/Hg higher usual* ABP

		Adjusted for cohort, age and sex			Adjusted for cohort, age, sex, smoking (Y/N), alcohol (Y/N), weight (relative to height)		
Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square	log(HR)	SE(log(HR))	chi-square
All vascular	**11,956**	**0.416**	**0.009**	**1987.3**	**0.412**	**0.010**	**1767.2**
IHD	**7,088**	**0.391**	**0.012**	**1015.8**	**0.385**	**0.013**	**893.9**
Acute MI	**4,153**	**0.407**	**0.016**	**651.4**	**0.401**	**0.017**	**571.0**
Old MI	**2,409**	**0.379**	**0.021**	**314.2**	**0.373**	**0.022**	**278.5**
All CVA	**2,411**	**0.473**	**0.020**	**550.9**	**0.485**	**0.021**	**522.7**
Subarachnoid haem	**163**	**0.487**	**0.082**	**35.7**	**0.540**	**0.085**	**40.3**
Haem stroke	**453**	**0.624**	**0.045**	**193.7**	**0.626**	**0.048**	**171.3**
Isch stroke	**599**	**0.458**	**0.040**	**130.7**	**0.470**	**0.042**	**125.1**
Unclassified stroke	**1,196**	**0.421**	**0.029**	**213.0**	**0.432**	**0.030**	**203.5**
Other vascular	**2,457**	**0.429**	**0.021**	**435.2**	**0.413**	**0.022**	**368.6**
Atherosclerosis	**150**	**0.286**	**0.083**	**11.8**	**0.301**	**0.087**	**12.0**
Heart Failure	**372**	**0.400**	**0.055**	**53.5**	**0.395**	**0.058**	**45.9**
Aortic aneurysm	**314**	**0.433**	**0.058**	**55.0**	**0.431**	**0.062**	**48.9**
Chronic rheum. heart disea	89	-0.074	0.125	0 4	-0.062	0.131	0.2
All cancer	**7,823**	**0.040**	**0.013**	**8.8**	**0.063**	**0.014**	**20.3**
Lung	**1,945**	**-0.098**	**0.028**	**12.1**	-0.003	0.029	0.0
Colorectal	**917**	**0.137**	**0.038**	**13.1**	**0.118**	**0.040**	**8.7**
Stomach	677	0.034	0 045	0 6	0.050	0.047	1.1
Prostate	558	0.028	0.051	0 3	0.021	0.053	0.2
Breast	316	0.077	0.066	1.4	0.046	0.069	0.4
Pancreatic	383	0.008	0.061	0.0	0.014	0.063	0.0
Leukaemia	218	0 040	0.080	0.3	0.035	0.084	0.2
Brain	169	-0.002	0.099	0.0	-0.003	0.104	0.0
Bladder	221	-0.010	0.081	0.0	-0.011	0.085	0.0
Renal	187	0.111	0.087	1.6	0.093	0.091	1.1
Ovarian	103	0.167	0.112	2.2	0.180	0.117	2.3
Melanoma	84	-0 062	0.142	0.2	-0.106	0.150	0.5
"Alcohol related" cancers							
Upper aerodigestive	**441**	**0.191**	**0.053**	**12.9**	**0.284**	**0.054**	**27.1**
Oesophagus	**261**	**0.205**	**0.069**	**8.9**	**0.279**	**0.071**	**15.5**
Oral	137	0 179	0 094	3.6	**0.273**	**0.096**	**8.0**
Larynx	43	0 141	0.182	0.6	0.350	0.185	3.6
Liver	**172**	**0.247**	**0.087**	**8.1**	0.134	0.092	2.1
COPD	892	**-0.113**	**0.040**	7.8	0 082	0.041	4.1
Suicide & trauma	379	0.158	0.063	6.3	**0.215**	**0.065**	10.9
Pneumonia	627	0.112	0.045	6.3	**0.170**	**0.046**	**13.6**
Liver	400	**0.383**	**0.054**	**51.3**	**0.331**	**0.057**	**33.5**
Renal	**297**	**0.446**	**0.057**	**60.9**	**0.372**	**0.061**	**37.2**
Diabetes	**301**	**0.479**	**0.056**	**72.3**	**0.347**	**0.062**	**31.6**
Peptic ulcer	101	0 166	0.113	2.1	0.226	0.118	3.7
Dementia	147	0.042	0 095	0.2	0.116	0.097	1.4
TB	56	-0.401	0.168	5.7	-0.077	0.167	0.2
Septicaemia	86	**0.315**	**0.111**	8.1	0.298	0.118	6.4
Parkinsons	105	-0.038	0.114	0.1	-0.003	0.117	0.0
All except vascular, renal, liver	**16,149**	**0.096**	**0.009**	**109.9**	**0.127**	**0.010**	**176.8**
All causes	**29,725**	**0.247**	**0.006**	**1505.6**	**0.259**	**0.007**	**1513.7**

* Mean time to death=11 years Therefore, RDR=0 58

Results with p<0.01 (chi-square>6 6) are In bold

All analyses are adjusted for cohort, sex and age

Chapter 8: Cholesterol and premature death

The importance of cholesterol fractions

Chapter summary

Law et al. (1994) have reported an approximately log-linear association of increasing risk of ischaemic heart disease (IHD) with increasing levels of total cholesterol throughout the range studied. The strength of this association was dependent on the age at risk, being about twice as great among individuals aged under 45 years as among those aged over 65. However, the initial report from the Prospective Studies Collaboration (1995), indicated no clear relationship between total stroke risk and total cholesterol. This might be because there are opposing effects of increasing ischaemic stroke risk and decreasing haemorrhagic stroke risk with increasing total cholesterol (Eastern Stroke and Coronary Heart Disease Collaborative Research Group, 1998); or, because there are opposing effects of high density lipoprotein (HDL) and non-HDL cholesterol; or, because any effects in middle age are diluted by different results at older ages. There have also been claims of modest inverse associations between total cholesterol and various other causes of death including cancer, respiratory disease, digestive disease, trauma and the combination of all vascular causes other than stroke or IHD (Jacobs et al., 1992). However, the extent to which these inverse associations may be explained by long-term effects of pre-clinical disease on blood lipids (that is, reverse-causality) by confounding due to other risk factors, or by unduly selective emphasis on particular findings, without consideration of the totality of evidence from all studies, is still to be resolved. Moreover, the relative importance of the separate cholesterol fractions to disease risk remains uncertain.

Using data on 64,000 vascular deaths, 30,000 cancer deaths and 42,000 deaths from other causes, this chapter investigates the age-specific relationships of total cholesterol and, where the data are available, HDL cholesterol and the remaining "non-HDL" cholesterol with IHD, stroke and various other causes of death.

1 Results: the relevance of total cholesterol, of HDL cholesterol and of non-HDL cholesterol to various causes of death

1.1 Ischaemic heart disease

For ischaemic heart disease (IHD), where a strong relationship with total cholesterol is not in doubt, the analyses have been presented in a series of steps from the least specific to the most specific. First, in all ages together without correction for the regression dilution bias (i.e. the association of the *baseline* total cholesterol measurement with IHD); then, with correction for it (i.e. the association of the *usual* cholesterol with IHD); then, with the corrected results subdivided by age (to compare the effects in middle with those in old age); and finally, separating HDL and non-HDL cholesterol.

Total cholesterol

Uncorrected analyses: Analysis of men and women irrespective of age showed a strong positive and continuous log-linear association between IHD risk and baseline measurements of total cholesterol. Analyses that were not corrected for the regression dilution bias showed that, during an average 13 years of follow-up, a difference of 3 mmol/l in the measured value at baseline of total cholesterol corresponded to a two-fold difference in IHD risk (Figure 1(a)). Although the strength

of the association between total cholesterol and IHD risk differed substantially between studies (see later), it was not appreciably altered by adjustment for blood pressure, smoking and relative weight.

Corrected analyses: The use of uncorrected baseline measurements tends to attenuate any real associations between the usual level of a risk factor and disease. As a consequence, the uncorrected relationship shown in Figure 1(a) underestimated the true relationship between the usual level of total cholesterol and IHD death by about one-third. Hence, after appropriate time-dependent correction for the regression dilution bias, a difference of only 2 mmol/l in the usual total cholesterol corresponded to a two-fold difference in IHD risk (Table 1). Furthermore, this relationship was approximately log-linear throughout the range 1 mmol/l below to 1 mmol/l above the mean cholesterol in each cohort (Figure 1(b)).

However, this relatively narrow range did not represent the true range of the total cholesterol levels available from this collection of individuals from different populations. Yet, by dividing individuals within each cohort into 15 equally-sized groups (according to increasing levels of baseline total cholesterol), and plotting the relative hazard for each of these groups against the mean usual total cholesterol level, the relationships in the extremes of the total cholesterol distributions could also be examined. Within each cohort, the relationship was positive over the range where there was plenty of evidence (of about 2 mmol/l). However, by combining information from many cohorts with different means, a wider range could be spanned in the upper part of the range. Even within that upper range, still higher total cholesterol continued to be associated with still higher risk (Figure 2). At the other extreme, though, there was some suggestion of a levelling in the risk relationship in the very

lowest fifteenth that requires further investigation (particularly of reverse-causality, in which the disease itself reduces cholesterol). If, ignoring this lowest group, the relationship is similar in different populations, this suggests a log-linear relationship from around 3 mmol/l in East Asian populations, where cholesterol levels are typically low (such as those in the Shanghai study with a mean total cholesterol of 4.2 mmol/l), up to 8.5 mmol/l in North European populations, where cholesterol levels are typically high (such as in the Oslo study with a mean total cholesterol of 6.9 mmol/l).

As with many risk factor-disease relationships, the proportional relevance of total cholesterol to IHD risk was much stronger in middle age than in old age (Figure 3). Among those dying at ages 40-59, the average age at death was 53 and within this age group there was a 64% increased risk with each 1 mmol/l higher total cholesterol. Hence, a difference of only 1.4 mmol/l in total cholesterol was associated with a two-fold difference in IHD risk at these ages. This contrasts with the 3 mmol/l estimated from the initial inappropriate analysis which combined all ages and made no correction for the regression dilution bias. However, although less steep, the association was still clearly positive even among those dying in their seventies, with each 1 mmol/l higher total cholesterol conferring 25% higher risk. Within each age group, total cholesterol was of similar relative importance for men and women. Since the absolute rates of IHD are considerably higher at older ages, total cholesterol is of substantial absolute importance to IHD risk, not only in middle age, but also in old age.

Cholesterol fractions (HDL and non-HDL)

The fundamental importance of lipid fractions to IHD risk has been reported in several studies (Gordon et al., 1977; Stampfer et al., 1991), but, due to lack of numbers in most individual studies, their age-specific relevance could not be assessed reliably. Data from not only total cholesterol, but also HDL measurements made at a baseline visit were available for 180,000 individuals in the PSC, of whom 14,000 died during follow-up. 5,600 of these deaths were caused by vascular diseases, of which 3,500 were attributed to IHD.

Taking all ages together, 1 mmol/litre higher usual non-HDL cholesterol corresponded to 63% (SD: 0.02) higher risk of IHD (Table 2). However, most relevant is the age-specific analysis, and this is presented in Figure 4. The proportional relevance of non-HDL cholesterol was greatest in the younger age group (40-59), where a difference of 1.3 mmol/l in usual non-HDL cholesterol corresponded to a two-fold difference in IHD. However, even at older ages the relationships were strong. A difference of 1.6 mmol/l in usual non-HDL cholesterol corresponded to a two-fold difference in IHD mortality at ages 60-69 and, although the data were more sparse, the relationship appeared to be about as strong at 70-79 as at 60-69. HDL cholesterol, by contrast, showed similarly strong relationships at all ages, albeit in the opposite direction. Thus, HDL cholesterol was inversely and continuously associated with IHD risk, with a difference of 0.5 mmol/l in the usual HDL cholesterol corresponding to a two-fold difference in risk (Figure 5, Table 3). For both non-HDL and HDL cholesterol, these associations were similar for men and women (Figures 4 and 5). Further analyses are required to investigate this apparent inconsistency of

decreasing relevance of total cholesterol and non-HDL cholesterol to IHD risk, but no such trend for HDL-cholesterol.

The following text-table shows how inappropriate analyses of total cholesterol (rather than the cholesterol fractions) combining all ages (instead of being age-specific) and making no correction for the regression dilution bias, misrepresented the important and opposing effects of HDL and non-HDL cholesterol as causes of IHD, particularly in middle age: -

Difference in cholesterol (mmol/l) associated with halving of risk of *premature* death from IHD (N~40,000) estimated from each analysis

	mmol/l
Baseline total cholesterol, **all ages** (i.e. without correction for regression dilution)	**3.0** ↓
Usual total cholesterol, **all ages** (i.e. with correction for regression dilution)	2.0 ↓
Usual non-HDL cholesterol, **all ages**	1.6 ↓
Usual non-HDL cholesterol, **age 60-79** (N~2,300) (i.e. age-specific, with correction for regression dilution)	**1.6** ↓
Usual non-HDL cholesterol, **age 40-59** (N~900)	**1.3** ↓
Usual HDL **cholesterol**, all ages (N~3,200)*	**0.5** ↑

* Note: the relevance of HDL cholesterol is similar at all ages

Hence, in assessing the role of cholesterol in IHD mortality, inclusion of a negative association with HDL cholesterol, neglect of the regression dilution bias and lack of age-specific analyses led to serious underestimation of the importance of blood lipids, particularly in middle age.

For example, if favourable changes of about 0.25 mmol/l in usual HDL cholesterol and 0.75 mmol/l in usual non-HDL cholesterol could be made, the uncorrected analyses in Figure 1(a) would misleadingly suggest that the resulting decrease of

approximately 0.5 mmol/l in *total* cholesterol would reduce IHD risk by only 11%. Yet, the corrected analyses in Figures 4 and 5 suggest that it would reduce the risk by more than half (55%) at ages 40-59, and by about half at ages 60-69 and 70-79. Therefore, as causes of premature death in early middle age, blood lipids are about five times more important than suggested by the inappropriate uncorrected analyses of IHD and total cholesterol combining all ages.

1.2 Cholesterol and stroke risk

Total cholesterol

After adjustment for study, sex and age, there was no overall association between total cholesterol and total stroke (Table 1, Figure 6), although there was substantial heterogeneity between the individual study estimates (see later). Previous analyses have observed apparent differences in the direction of the associations between total cholesterol and haemorrhagic stroke and ischaemic stroke, although these trends were not definitive (Eastern Stroke and Coronary Heart Disease Collaborative Research Group, 1998). Similar trends, which were also not definitive, were observed in the present analyses (Figure 7). Age-specific analyses of these data suggested that there may be an increased risk of ischaemic stroke with higher total cholesterol in both men and women aged under seventy (χ^2=15.7), but not beyond 70. Conversely, there was no significant association between total cholesterol and haemorrhagic stroke in middle age, and some suggestion of a small inverse association at older ages. These relationships were similar in men and women. Figure 8 shows the strength of the age- and sex-specific relative hazards of (a) total stroke, (b) ischaemic stroke, and (c) haemorrhagic stroke, associated with 1 mmol/l

higher usual total cholesterol. Although the relative hazards for total stroke are more statistically reliable, the age-specific relationships with the stroke sub-types are more clinically meaningful. There were sufficient numbers of both ischaemic and haemorrhagic stroke deaths to allow appropriately age-dependent analyses, and so the real relationships of total cholesterol with the stroke sub-types began to emerge.

Cholesterol fractions (HDL and non-HDL)

There were 974 stroke deaths among individuals with HDL-cholesterol measurements reported at baseline: 69 subarachnoid haemorrhages, 241 haemorrhagic stroke (primary intra-cerebral haemorrhages), 156 ischaemic strokes and 508 strokes of unknown aetiology. Of these, only 182 occurred in early middle age, of which 30 were attributed to subarachnoid haemorrhage, 72 to haemorrhagic stroke and 24 to ischaemic stroke. Consequently, there were too few data to assess reliably the sex-specific, and more importantly, the age-specific relationships between cholesterol fractions and stroke sub-types. Thus, it was only possible to calculate statistically reliable log-hazard ratios for total, haemorrhagic and ischaemic stroke associated with a 1 mmol/l higher non-HDL cholesterol and 0.1 mmol/l higher HDL cholesterol, which were not subdivided by age or sex. These are set out in Tables 2 and 3, respectively. The shape of these relationships is shown, along with total cholesterol in Figures 9-11. There was no evidence of any association between non-HDL cholesterol and stroke of any type (Table 2, Figures 9-11), and there was only weak evidence that higher HDL cholesterol may be associated with a slightly lower risk of total stroke (Figure 9, χ^2=4.4; p=0.04) and ischaemic stroke (Figure 10, χ^2=4.6;

p=0.03). There was no evidence for any association between HDL cholesterol and haemorrhagic stroke (Figure 11).

The cholesterol-stroke relationships at all ages and in early middle age have been summarised below. To avoid undue emphasis on apparently spectacular, but statistically insignificant results, only those relationships significant at the 5% level have been shown.

% (SE) change in stroke risk for given higher usual levels of total, HDL and non-HDL cholesterol

Cause of death	Among all 1.1 million[1] N	total cholesterol (mmol/l)	Among 178,000 with HDL measured N	HDL cholesterol (0.1 mmol/l)	non-HDL cholesterol (mmol/l)
<u>All ages</u>					
Stroke	13,229		974	-3.0 (1.5)*	
Subarachnoid haemorrhage	974	-	69		
Primary intracebral haemorrhage	2,938	-6.3 (2.8)*	241		
Ischaemic	2,552	6.6 (2.7)*	156	-7.8 (3.7)*	
<u>Age at risk: 40-59 years</u>					
Stroke	1,920	13.2 (3.0)****	182	-	23.4 (9.0)*
Subarachnoid haemorrhage	471		30		-
Primary intracebral haemorrhage	660		72		
Ischaemic	246	39.9 (7.1)****	24		

-, *, **, ***, **** correspond to two-sided p-values of >0.05, ≤0.05, ≤0.01, ≤0.001, ≤0.0001

[1] Results were similar among the 178,000 with HDL measurements

There was some suggestion of a positive association between total cholesterol and ischaemic stroke, and a negative association between total cholesterol and haemorrhagic stroke. However, although no independent relationships with the cholesterol fractions were observed, this does not necessarily mean that such relationships do not exist. Despite this uniquely large data set, there were still too few deaths from strokes of known type among those with HDL cholesterol measurements to allow further statistically reliable investigation of these relationships.

Half the strokes in this data set were unclassified, and of those that were classified, many are likely to have been misclassified, having been ascertained mainly from death certificates that often do not distinguish reliably between the pathological types of stroke. Such misclassification of stroke sub-types will have caused systematic under-estimation of any differences between haemorrhagic and ischaemic strokes in their associations with cholesterol and blood pressure.

1.3 Cholesterol and non-vascular causes of death

Cholesterol and cancer

Despite decades of research, there still remains uncertainty about whether there is any independent association between blood cholesterol and the incidence of various types of cancer. The concerns that very low levels of total cholesterol may be associated with an increased risk of malignancy were not substantiated in the Shanghai study (either for all cancer or for those specific cancers where data were available), where cholesterol levels as low as 3 mmol/l were "typical" (Chen et al., 1991). In the PSC data, there was a highly significant, but shallow, negative trend for increasing risk of total cancer death with lower levels of usual total cholesterol. Each mmol/l higher total cholesterol was associated with 7% fewer cancer deaths (χ^2=75; Figure 12). This negative trend was largely due to negative associations with lung, stomach, upper aero-digestive and liver cancers (Table 1). When deaths within five years of the baseline measurement were removed to minimise the effects of reverse-causality (in which the disease itself reduces cholesterol), these associations weakened or disappeared (Table 4). This suggests that reverse-causality was responsible for at least some of the observed associations. There have been some

reports of inverse relationships with cancer of the large intestine (Delahaye et al., 1992), but there was no evidence of any such association in these data (Table 1, Figure 13).

Both non-HDL and HDL cholesterol were negatively associated with total cancer mortality ($p<0.001$; Tables 2 and 3), although the association with non-HDL cholesterol attenuated to insignificance after deaths within 5 years of the cholesterol measurement were removed from the analysis (Table 5, Figure 12). The association with HDL cholesterol, however, remained, and there was a 3% reduction in total cancer risk for every 0.1 mmol/l higher HDL cholesterol (Table 6, Figure 12). There were no significant relationships at the 1% level for HDL cholesterol with any of lung (Figure 14), stomach (Figure 15) or liver (Figure 16) cancers: Table 6.

Upper aero-digestive cancer was positively related to HDL cholesterol. Moreover, this relationship was sustained beyond 5 years, with every 0.1 mmol/l higher HDL cholesterol associated with a 12% higher risk ($\chi^2=13$; Table 3, Figure 17). The incidence of upper aero-digestive cancers is strongly related to alcohol consumption (Thun et al., 1997), and heavy drinking also causes an increase in HDL cholesterol (Castelli et al., 1977). Thus, at least some of this association must have been due to confounding with alcohol. However, of the 130,000 participants defined as "not current" drinkers using the data provided, almost half could not be defined reliably as never-, ex- or even occasional- drinkers. Ideally, these relationships should be re-assessed among lifetime abstainers. However, with only 18 upper aero-digestive cancer deaths among the 47,000 individuals with HDL cholesterol measurements reported as never drinkers (of whom a substantial proportion will have been misclassified), this was not possible.

Cholesterol and non-vascular, non-cancer mortality

In these data, there was a strong and positive relationship between total cholesterol and diabetic deaths, which was not attenuated by removal of deaths within 5 years of the baseline cholesterol measurement (Figure 18). Many of these diabetic deaths would have had a large vascular component, and this was reflected in the hazard ratio (30%(SE:5) increased risk for each 1 mmol/l higher usual total cholesterol (χ^2=32: Table 1)), which was similar to that for all vascular deaths (26%(SE:0.5) increased risk for each 1 mmol/l higher usual total cholesterol (χ^2=1904)). The association between non-HDL cholesterol and diabetic deaths was also comparable to that between non-HDL cholesterol and all vascular deaths (Table 2, Figure 18). The negative relationship between HDL cholesterol and diabetic deaths was much stronger (27%(SE:5) lower risk per 0.1 mmol/l higher HDL cholesterol (χ^2=43): Table 3, Figure 18) than that for all vascular deaths (11%(SE:0.7) lower risk per 0.1 mmol/l higher usual HDL cholesterol (χ^2=279)).

Despite previous reports to the contrary (Jacobs et al., 1992), most other non-vascular, non-cancer deaths were not associated with total cholesterol, non-HDL cholesterol or HDL cholesterol (Tables 1-3, Figure 19). Reported associations of increased suicide and violence with low levels of cholesterol, based on small numbers of events, have attracted much publicity (Kaplan et al., 1997). However, the present analyses based on over 2,000 such deaths, showed no significant associations with any of total, HDL or the remaining non-HDL cholesterol even at low levels of total and non-HDL cholesterol (Figure 20). Deaths from liver diseases were inversely related to total cholesterol, but as the liver is where most cholesterol is synthesised this could reflect reverse-causality, and the statistical significance of this

association was greatly reduced after removal of deaths within 5 years of the cholesterol measurement.

Total cholesterol was negatively associated with mortality from four major lung diseases: lung cancer, chronic obstructive pulmonary disease (COPD), pneumonia, and pulmonary tuberculosis (TB). Furthermore, the strength of these relationships was not attenuated greatly by removal of deaths within 5 years of the cholesterol measurement (Table 4). The inverse association with COPD-related death was of particular interest, because the disease is so common in Eastern populations where cholesterol levels are commonly low (and still accounts for 3 million of the 40 million adult deaths worldwide each year). Beyond 5 years of follow-up, COPD remained significantly, and negatively, related to total cholesterol, and associations with HDL and non-HDL cholesterol were both opposite to those seen for IHD. For non-HDL cholesterol, there was a 40% lower risk of COPD-related death for each 1 mmol/l higher usual non-HDL cholesterol (χ^2=26: Table 5, Figure 21). COPD (together with upper aero-digestive cancer: see above) was also the only cause of death that sustained a clearly positive relationship with HDL cholesterol beyond 5 years of follow-up, with each 0.1 mmol/l higher HDL associated with an 8% higher risk of COPD-related death (χ^2=22: Table 6). Most of the COPD-related deaths that occurred in cohorts from Europe, North America or Australia would have been caused by tobacco. Yet, adjusting for smoking at baseline (yes or no) removed little of the observed associations. Although the crude smoking variable used was insensitive, this suggests that the associations may be independent of tobacco consumption. However, the relationships were confounded by weight, so that

adjustment for relative weight substantially attenuated the risk relationships over all follow-up:

% (SE) change in COPD-related mortality risk for given changes in usual levels of total, HDL and non-HDL cholesterol

Adjustment	Cohort, age, sex	Cohort, age, sex + smoking (yes/no)	Cohort, age & sex + relative weight
per mmol/l ↑ total cholesterol (2,734 deaths*)	-27.6 (3.3) $\chi^2 = 69$	-26.4 (3.3) $\chi^2 = 63$	-18.1 (3.3) $\chi^2 = 30$
per mmol/l ↑ non-HDL cholesterol (335 deaths*)	-51.5 (10.2) $\chi^2 = 25$	-51.4 (10.8) $\chi^2 = 22$	-36.0 (10.7) $\chi^2 = 11$
Per 0.1 mmol/l ↓ HDL cholesterol (335 deaths*)	-7.8 (1.6) $\chi^2 = 22$	-12.8 (2.2) $\chi^2 = 35$	-6.2 (1.5) $\chi^2 = 16$

* *all* analyses included only those participants with data on total cholesterol, blood pressure, smoking (yes/no), height and weight

1.4 Joint effects of blood pressure and total cholesterol on IHD and stroke risk

Irrespective of the baseline level of total cholesterol, there were strong, approximately log-linear associations of increasing risk of both IHD and stroke with increasing SBP (data not shown). Furthermore, for IHD, there was an approximately log-linear association of increasing risk with increasing levels of total cholesterol, irrespective of the baseline level of SBP (Figure 22(a)). However, for total stroke, there was some suggestion of a negative association with total cholesterol among those with the highest SBP measurements at baseline (Figure 22(b)). Further investigation of this interaction belongs to the wider scope of the PSC. With less than 200 haemorrhagic or ischaemic strokes in the lowest fifth of baseline SBP, there were too few data to assess reliably these interactions for the stroke sub-types.

2 Implications for further research

Limited information on cholesterol fractions

Despite the uniquely large number of individuals with data on HDL cholesterol (178,000 individuals), far fewer data were available on LDL cholesterol (about 50,000 individuals) and even fewer on other sub-fractions such as lipoprotein (a), very low density lipoproteins, or triglycerides. Previous research has suggested that indices of cholesterol other than HDL and non-HDL, such as the total to HDL cholesterol ratio, may be more predictive of risk than either measurement alone. However, this inappropriately mixes two different quantities with opposing effects on risk.

Controlling for confounders

Some preliminary investigation of the effects of confounders on the cholesterol relationships were attempted - in particular, the confounding effects of smoking, alcohol and relative weight. Very crude adjustment for smoking and drinking habits at baseline (yes or no) made little difference to the observed associations. Adjustment for relative weight, however, attenuated those associations believed to be due to reverse-causality (for example, COPD and non-alcohol related cancers). For each cause of death, the effect of adjusting simultaneously for smoking (yes or no), alcohol (yes or no), height and weight on the log hazard ratios associated with total cholesterol (Table 7), non-HDL (Table 8) and HDL (Table 9) cholesterol, was investigated in the subset of individuals with data available on each factor. However, more detailed investigation of these confounding effects belongs to the wider scope of the PSC.

Heterogeneity

There was some concern about combining data from diverse populations from different eras, and with different life experiences. To avoid such issues, individuals were ranked *within* their study according to baseline measurements of the risk factor of interest, and then divided into similar-sized groups for analysis. Thus, direct comparisons between individuals were made *only within* studies and *not between* studies. Yet, there remained significant heterogeneity between the studies in the risk relationships for both IHD (Figure 23) and total stroke (Figure 24), which could not be explained fully by different measurement techniques (for example, plasma versus serum samples, fasting versus non-fasting samples), different years of baseline measurements (cohort effects and the introduction of effective cholesterol lowering therapies), mean values at baseline or the proportion of females. A major contributing factor to this observed heterogeneity would be the different proportional effects at different ages, which cannot be explored reliably due to inadequate numbers of deaths at particular ages within most individual studies. For total stroke, if the trends for qualitatively different relationships in ischaemic and haemorrhagic strokes were real, then different proportions of haemorrhagic strokes in different populations would have contributed to the observed heterogeneity in the total stroke results.

3 Discussion

The primary aim of this chapter was to investigate the real relationships between cholesterol and various causes of death by applying appropriate statistical techniques with sufficient data. Only by taking proper account of the regression dilution bias; the full range of cholesterol values, as well as the opposing effects of

HDL and the remaining non-HDL cholesterol; and, in particular, by undertaking age-specific analyses, was it possible to establish reliably the relationship between cholesterol, its sub-fractions and IHD. These analyses have demonstrated that, as a cause of premature death due to IHD, blood lipids are about five times as important as has been implied by inappropriate analyses of baseline total cholesterol in all ages together. Moreover, the strengths of the relationships were similar in men and women in each of the age groups studied. Thus, the importance of HDL and the remaining non-HDL cholesterol as determinants of risk has long been underestimated.

The relationships between cholesterol and stroke were more complex, and only by studying age-specific relationships between cholesterol fractions and stroke sub-types will the real relationships emerge. Despite the uniquely large number of individuals with data available on HDL cholesterol and stroke sub-types, there remained insufficient data to assess reliably the age-specific risk relationships. Unlike the blood pressure relationships, the relationships between total cholesterol and haemorrhagic stroke, and between total cholesterol and ischaemic stroke, appeared qualitatively different. There was evidence of increasing ischaemic stroke risk with increasing total cholesterol at younger ages, while there seemed to be a decreasing haemorrhagic stroke risk with increasing total cholesterol at older ages. When men and women of all ages were analysed together, there were no significant relationships between HDL or the remaining non-HDL cholesterol with either total stroke, or with either of the stroke sub-types. Although there was a trend towards an inverse relationship between HDL cholesterol and ischaemic stroke, these non-age-specific associations should be viewed with extreme caution. Thus, the real sex- and, more importantly, age-specific relationships between cholesterol fractions and stroke

sub-types await the availability of yet more information on sub-typed stroke deaths among individuals with baseline measurements of cholesterol fractions.

With regard to other causes of death, the associations with diabetic deaths were comparable to those with IHD, probably due to the large vascular component in the majority of such deaths. There were too few deaths ascribed to diabetes to determine independently the age- and sex-specific relationships. However, by extrapolating the IHD relationships (which is reasonable given the close compatibility with the diabetic-related relationships in all ages combined), non-HDL cholesterol and HDL cholesterol are likely to be much more important determinants of mortality risk among these individuals than has previously been appreciated. For the various cancers studied, the negative associations with total cholesterol were due primarily to negative associations with HDL cholesterol, but these largely disappeared within 5 years of the cholesterol measurements, suggesting reverse-causality. The only relationship sustained beyond 5 years of follow-up was that between HDL cholesterol and upper aero-digestive cancer. However, upper aero-digestive cancer risk is strongly related to alcohol, which in turn is positively associated with HDL cholesterol. Therefore, this relationship may be attributed largely to confounding with alcohol consumption patterns, although this remains unclear.

The most important non-vascular cause of death for which the relationships did not attenuate markedly within 5 years of follow-up was COPD-related death. There was a positive association with HDL cholesterol, and a negative association with non-HDL cholesterol. In these data, COPD was strongly and inversely related to relative weight (a four-fold difference in risk between those in the bottom fifth of their cohort's distribution of body mass index (i.e. weight relative to the square of height) and those

in the top fifth), although cholesterol was only weakly correlated with relative weight (ρ=0.17). After controlling for relative weight, 60% of the observed association with total cholesterol was removed (adjusted χ^2=35 compared with unadjusted χ^2=86) and 50% of the observed association with non-HDL cholesterol was removed (adjusted χ^2=52 compared with unadjusted χ^2=25). This suggests the strong relationships observed between total and non-HDL cholesterol and COPD-related death may be due at least partly to reverse-causality. Crude adjustment for smoking removed little of the observed cholesterol-COPD associations in these data, although the available smoking data were not sufficiently sensitive to control adequately for any confounding.

The availability of large numbers of non-vascular deaths has avoided spurious relationships being observed merely by the play of chance. Removal of deaths within 5 years of the cholesterol measurement, together with preliminary attempts to control for confounding factors, attenuated most of the remaining observed inverse relationships in which chronic disease might have caused low cholesterol.

4 Conclusions

In conclusion, HDL and the remaining non-HDL cholesterol are far more important determinants of IHD death than has previously been realised, particularly at younger ages. Furthermore, the extent to which risk is diminished by prolonged higher HDL cholesterol or prolonged lower non-HDL cholesterol appears independent of the range being studied. Only through quantification of the important opposing effects of the cholesterol fractions (and the consequent need to move away from assessing total cholesterol levels) together with the acceptance of no clinically relevant inverse

associations with cancer or other non-vascular mortality (except the remarkable associations with COPD), will the importance of low HDL-cholesterol and high non-HDL cholesterol as determinants of premature mortality be fully realised.

Figures and Tables

Table 1: Strength of the associations of *usual* total cholesterol with various causes of death:
log hazard ratios per mmol/l higher usual total cholesterol

Table 2: Strength of the associations of *usual* non-HDL cholesterol with various causes of death:
log hazard ratios per one mmol/l higher usual non-HDL cholesterol

Table 3: Strength of the associations of *usual* HDL cholesterol with various causes of death:
log hazard ratios per 0.1 mmol/l higher usual HDL cholesterol

Table 4: Strength of the associations of *usual* total cholesterol with various causes of death, excluding deaths within the first 5 years:
log hazard ratios per one mmol/l higher usual total cholesterol

Table 5: Strength of the associations of *usual* non-HDL cholesterol with various causes of death, excluding deaths within the first 5 years:
log hazard ratios per one mmol/l higher usual non-HDL cholesterol

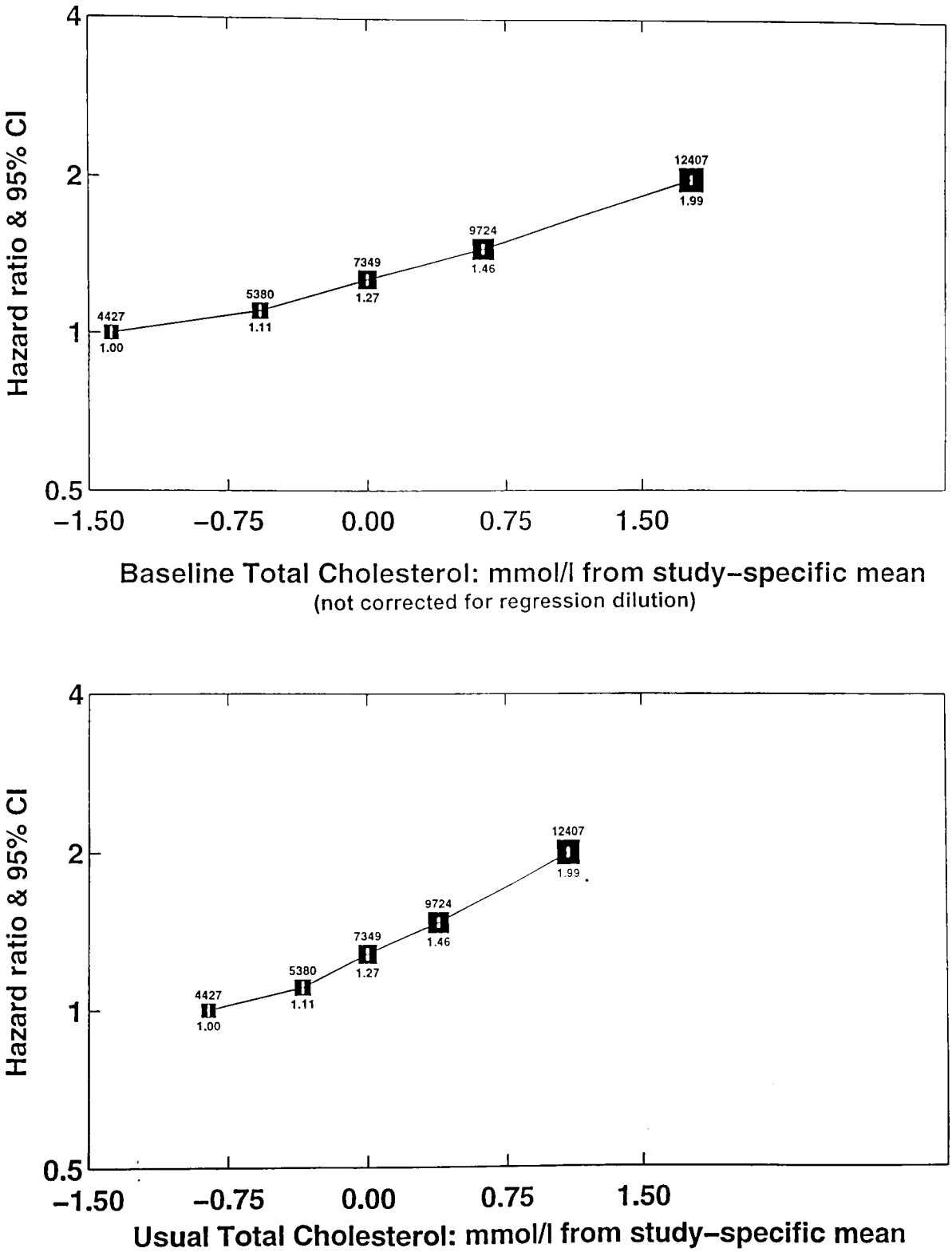
Table 6: Strength of the associations of *usual* HDL cholesterol with various causes of death, excluding deaths within the first 5 years:
log hazard ratios per 0.1 mmol/l higher usual HDL cholesterol

Table 7: Strength of the associations of *usual* total cholesterol with various causes of death, after controlling for smoking (yes/no), drinking (yes/no) and obesity at baseline, and excluding deaths within the first 5 years:
log hazard ratios per one mmol/l higher usual total cholesterol

Table 8: Strength of the associations of *usual* non-HDL cholesterol with various causes of death, after controlling for smoking (yes/no), drinking (yes/no) and obesity at baseline, and excluding deaths within the first 5 years:
log hazard ratios per one mmol/l higher usual non-HDL cholesterol

Table 9: Strength of the associations of *usual* HDL cholesterol with various causes of death, after controlling for smoking (yes/no), drinking (yes/no) and relative weight at baseline, and excluding deaths within the first 5 years:
log hazard ratios per 0.1 mmol/l higher usual HDL cholesterol

Figure 1: Proportional IHD risks (a) by baseline and (b) by usual total cholesterol



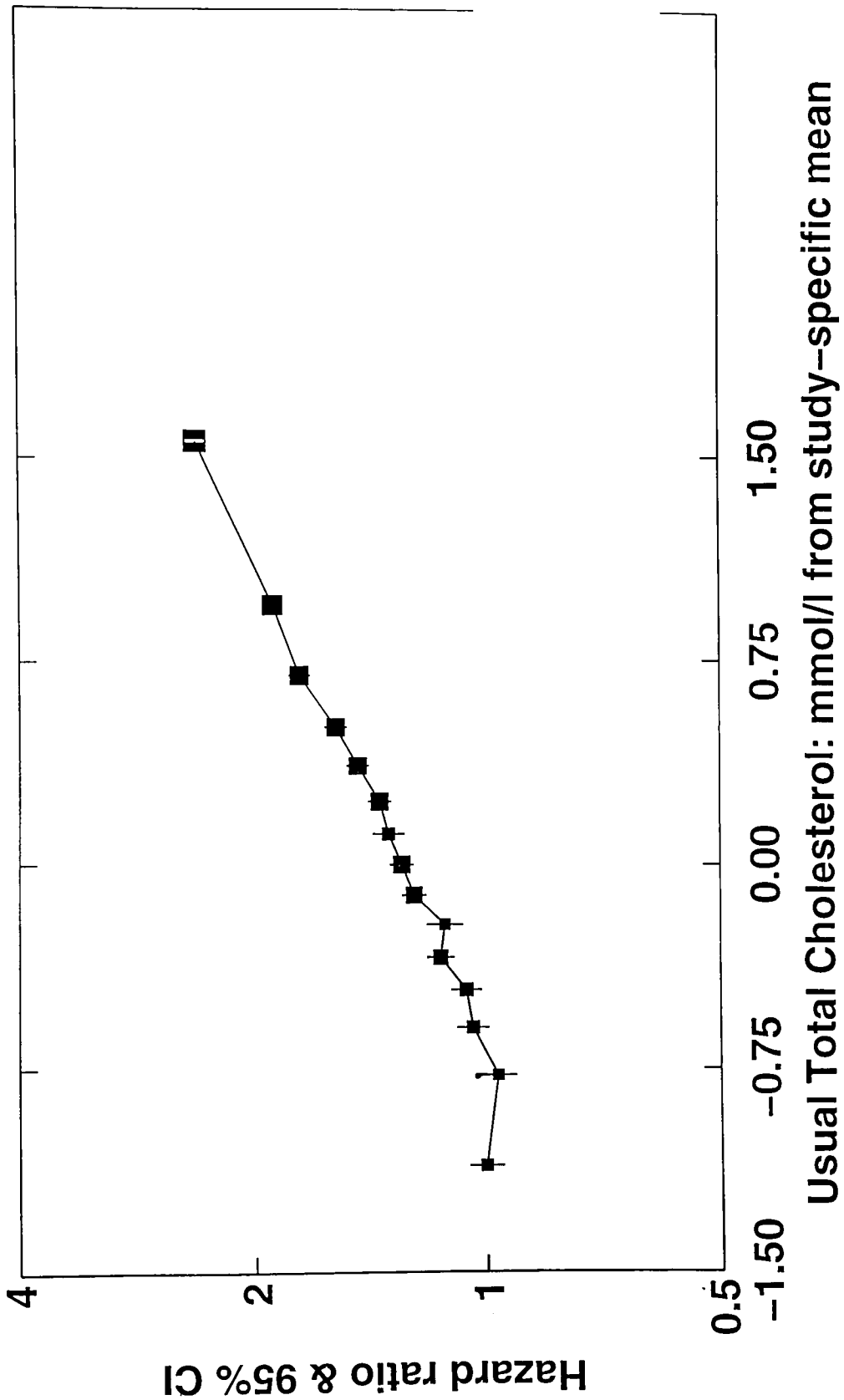
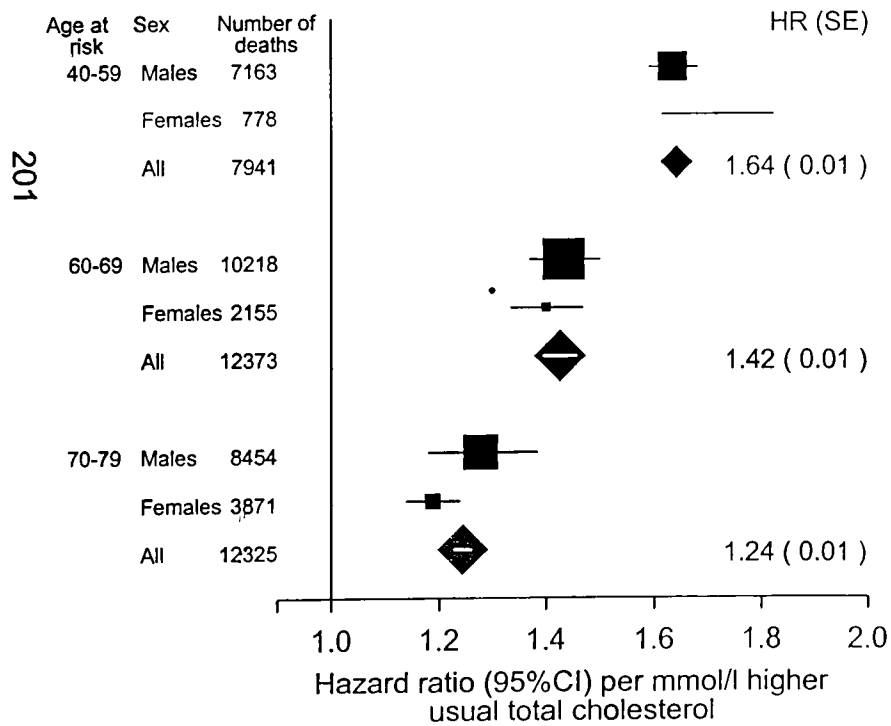


Figure 2: Proportional IHD risks throughout the entire range

Ischaemic heart disease
(Total: 39287 deaths)

Figure 3: Strength and shape of the age- and sex-specific associations between usual total cholesterol and IHD risk

Ischaemic heart disease
(Total: 32,639 deaths)



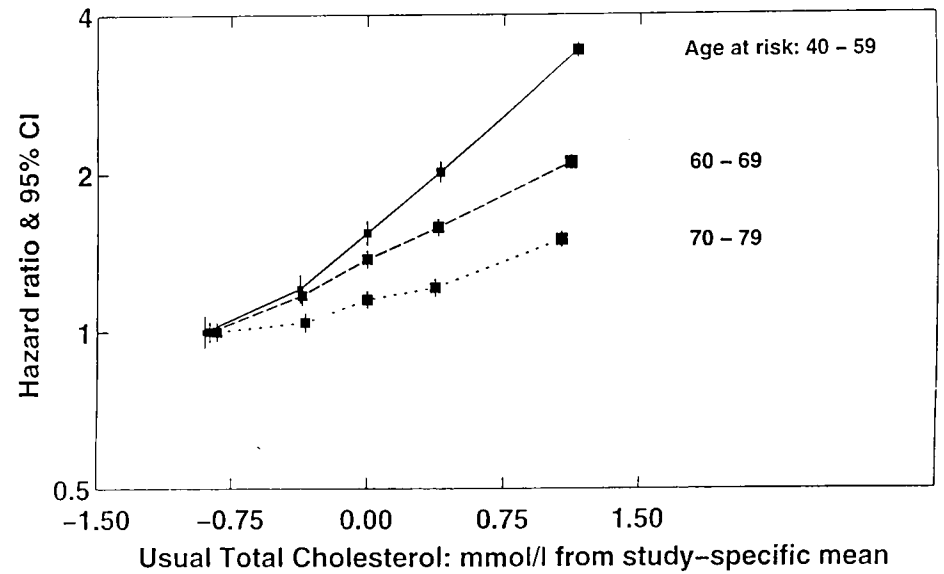
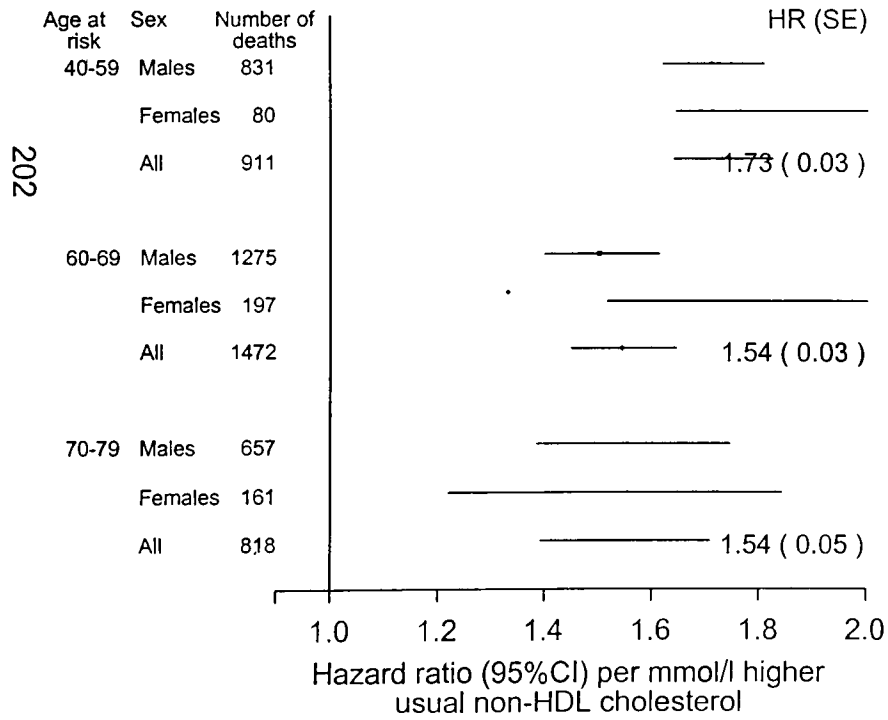


Figure 4: Strength and shape of the age- and sex-specific associations between usual non-HDL cholesterol with IHD risk

Ischaemic heart disease
(Total: 3,201 deaths)



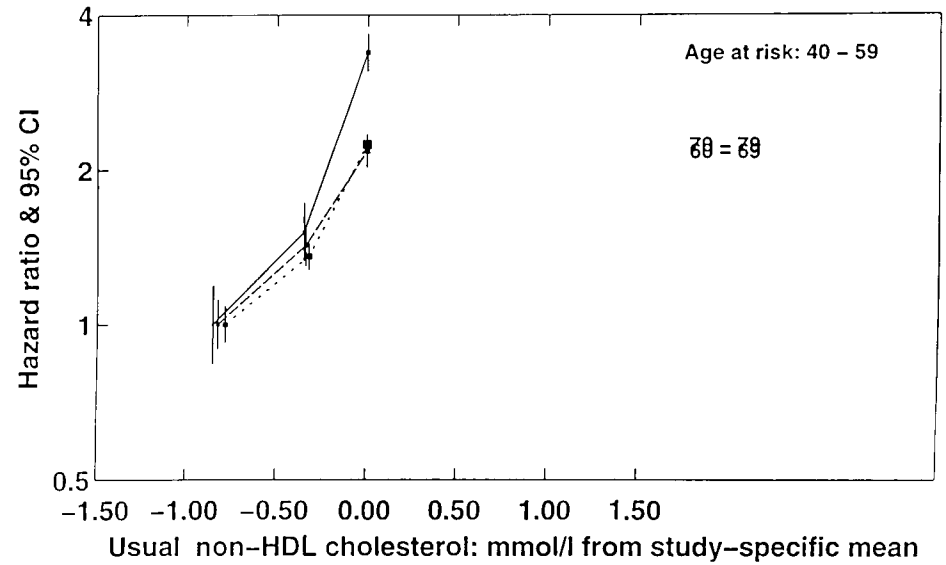
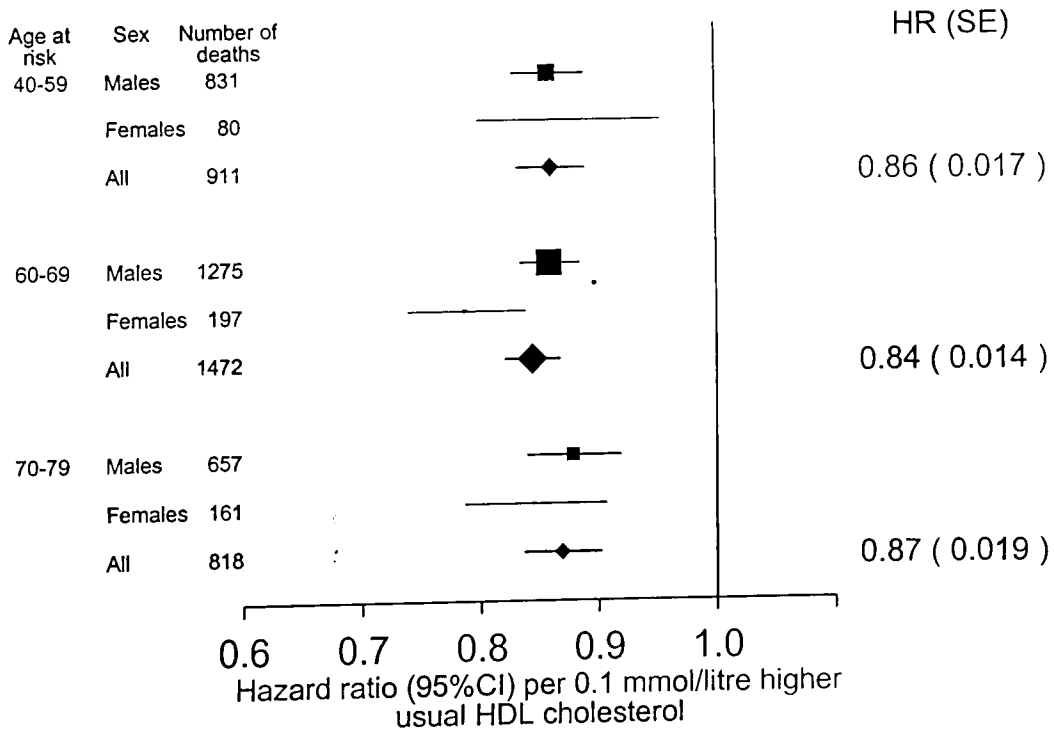
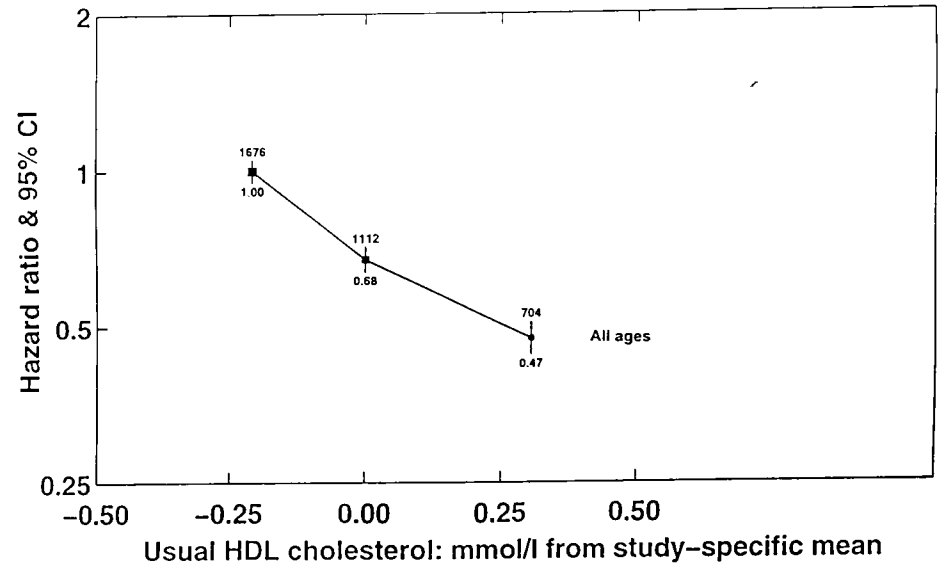


Figure 5: Strength and shape of the age- and sex-specific associations between usual HDL cholesterol and IHD risk

Ischaemic heart disease
(Total: 3,201 deaths)





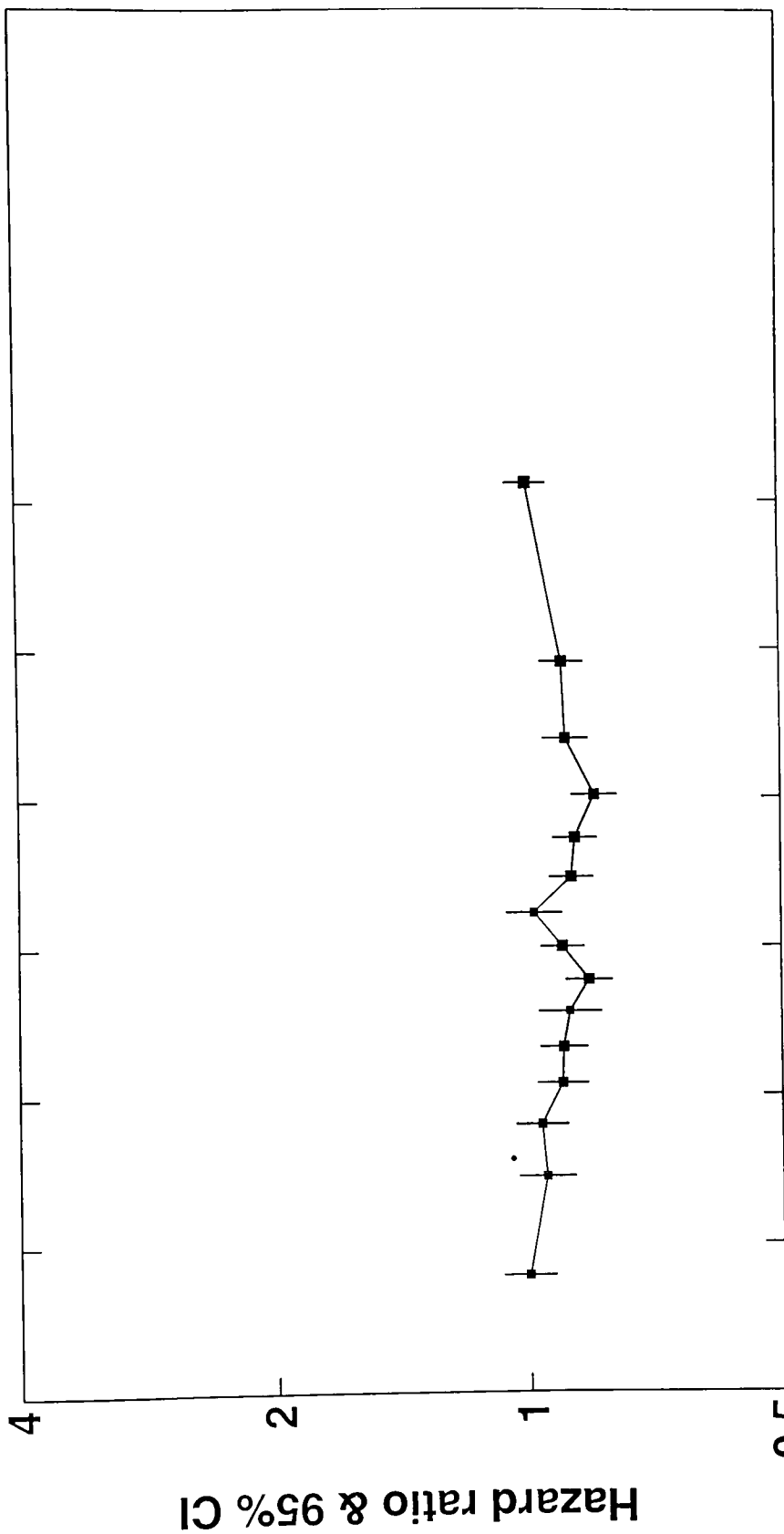


Figure 6: Proportional stroke risks throughout the entire range of usual total cholesterol

Figure 7: Proportional risks of ischaemic and haemorrhagic stroke by usual total cholesterol

(Total: 5490 deaths)

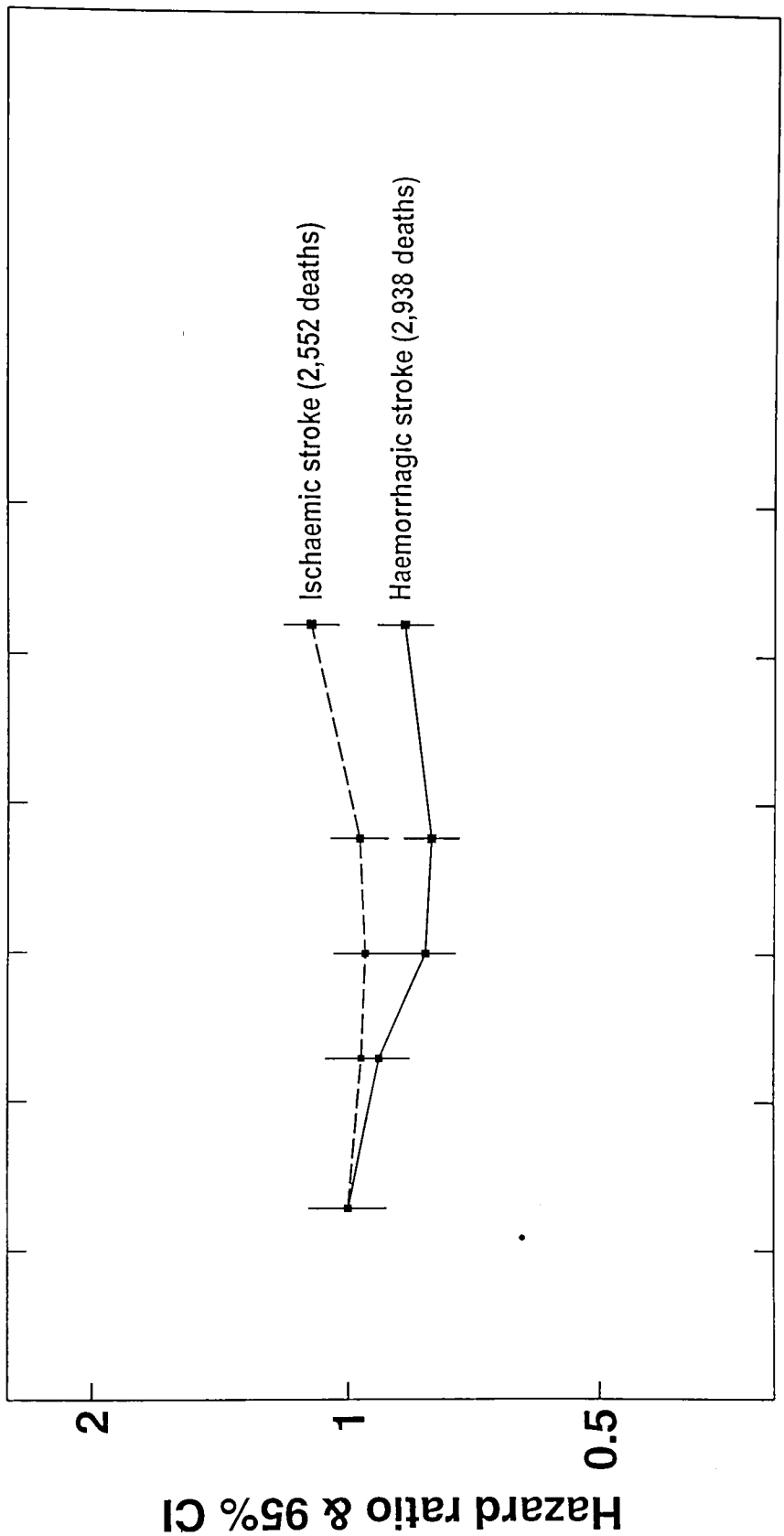


Figure 8: Strength of the age- and sex-specific associations between usual total cholesterol and total stroke risks, haemorrhagic stroke risks, and ischaemic stroke risks

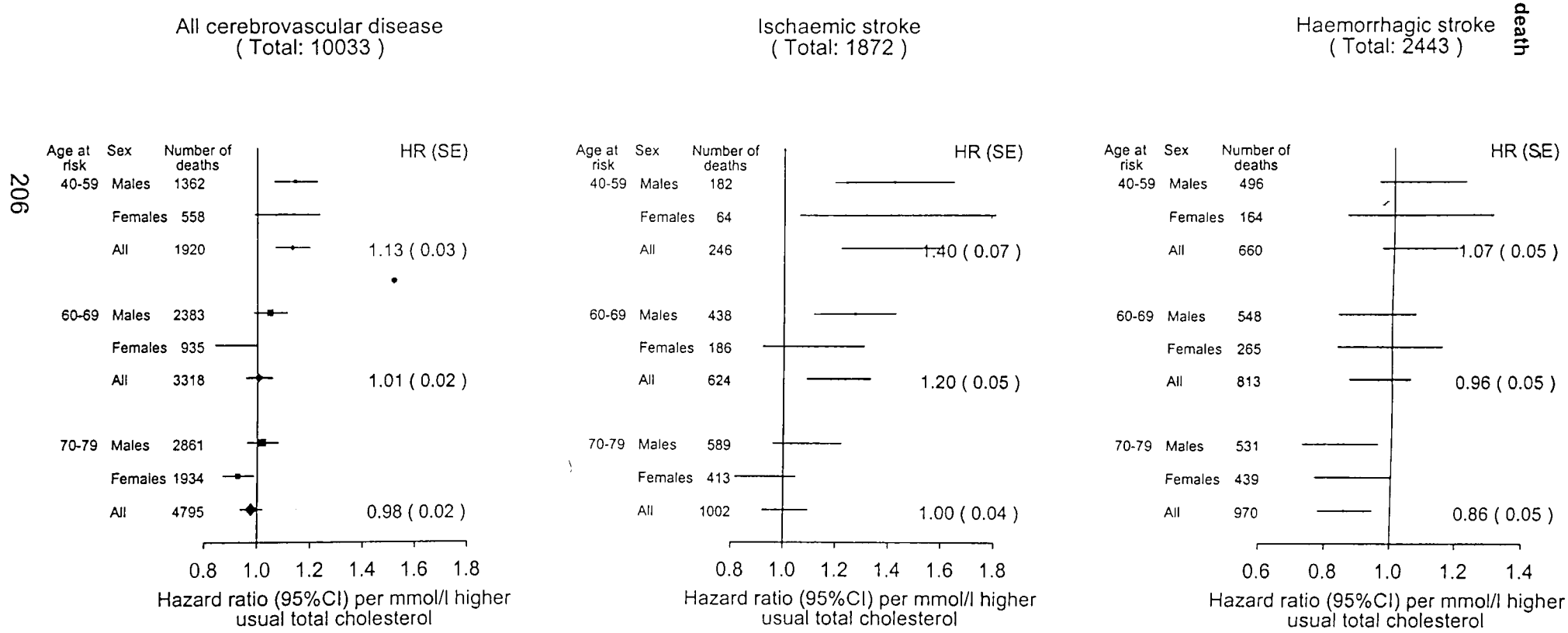
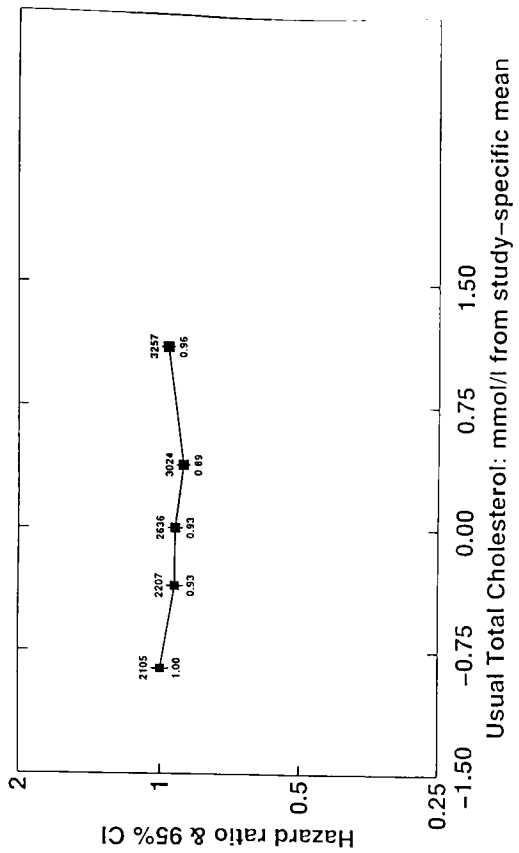
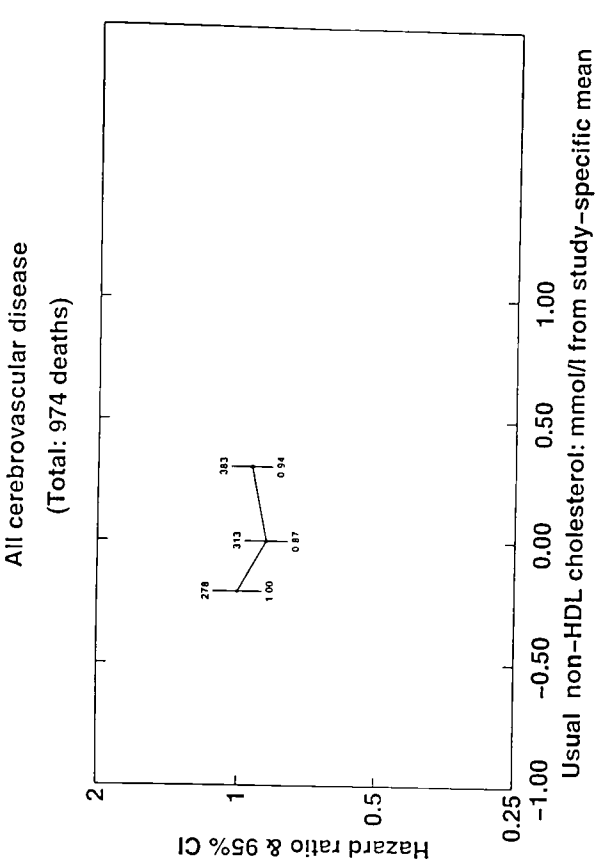


Figure 9: Proportional stroke risks by usual total, non-HDL and HDL cholesterol





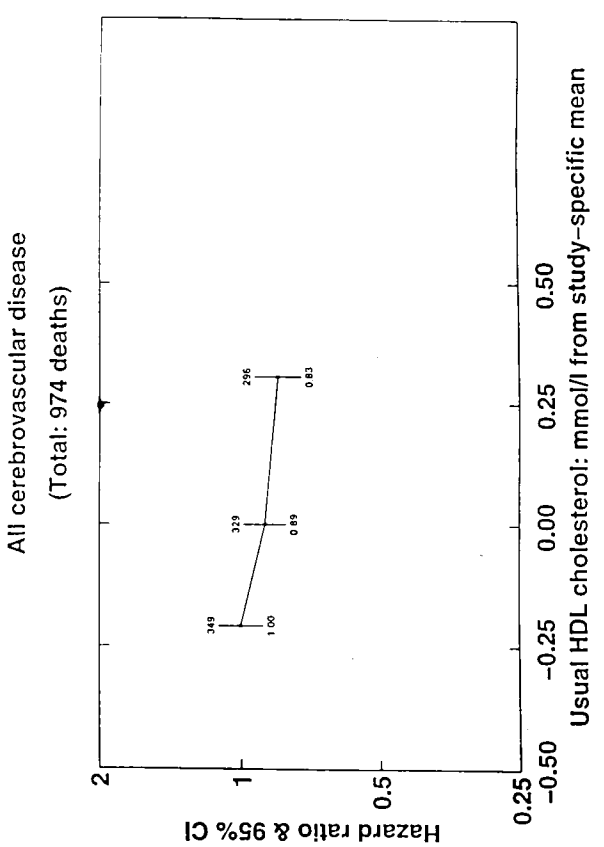
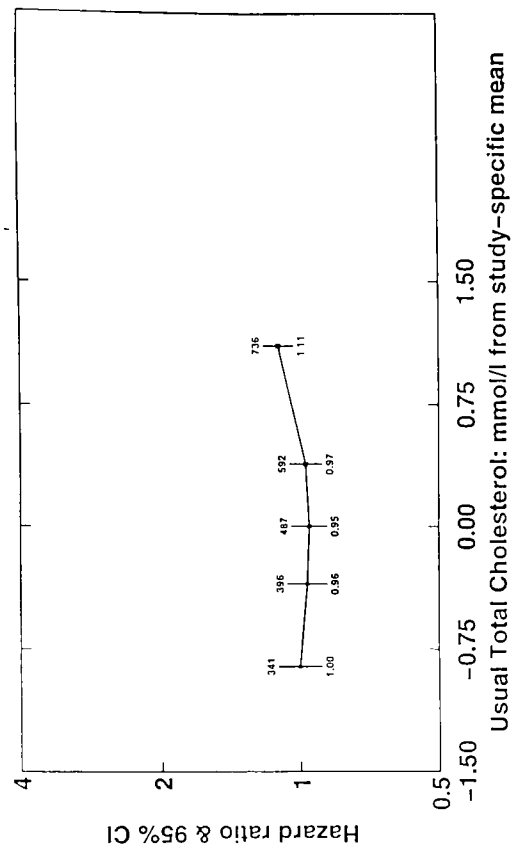
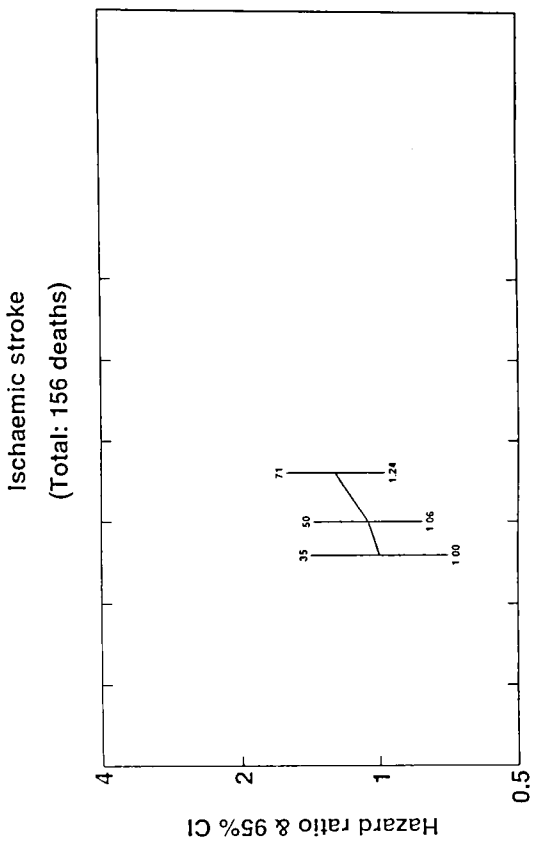


Figure 10: Proportional ischaemic stroke risks by usual total, non-HDL and HDL cholesterol





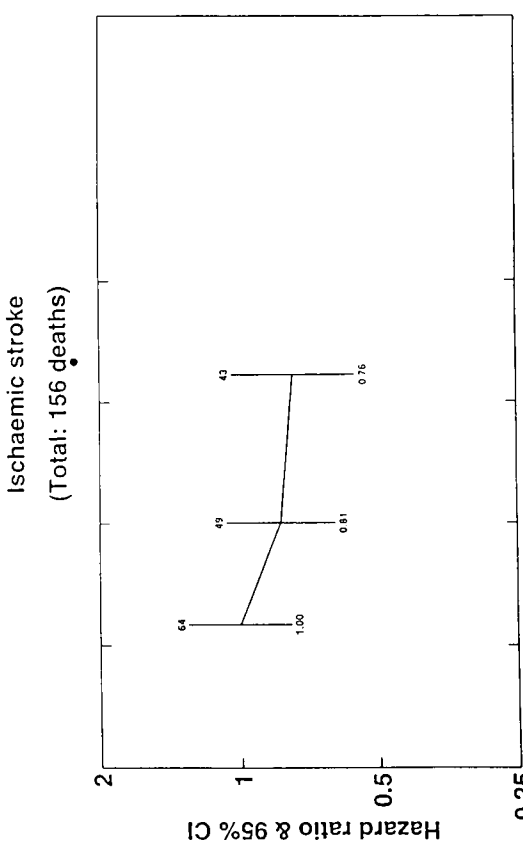
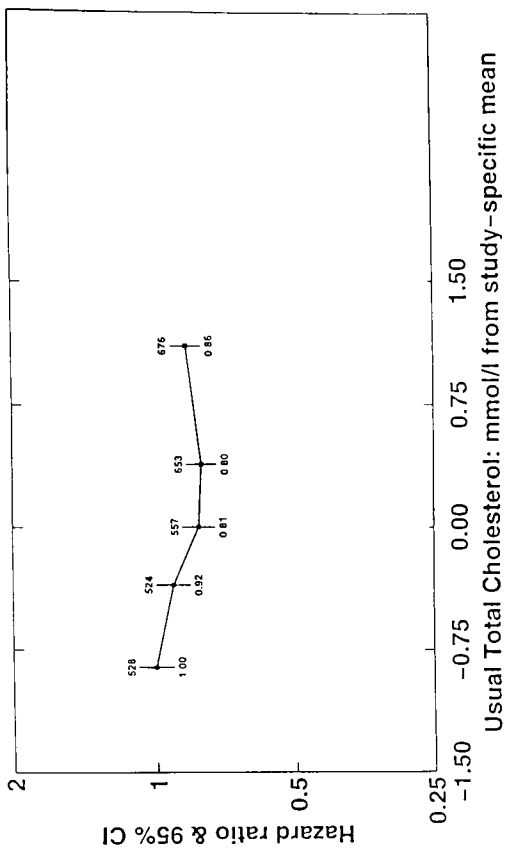
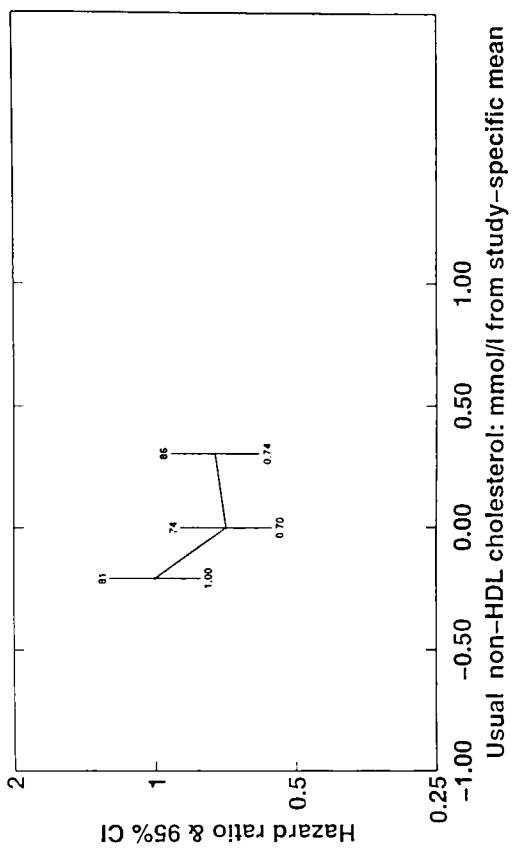


Figure 11: Proportional haemorrhagic stroke risks by usual total, non-HDL and HDL cholesterol





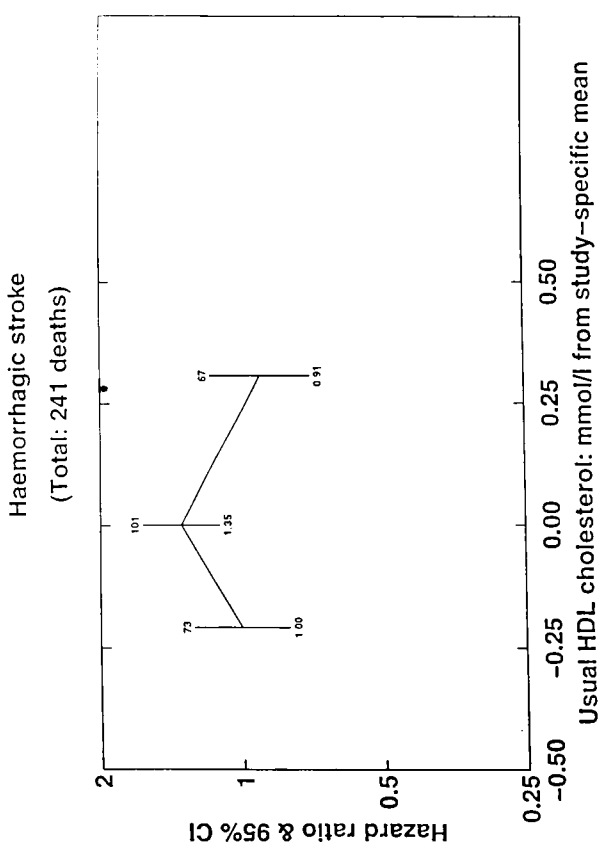
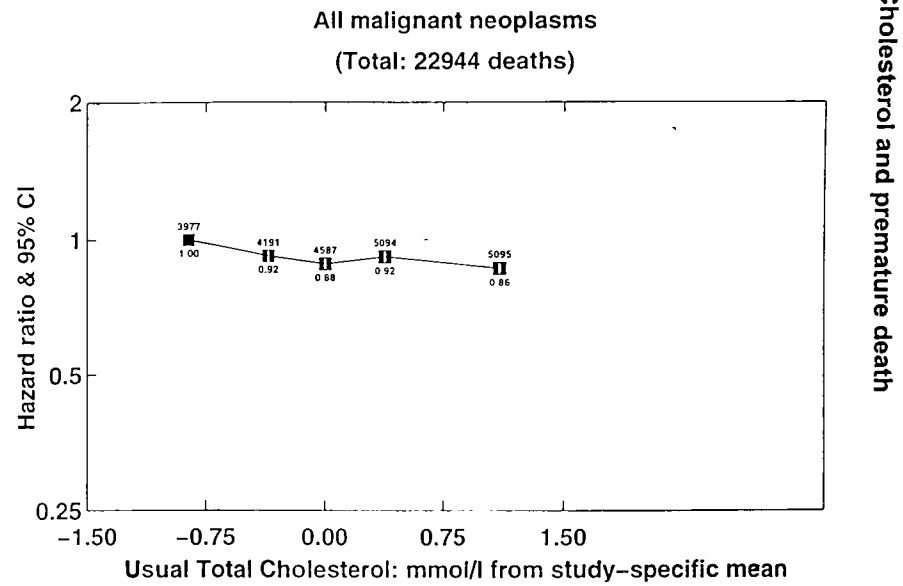
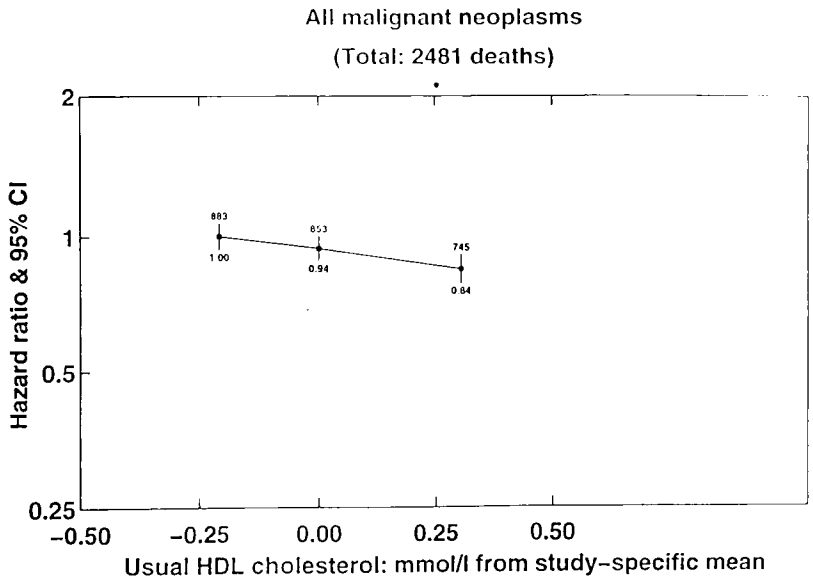


Figure 12: Proportional risk of total cancer mortality by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements





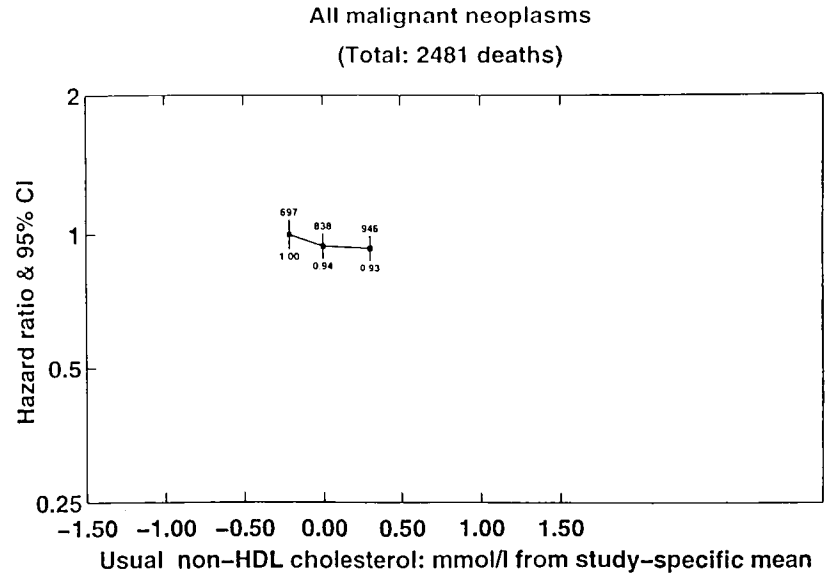
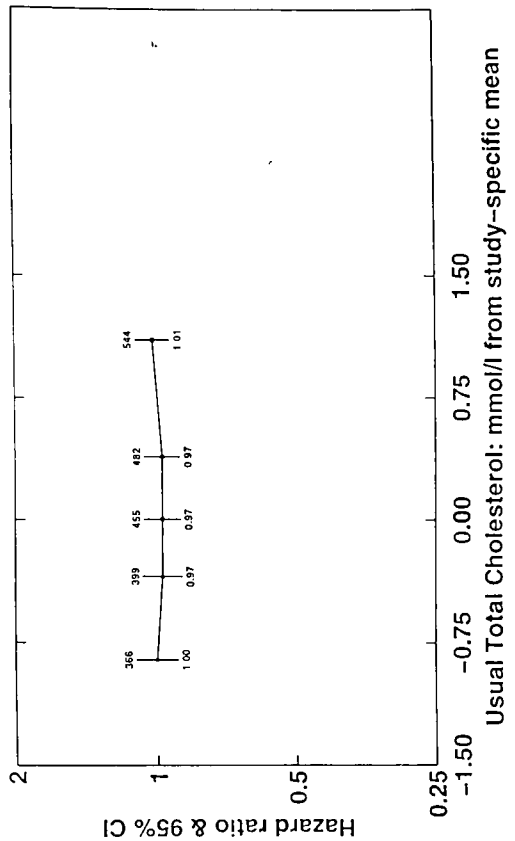
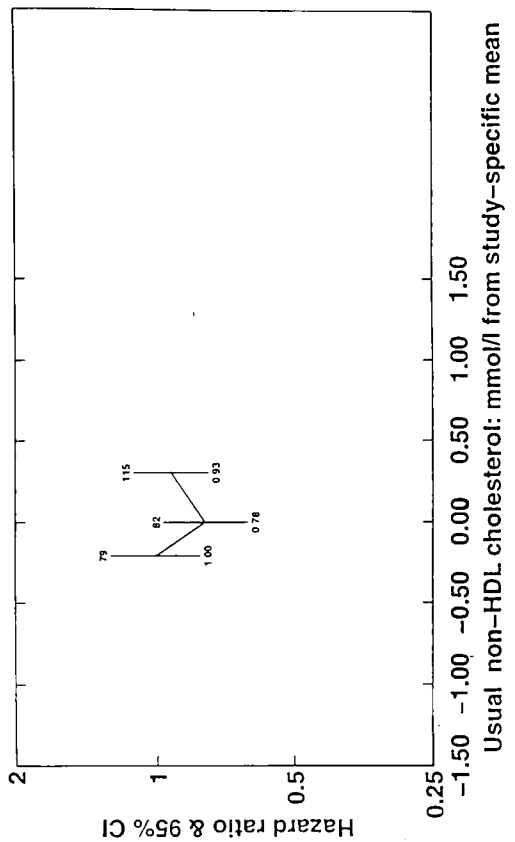


Figure 13: Proportional colorectal (large intestine) cancer risks by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements





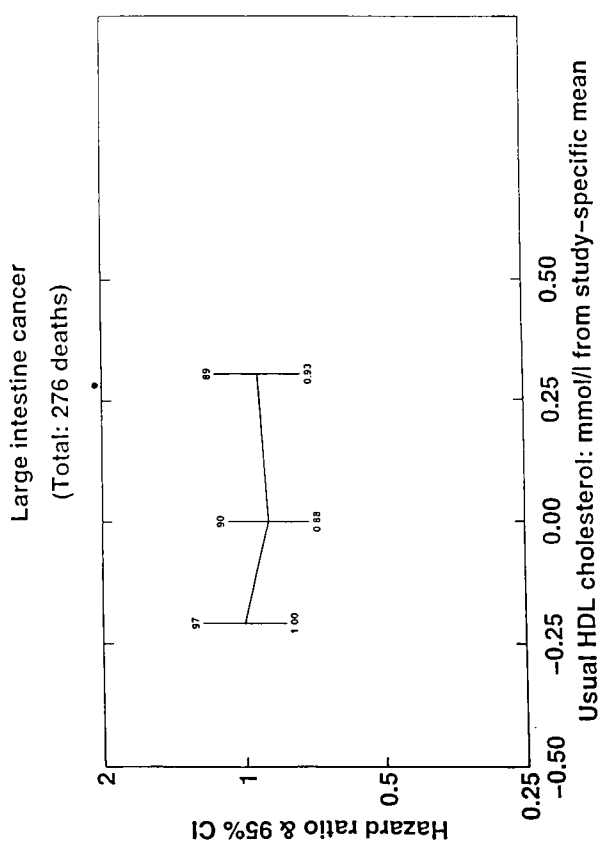


Figure 14: Proportional lung cancer risks by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements

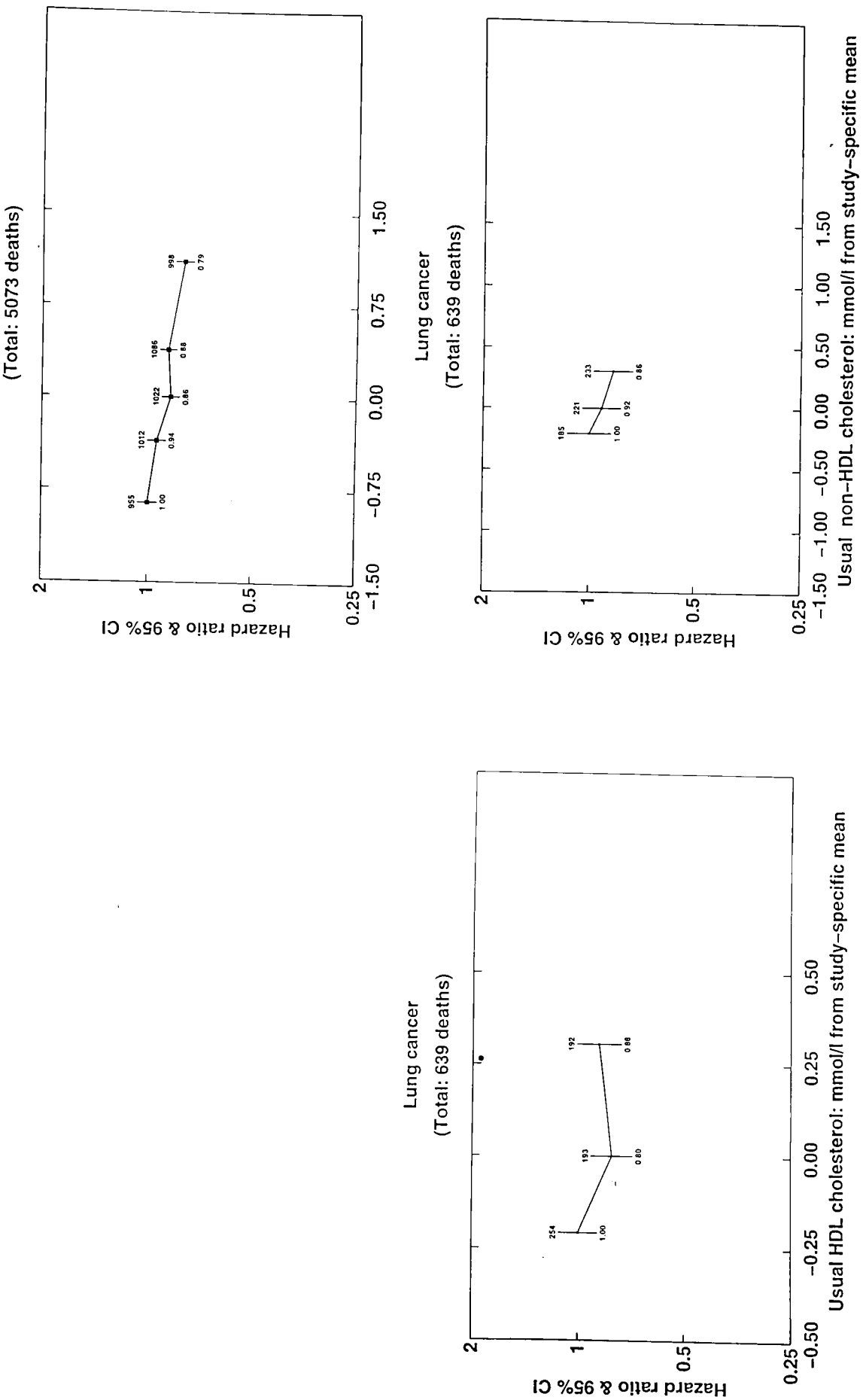
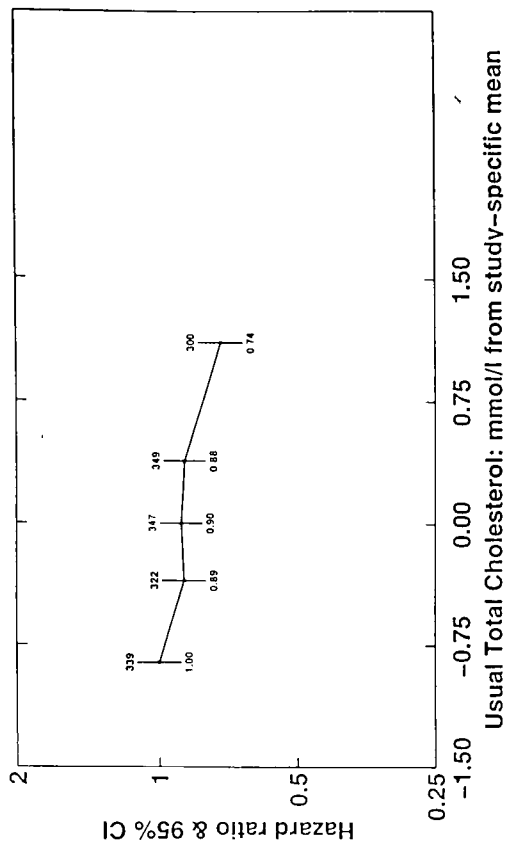
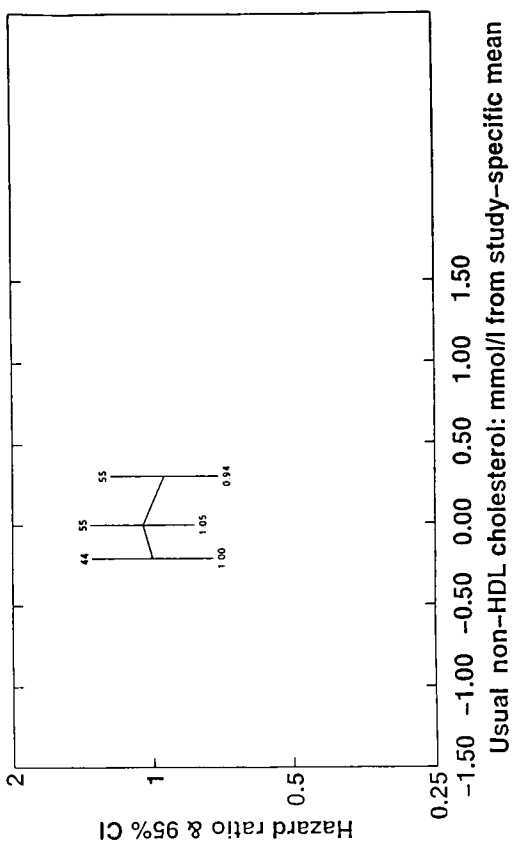


Figure 15: Proportional stomach cancer risks by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements





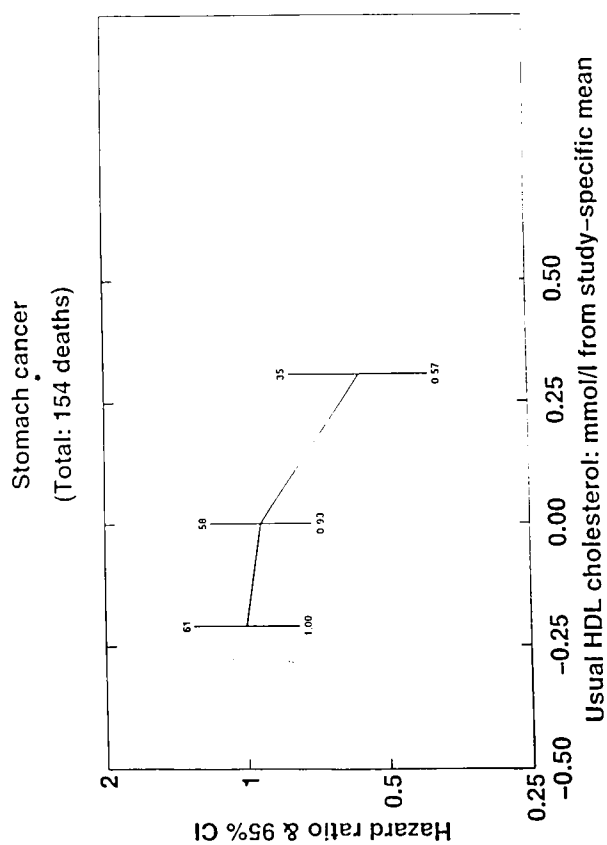
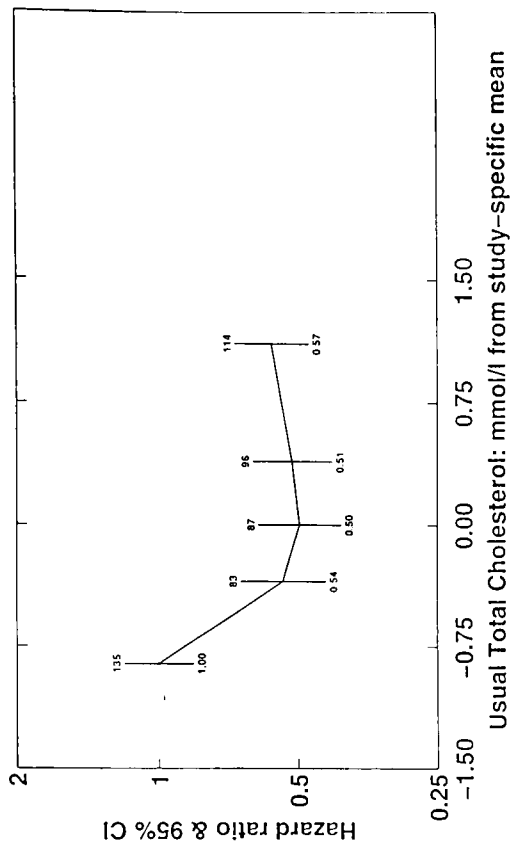
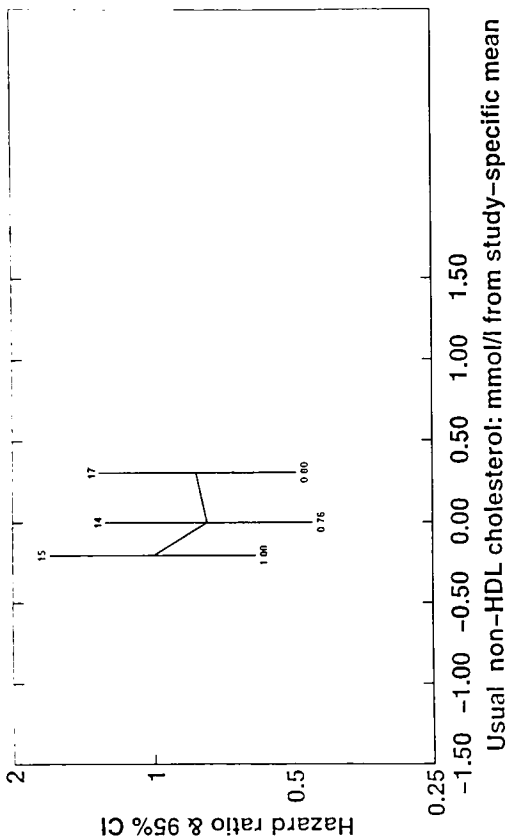


Figure 16: Proportional liver cancer risks by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements





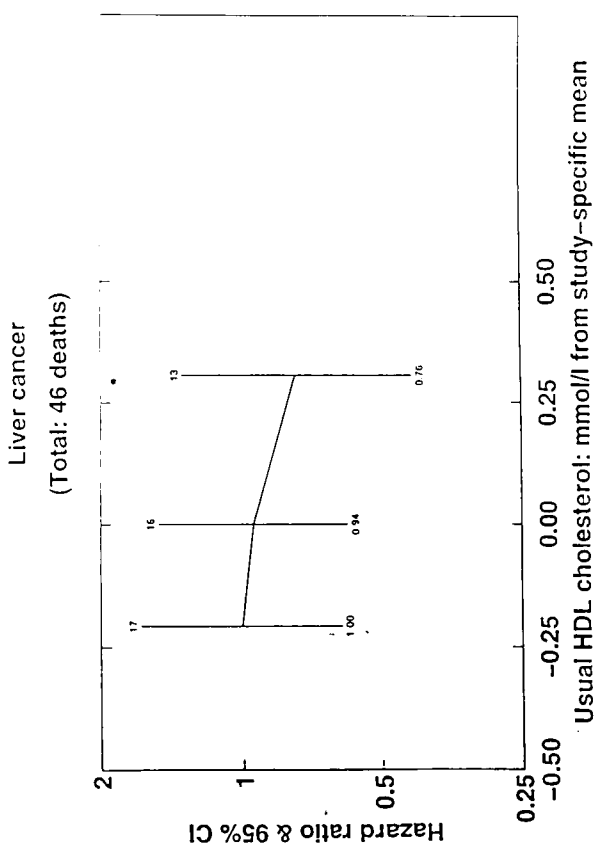
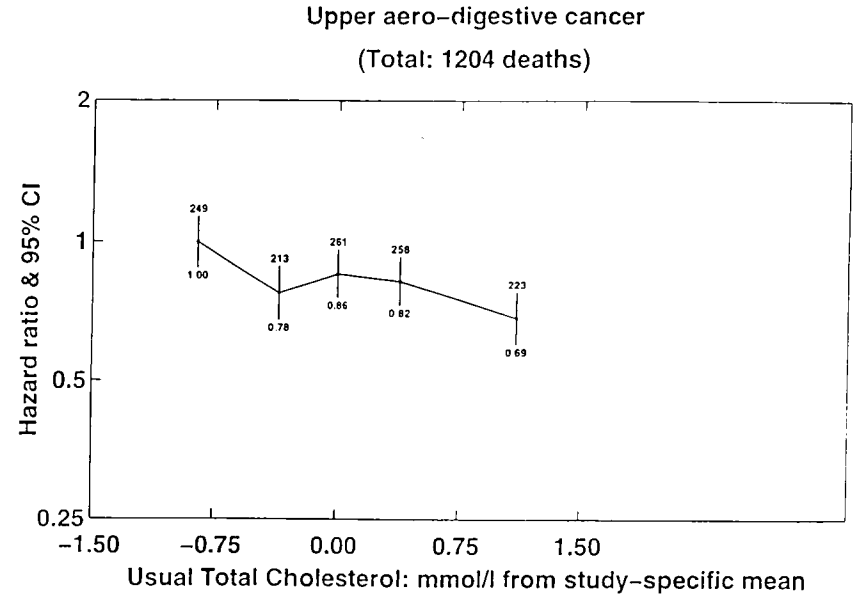
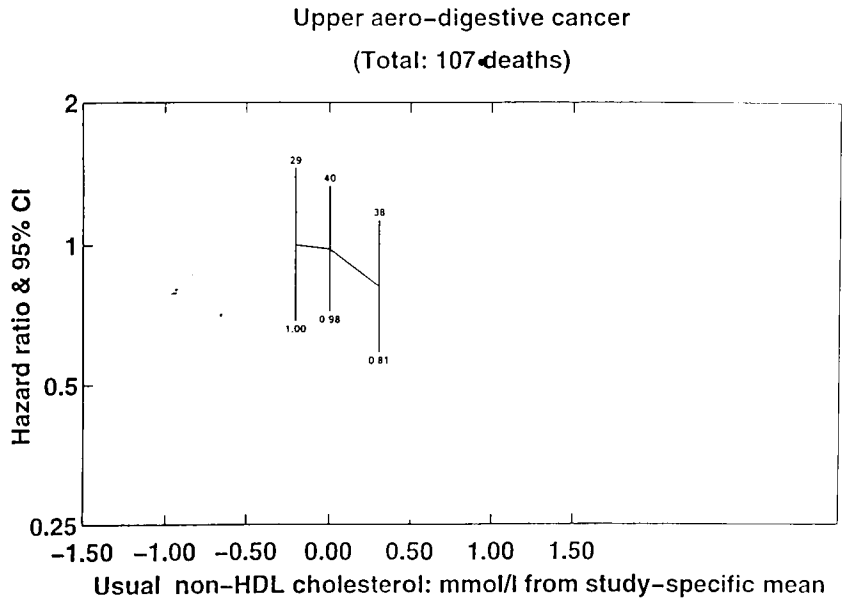


Figure 17: Proportional upper-aerodigestive cancer risks by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements





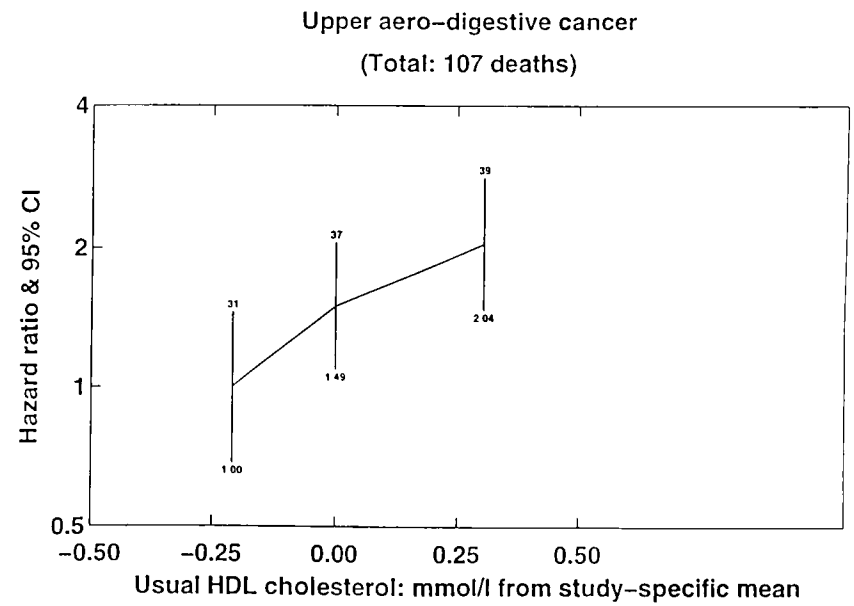
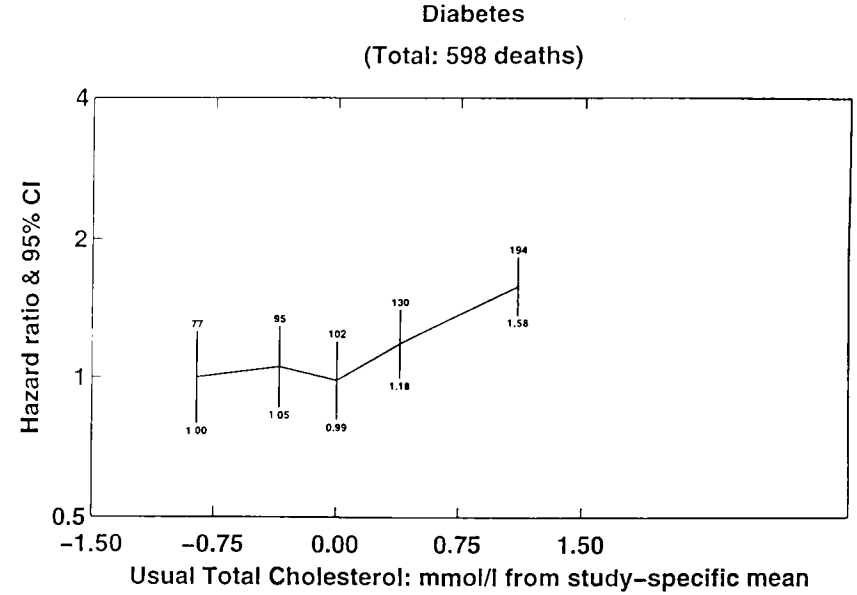
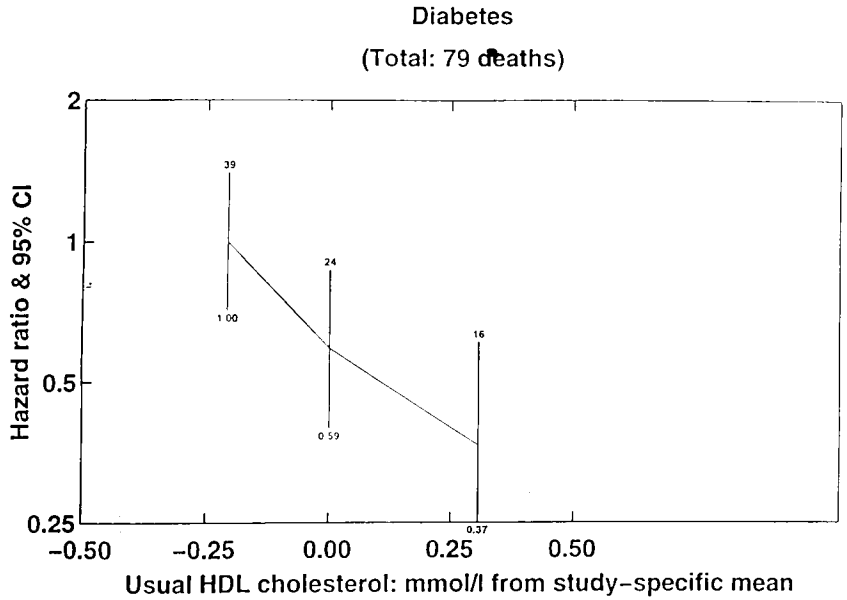


Figure 18: Proportional diabetes related mortality risks by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements





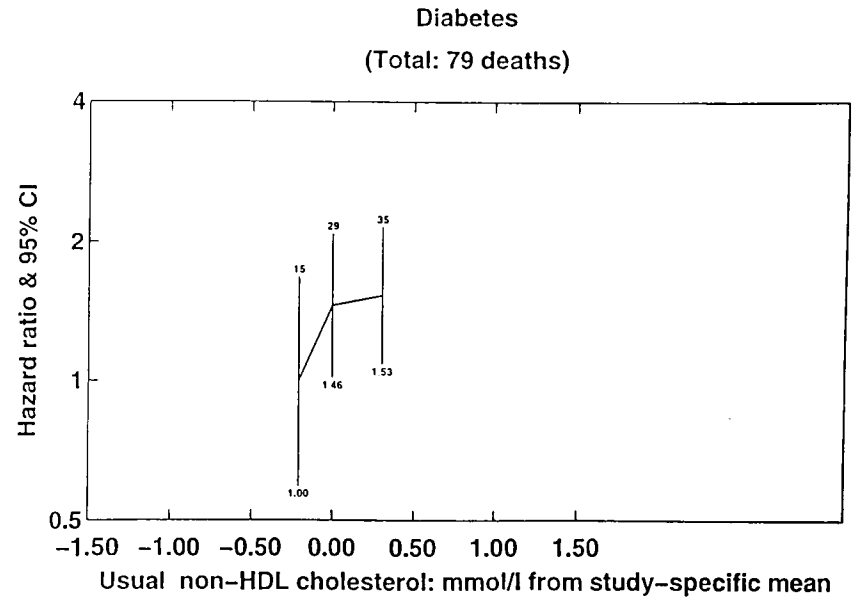
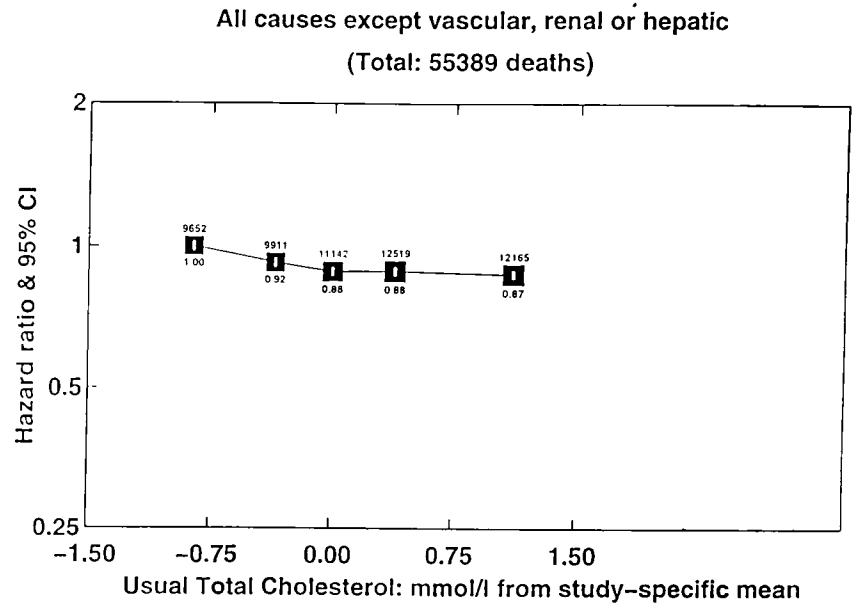
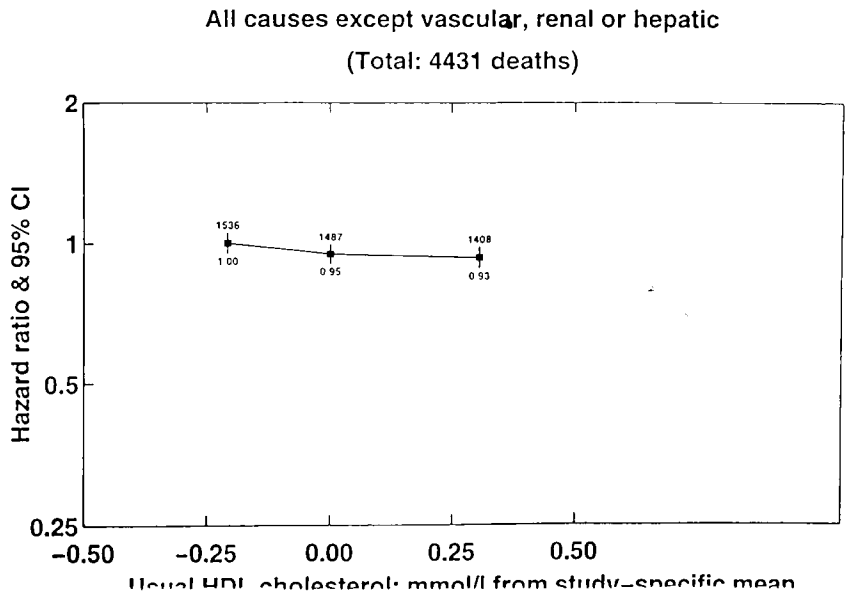


Figure 19: Proportional risks of death by any cause except vascular, renal or hepatic diseases by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements





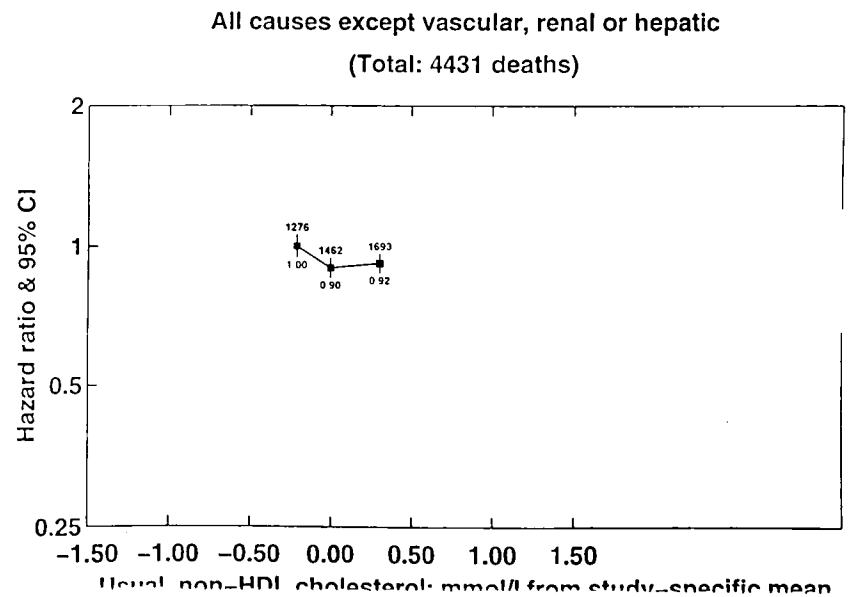
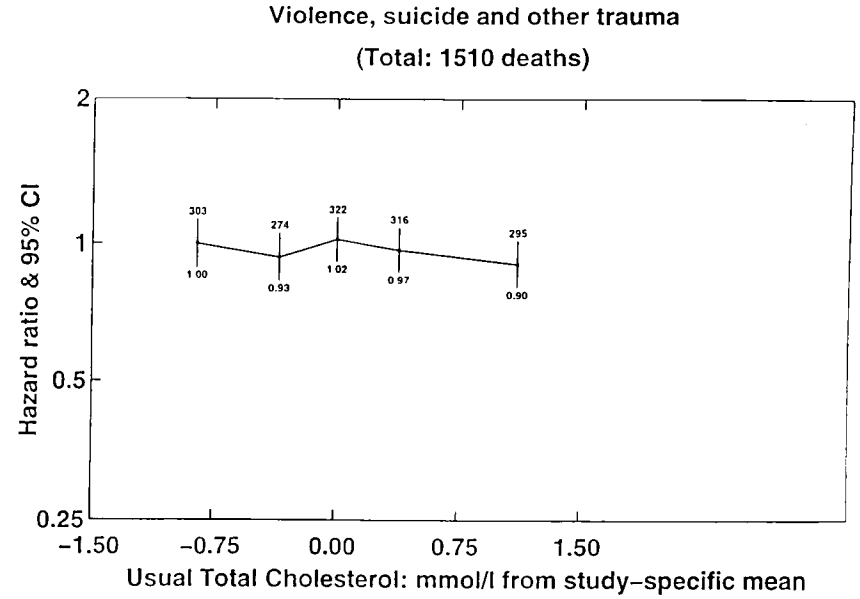
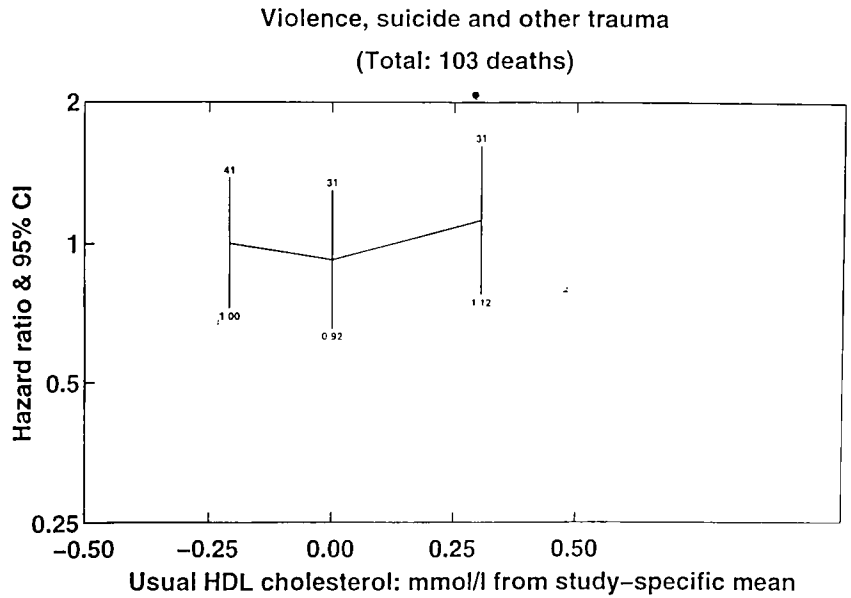


Figure 20: Proportional risks of death by violence, suicide or trauma by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements





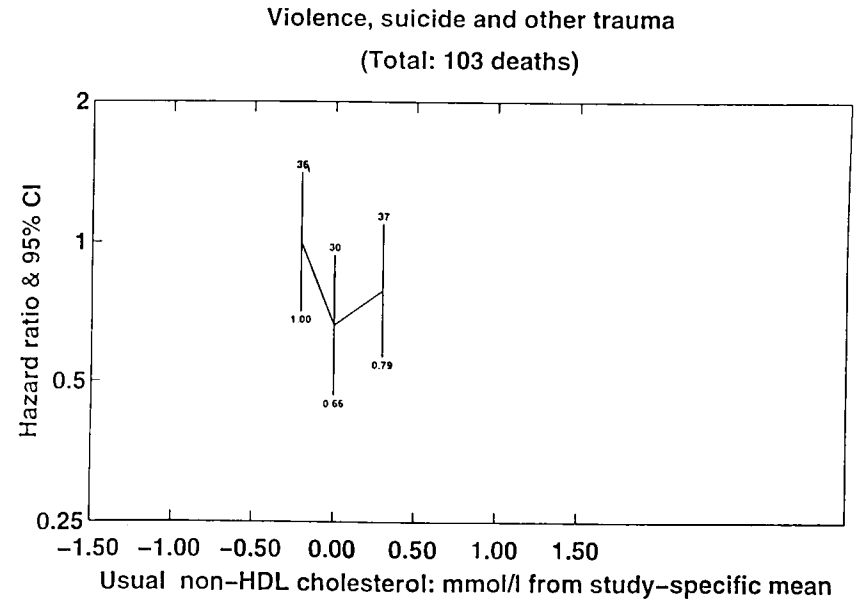
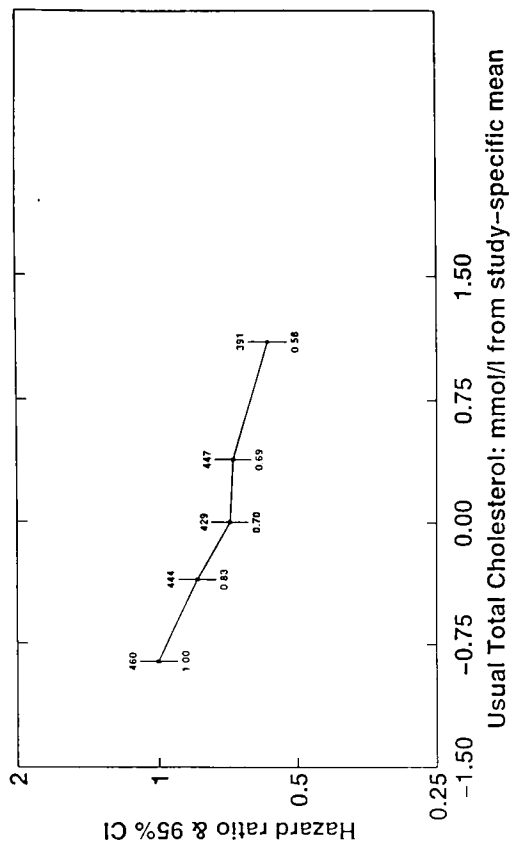
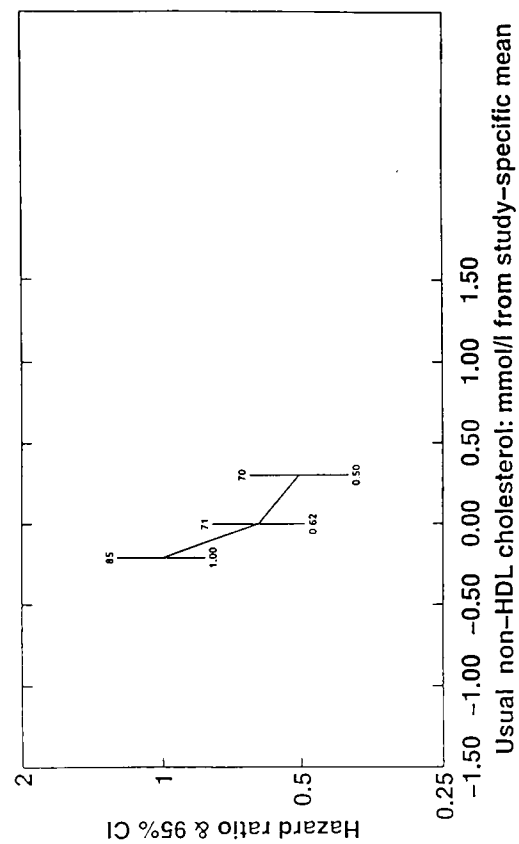


Figure 21: Proportional COPD related mortality risks by usual total, non-HDL and HDL cholesterol, excluding deaths within 5 years of the cholesterol measurements





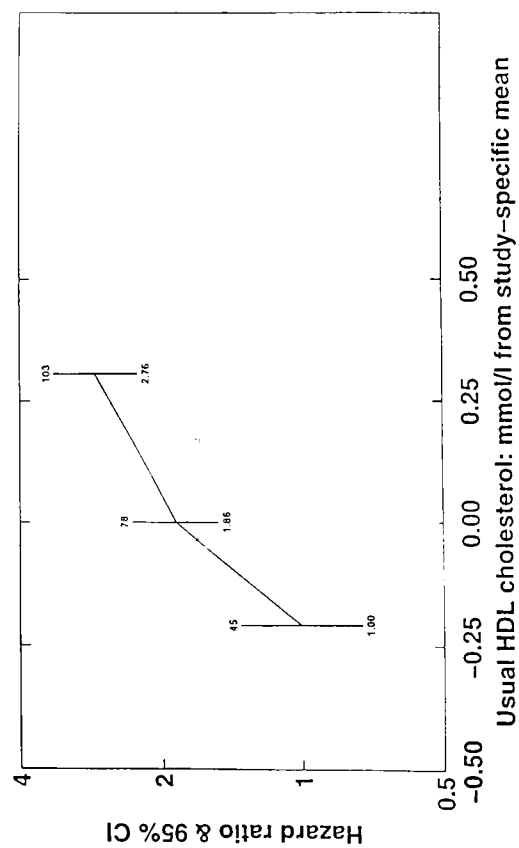


Figure 22: Proportional risks of (a) IHD and (b) total stroke, by usual total cholesterol, within fifths of SBP defined by baseline measurements

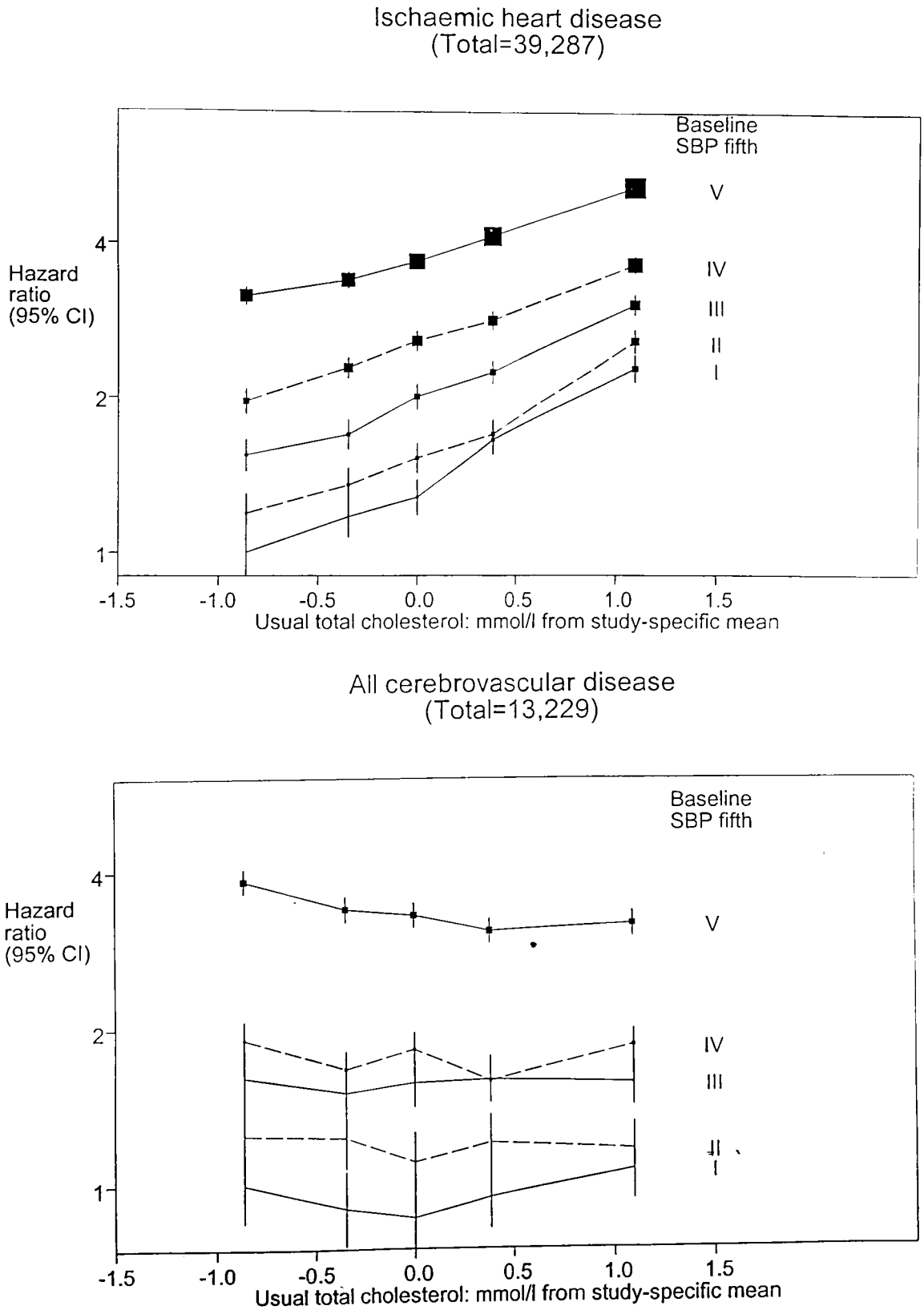


Figure 23: Strength of the study-specific associations between *usual* total cholesterol and IHD risk, with study- and duration-specific corrections for regression dilution

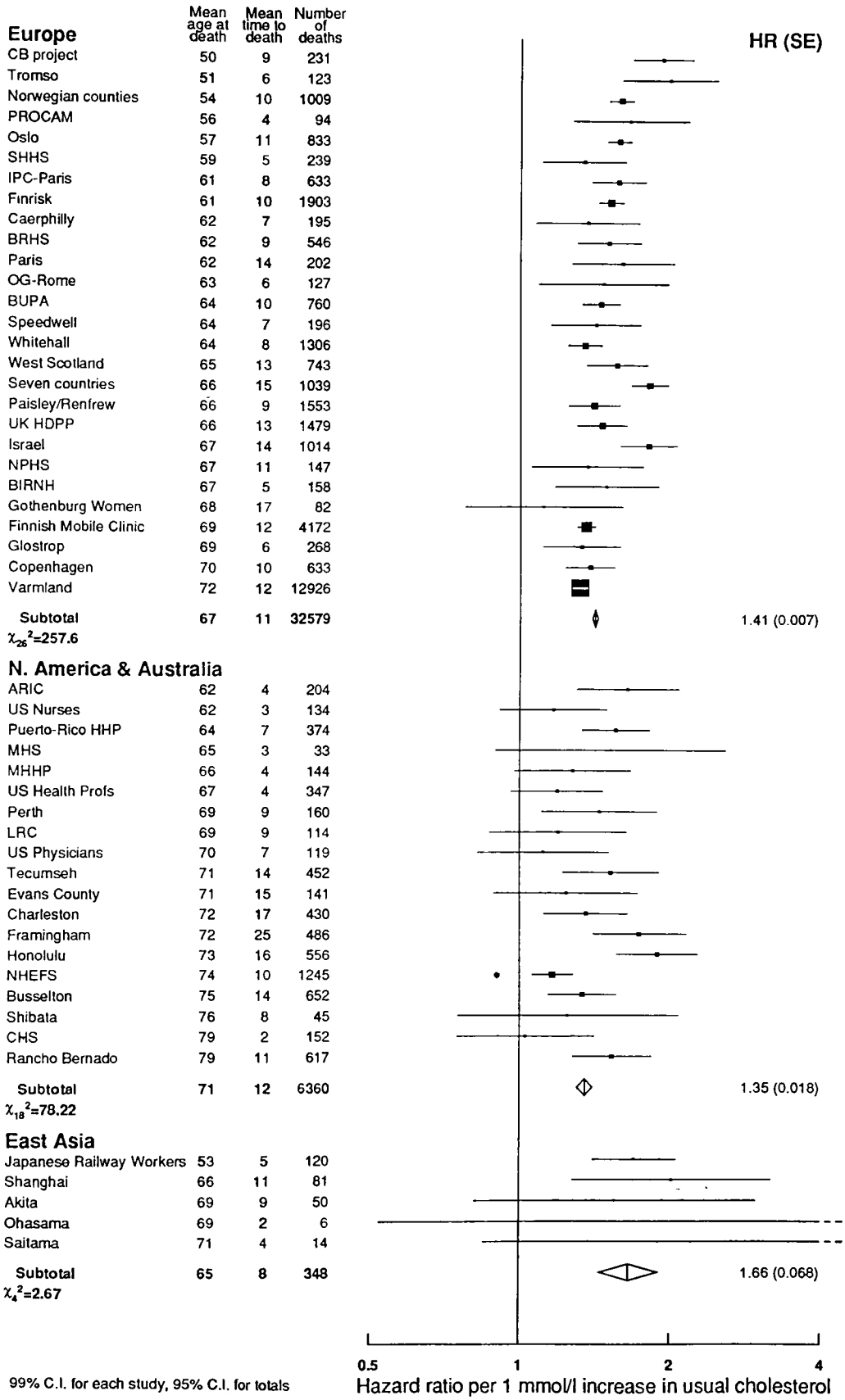


Figure 24: Strength of the study-specific associations between *usual* total cholesterol and all stroke risk, with study- and duration-specific corrections for regression dilution

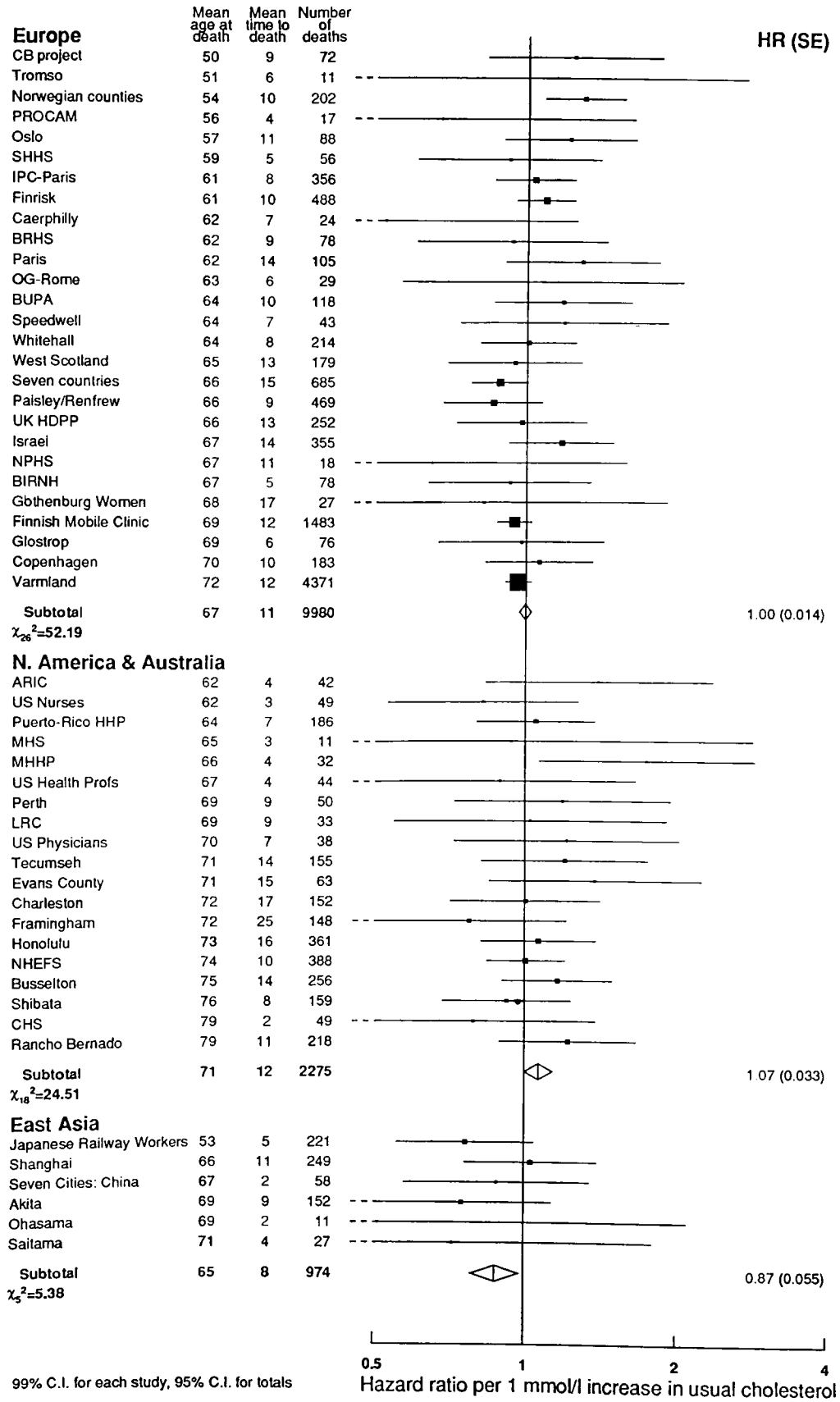


Table 1: Strength of the associations of *usual* total cholesterol with various causes of death

log hazard ratios per 1 mmol/l higher *usual** total cholesterol

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**64,002**	**0.232**	**0.005**	**1904.4**
IHD	**39,287**	**0.342**	**0.006**	**2847.0**
Acute MI	**14,421**	**0.412**	**0.010**	**1826.1**
Old MI	**6,479**	**0.308**	**0.016**	**379.1**
All CVA	13,229	0.001	0.013	0.0
Subarachnoid haem	974	0.005	0.045	0.0
Haem stroke	2,938	-0.064	0.027	5.6
Isch stroke	2,552	0.063	0.026	5.7
Unclassified stroke	6,765	0.001	0.018	0.0
Other vascular	**11,486**	**0.072**	**0.013**	**30.2**
Atherosclerosis	1,657	-0.052	0.031	2.9
Heart Failure	931	0.076	0.050	2.3
Aortic aneurysm	**813**	**0.297**	**0.045**	**43.2**
Chronic rheum. heart disease	368	-0.157	0.076	4.2
All cancer	**29,227**	**-0.074**	**0.008**	**75.0**
Lung	**6,473**	**-0.103**	**0.019**	**30.9**
Colorectal	2,883	-0.018	0.027	0.4
Stomach	**2,163**	**-0.153**	**0.033**	**21.1**
Prostate	1,383	-0.063	0.041	2.4
Breast	1,341	0.000	0.039	0.0
Pancreatic	1,297	0.056	0.040	2.0
Leukaemia	**779**	**-0.176**	**0.054**	**10.6**
Brain	688	0.089	0.053	2.8
Bladder	678	-0.059	0.057	1.0
Renal	606	0.044	0.058	0.6
Ovarian	495	-0.027	0.063	0.2
Melanoma	312	0.044	0.079	0.3
"Alcohol related" cancers				
Upper aerodigestive	**1,503**	**-0.141**	**0.039**	**13.1**
Oesophagus	**776**	**-0.176**	**0.055**	**10.3**
Oral	472	-0.075	0.069	1.2
Larynx	255	-0.163	0.095	3.0
Liver	**660**	**-0.342**	**0.063**	**29.2**
COPD	**2,581**	**-0.277**	**0.030**	**86.6**
Suicide & trauma	2,057	-0.052	0.030	3.1
Pneumonia	**1,502**	**-0.116**	**0.039**	**8.8**
Liver	**1,467**	**-0.197**	**0.040**	**24.6**
Renal	914	0.011	0.047	0.1
Diabetes	**747**	**0.269**	**0.048**	**32.1**
Peptic ulcer	376	0.024	0.076	0.1
Dementia	271	-0.052	0.091	0.3
TB	**236**	**-0.679**	**0.107**	**40.1**
Septicaemia	224	-0.077	0.099	0.6
Parkinsons	200	-0.134	0.107	1.6
All except vascular, renal, liver	**68,755**	**-0.074**	**0.006**	**172.9**
All causes	**136,466**	**0.079**	**0.004**	**427.2**

* Mean time to death=11 years. Therefore, RDR=0.63

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 2: Strength of the associations of *usual* non-HDL cholesterol with various causes of death, log hazard ratios per 1 mmol/l higher *usual** non-HDL cholesterol

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**5,646**	**0.377**	**0.018**	**454.0**
IHD	**3,492**	**0.486**	**0.020**	**595.0**
Acute MI	**2,305**	**0.485**	**0.024**	**399.6**
Old MI	**1,093**	**0.470**	**0.037**	**162.1**
All CVA	974	0.057	0.049	1.4
Subarachnoid haem	69	0.005	0.180	0.0
Haem stroke	241	-0.083	0.098	0.7
Isch stroke	156	0.125	0.119	1.1
Unclassified stroke	508	0.113	0.068	2.7
Other vascular	**1,180**	**0.211**	**0.042**	**24.7**
Atherosclerosis	35	-0.108	0.268	0.2
Heart Failure	143	-0.059	0.136	0.2
Aortic aneurysm	**123**	**0.440**	**0.104**	**17.8**
Chronic rheum. heart disease	63	0.177	0.173	1.0
All cancer	**4,017**	**-0.085**	**0.024**	**12.5**
Lung	1,035	-0.075	0.047	2.5
Colorectal	450	-0.029	0.070	0.2
Stomach	249	-0.150	0.101	2.2
Prostate	181	-0.250	0.119	4.4
Breast	216	-0.060	0.102	0.3
Pancreatic	244	-0.017	0.097	0.0
Leukaemia	121	-0.083	0.139	0.4
Brain	123	0.011	0.132	0.0
Bladder	110	0.019	0.145	0.0
Renal	104	-0.102	0.147	0.5
Ovarian	60	0.086	0.179	0.2
Melanoma	60	0.222	0.177	1.6
"Alcohol related" cancers				
Upper aerodigestive	171	-0.150	0.117	1.6
Oesophagus	106	0.007	0.146	0.0
Oral	41	-0.281	0.245	1.3
Larynx	24	-0.653	0.329	3.9
Liver	79	-0.343	0.181	3.6
COPD	**343**	**-0.583**	**0.079**	**54.4**
Suicide & trauma	184	-0.163	0.115	2.0
Pneumonia	184	-0.191	0.122	2.5
Liver	**173**	**-0.580**	**0.101**	**32.8**
Renal	112	0.154	0.140	1.2
Diabetes	**126**	**0.399**	**0.115**	**12.0**
Peptic ulcer	44	0.177	0.220	0.6
Dementia	32	-0.181	0.274	0.4
TB	14	-0.967	0.475	4.1
Septicaemia	41	-0.135	0.256	0.3
Parkinsons	21	-0.087	0.328	0.1
All except vascular, renal, liver	**6,928**	**-0.087**	**0.019**	**21.9**
All causes	**14,028**	**0.140**	**0.012**	**129.4**

* Mean time to death=11 years. Therefore, RDR=0.59

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 3: Strength of the associations of *usual* HDL cholesterol with various causes of death

log hazard ratios per 0.1 mmol/l higher *usual** HDL cholesterol

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**5,646**	**-0.112**	**0.007**	**279.2**
IHD	**3,492**	**-0.147**	**0.009**	**276.5**
Acute MI	**2,305**	**-0.140**	**0.011**	**165.6**
Old MI	**1,093**	**-0.153**	**0.016**	**96.4**
All CVA	974	-0.031	0.015	4.4
Subarachnoid haem	69	0.005	0.050	0.0
Haem stroke	241	0.010	0.026	0.1
Isch stroke	156	-0.078	0.037	4.6
Unclassified stroke	508	-0.046	0.022	4.3
Other vascular	**1,180**	**-0.087**	**0.014**	**38.8**
Atherosclerosis	35	-0.072	0.080	0.8
Heart Failure	143	-0.051	0.038	1.8
Aortic aneurysm	**123**	**-0.300**	**0.050**	**36.0**
Chronic rheum. heart disease	63	-0.163	0.065	6.3
All cancer	**4,017**	**-0.027**	**0.007**	**14.5**
Lung	1,035	-0.012	0.014	0.8
Colorectal	450	0.002	0.020	0.0
Stomach	249	-0.074	0.030	6.1
Prostate	181	0.039	0.025	2.3
Breast	216	-0.065	0.027	5.6
Pancreatic	244	-0.042	0.029	2.1
Leukaemia	**121**	**-0.129**	**0.045**	**8.3**
Brain	123	-0.030	0.040	0.5
Bladder	110	-0.035	0.046	0.6
Renal	104	-0.066	0.045	2.1
Ovarian	60	0.009	0.048	0.0
Melanoma	60	0.000	0.057	0.0
"Alcohol related" cancers				
Upper aerodigestive	**171**	**0.130**	**0.022**	**33.6**
Oesophagus	**106**	**0.112**	**0.034**	**10.9**
Oral	**41**	**0.112**	**0.041**	**7.3**
Larynx	**24**	**0.261**	**0.056**	**22.1**
Liver	79	-0.003	0.051	0.0
COPD	**343**	**0.082**	**0.013**	**37.5**
Suicide & trauma	184	0.045	0.019	5.6
Pneumonia	184	0.059	0.031	3.6
Liver	**173**	**0.056**	**0.013**	**17.4**
Renal	112	-0.052	0.045	1.4
Diabetes	**126**	**-0.316**	**0.048**	**42.5**
Peptic ulcer	44	0.001	0.070	0.0
Dementia	32	0.027	0.075	0.1
TB	14	-0.174	0.138	1.6
Septicaemia	41	0.077	0.066	1.4
Parkinsons	21	-0.114	0.103	1.2
All except vascular, renal, liver	6,928	-0.011	0.005	4.1
All causes	**14,028**	**-0.051**	**0.004**	**168.5**

* Mean time to death=11 years. Therefore, RDR=0.65

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 4: Strength of the associations of *usual* total cholesterol with various causes of death, after removing deaths within 5 years of the cholesterol measurement

log hazard ratios per 1 mmol/l higher *usual** total cholesterol

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**52,443**	**0.231**	**0.006**	**1528.3**
IHD	**32,241**	**0.332**	**0.007**	**2163.5**
Acute MI	**11,207**	**0.401**	**0.011**	**1305.8**
Old MI	**5,123**	**0.319**	**0.018**	**330.8**
All CVA	10,947	0.011	0.014	0.7
Subarachnoid haem	710	0.069	0.051	1.8
Haem stroke	2,124	-0.048	0.032	2.3
Isch stroke	2,085	0.046	0.029	2.4
Unclassified stroke	6,028	0.011	0.019	0.3
Other vascular	**9,255**	**0.092**	**0.015**	**39.3**
Atherosclerosis	1,389	-0.030	0.034	0.8
Heart Failure	746	0.059	0.056	1.1
Aortic aneurysm	**687**	**0.295**	**0.049**	**36.1**
Chronic rheum. heart disease	242	-0.152	0.094	2.6
All cancer	**22,944**	**-0.045**	**0.010**	**22.4**
Lung	**5,073**	**-0.084**	**0.021**	**16.1**
Colorectal	2,246	0.028	0.031	0.8
Stomach	**1,657**	**-0.114**	**0.038**	**9.2**
Prostate	1,211	-0.052	0.043	1.4
Breast	906	0.068	0.047	2.1
Pancreatic	972	0.068	0.046	2.2
Leukaemia	600	-0.110	0.061	3.3
Brain	484	0.117	0.063	3.5
Bladder	566	-0.062	0.063	1.0
Renal	**460**	**0.168**	**0.064**	**6.9**
Ovarian	343	-0.051	0.077	0.4
Melanoma	239	-0.052	0.093	0.3
"Alcohol related" cancers				
Upper aerodigestive	**1,204**	**-0.138**	**0.044**	**9.9**
Oesophagus ca.	**620**	**-0.197**	**0.062**	**10.2**
Oral	375	-0.050	0.076	0.4
Larynx	209	-0.127	0.105	1.5
Liver	**515**	**-0.248**	**0.070**	**12.4**
COPD	**2,171**	**-0.272**	**0.032**	**70.0**
Suicide & trauma	1,510	-0.022	0.034	0.4
Pneumonia	1,313	-0.101	0.042	5.9
Liver	**1,132**	**-0.115**	**0.045**	**6.6**
Renal	729	0.021	0.052	0.2
Diabetes	**598**	**0.275**	**0.053**	**27.0**
Peptic ulcer	288	0.163	0.083	3.9
Dementia	255	-0.027	0.093	0.1
TB	**168**	**-0.617**	**0.127**	**23.7**
Septicaemia	167	0.020	0.112	0.0
Parkinsons	186	-0.127	0.111	1.3
All except vascular, renal, liver	**55,394**	**-0.051**	**0.006**	**65.7**
All causes	**110,655**	**0.091**	**0.004**	**457.4**

* Mean time to death=11 years. Therefore, RDR=0.63

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 5: Strength of the associations of *usual* non-HDL cholesterol with various causes of death, after removing deaths within 5 years of the cholesterol measurement

log hazard ratios per 1 mmol/l higher *usual** non-HDL cholesterol

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**3,663**	**0.381**	**0.022**	**308.1**
IHD	**2,292**	**0.483**	**0.024**	**392.2**
Acute MI	**1,514**	**0.466**	**0.030**	**236.7**
Old MI	**738**	**0.508**	**0.042**	**143.3**
All CVA	637	0.090	0.060	2.2
Subarachnoid haem	32	0.247	0.242	1.0
Haem stroke	134	0.028	0.130	0.0
Isch stroke	94	0.052	0.158	0.1
Unclassified stroke	377	0.107	0.079	1.8
Other vascular	**734**	**0.208**	**0.053**	**15.2**
Atherosclerosis	23	-0.120	0.337	0.1
Heart Failure	92	-0.148	0.169	0.8
Aortic aneurysm	**86**	**0.423**	**0.120**	**12.5**
Chronic rheum. heart disease	32	0.085	0.248	0.1
All cancer	**2,481**	**-0.021**	**0.030**	**0.5**
Lung	639	-0.013	0.059	0.0
Colorectal	276	0.026	0.088	0.1
Stomach	154	-0.011	0.124	0.0
Prostate	121	-0.178	0.142	1.6
Breast	121	0.066	0.130	0.3
Pancreatic	148	0.023	0.123	0.0
Leukaemia	71	-0.080	0.179	0.2
Brain	75	-0.017	0.169	0.0
Bladder	73	-0.133	0.183	0.5
Renal	64	0.065	0.182	0.1
Ovarian	31	0.056	0.252	0.0
Melanoma	39	0.180	0.219	0.7
"Alcohol related" cancers				
Upper aerodigestive	107	-0.044	0.145	0.1
Oesophagus	67	0.139	0.173	0.6
Oral	26	-0.044	0.294	0.0
Larynx	14	-1.116	0.454	6.1
Liver	46	-0.103	0.225	0.2
COPD	**226**	**-0.515**	**0.102**	**25.6**
Suicide & trauma	103	-0.024	0.148	0.0
Pneumonia	129	-0.080	0.143	0.3
Liver	**120**	**-0.561**	**0.116**	**23.5**
Renal	80	0.239	0.159	2.3
Diabetes	**79**	**0.268**	**0.151**	**3.2**
Peptic ulcer	28	0.513	0.250	4.2
Dementia	25	-0.114	0.300	0.1
TB	9	-1.030	0.615	2.8
Septicaemia	28	-0.225	0.311	0.5
Parkinsons	15	0.095	0.362	0.1
All except vascular, renal, liver	4,431	-0.037	0.023	2.5
All causes	**9,169**	**0.164**	**0.015**	**117.5**

* Mean time to death=11 years. Therefore, RDR=0.59

Results with p<0.01 (chi-square>6.6) are in boid

All analyses are adjusted for cohort, sex and age

Table 6: Strength of the associations of *usual* HDL cholesterol with various causes of death, after removing deaths within 5 years of the cholesterol measurement

log hazard ratios per 0.1 mmol/l higher *usual** HDL cholesterol

Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square
All vascular	**3,663**	**-0.115**	**0.009**	**182.2**
IHD	**2,292**	**-0.145**	**0.011**	**169.6**
Acute MI	**1,514**	**-0.155**	**0.014**	**123.4**
Old MI	**738**	**-0.130**	**0.019**	**47.2**
All CVA	637	-0.076	0.020	14.2
Subarachnoid haem	32	-0.053	0.081	0.4
Haem stroke	134	-0.063	0.042	2.2
Isch stroke	94	-0.105	0.049	4.6
Unclassified stroke	377	-0.074	0.027	7.4
Other vascular	**734**	**-0.066**	**0.018**	**13.8**
Atherosclerosis	23	-0.040	0.098	0.2
Heart Failure	92	-0.041	0.049	0.7
Aortic aneurysm	**86**	**-0.282**	**0.060**	**22.4**
Chronic rheum. heart disease	32	-0.227	0.098	5.3
All cancer	**2,481**	**-0.033**	**0.009**	**12.6**
Lung	639	-0.012	0.018	0.4
Colorectal	276	-0.021	0.028	0.6
Stomach	154	-0.117	0.041	8.1
Prostate	121	0.033	0.033	1.0
Breast	121	-0.044	0.036	1.5
Pancreatic	148	-0.048	0.039	1.5
Leukaemia	**71**	**-0.134**	**0.059**	**5.1**
Brain	75	-0.045	0.053	0.7
Bladder	73	-0.038	0.059	0.4
Renal	64	-0.084	0.061	1.9
Ovarian	31	0.022	0.065	0.1
Melanoma	39	-0.040	0.076	0.3
"Alcohol related" cancers				
Upper aerodigestive	**107**	**0.111**	**0.031**	**13.1**
Oesophagus	**67**	**0.099**	**0.046**	**4.6**
Oral	**26**	**0.036**	**0.074**	**0.2**
Larynx	**14**	**0.313**	**0.070**	**20.0**
Liver	46	0.008	0.067	0.0
COPD	**226**	**0.079**	**0.017**	**22.0**
Suicide & trauma	103	0.025	0.041	0.4
Pneumonia	129	0.019	0.040	0.2
Liver	**120**	**0.059**	**0.013**	**22.2**
Renal	80	-0.068	0.055	1.5
Diabetes	**79**	**-0.206**	**0.058**	**12.8**
Peptic ulcer	28	0.087	0.085	1.0
Dementia	25	0.034	0.084	0.2
TB	9	-0.375	0.187	4.0
Septicaemia	28	0.134	0.078	2.9
Parkinsons	15	-0.075	0.117	0.4
All except vascular, renal, liver	4,431	-0.014	0.007	3.9
All causes	**9,169**	**-0.052**	**0.005**	**112.1**

* Mean time to death=11 years. Therefore, RDR=0.65

Results with p<0.01 (chi-square>6.6) are in bold

All analyses are adjusted for cohort, sex and age

Table 7: Strength of the associations of *usual* total cholesterol with various causes of death, after removing deaths within 5 years of the cholesterol measurement

For those individuals with duration, sex, status, SBP, DBP, total cholesterol AND smoking (Y/N), alcohol (Y/N), height and weight

log hazard ratios per 1 mmol/l higher *usual total cholesterol**

		Adjusted for cohort, age and sex			Adjusted for cohort, age, sex, smoking (Y/N), alcohol (Y/N), weight (relative to height)		
Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square	log(HR)	SE(log(HR))	chi-square
All vascular	**11,956**	**0.223**	**0.012**	**336.7**	**0.209**	**0.012**	**287.4**
IHD	**7,088**	**0.327**	**0.014**	**511.6**	**0.314**	**0.015**	**440.6**
Acute MI	**4,153**	**0.355**	**0.019**	**347.7**	**0.344**	**0.020**	**306.8**
Old MI	**2,409**	**0.290**	**0.026**	**127.4**	**0.277**	**0.026**	**112.0**
All CVA	2,411	0.035	0.030	1 4	0.027	0.030	0.8
Subarachnoid haem	163	0.044	0.113	0.1	0.012	0.114	0.0
Haem stroke	453	-0.074	0.071	1.1	-0.100	0.072	1.9
Isch stroke	599	0.144	0.058	6.2	0.141	0.059	5.7
Unclassified stroke	1,196	0.019	0.042	0.2	0.018	0.042	0.2
Other vascular	2,457	0.044	0.029	2.4	0.027	0.029	0.9
Atherosclerosis	150	0.062	0.119	0.3	0.106	0.125	0.7
Heart Failure	372	-0.114	0.082	1.9	-0.145	0.083	3.0
Aortic aneurysm	**314**	**0.237**	**0.074**	**10.4**	**0.233**	**0.078**	**8.8**
Chronic rheum. heart disea	89	-0.270	0.156	3.0	-0.270	0.160	2.9
All cancer	**7,823**	**-0.085**	**0.017**	**25.8**	**-0.083**	**0.017**	**24.1**
Lung	**1,945**	**-0.143**	**0.034**	**17.8**	**-0.108**	**0.034**	**10.1**
Colorectal	917	0.019	0.048	0 2	0.007	0.048	0.0
Stomach	677	-0.061	0.060	1 1	-0.058	0.060	0.9
Prostate	558	-0.050	0.063	0 6	-0.062	0.064	1.0
Breast	316	0.087	0.076	1.3	0.083	0.076	1.2
Pancreatic	383	-0.043	0.075	0.3	-0.035	0.075	0.2
Leukaemia	218	-0.253	0.103	6.1	-0.238	0.104	5.2
Brain	169	0.121	0.106	1.3	0.128	0.106	1.5
Bladder	221	-0.131	0 100	1.7	-0.129	0.102	1.6
Renal	187	-0.109	0.109	1 0	-0.125	0.110	1.3
Ovarian	103	-0.106	0.141	0 6	-0.118	0.142	0.7
Melanoma	84	0.064	0.150	0 2	0.043	0.153	0.1
"Alcohol related" cancers							
Upper aerodigestive	**441**	**-0.307**	**0.075**	**16.9**	**-0.265**	**0.076**	**12.3**
Oesophagus	261	-0.221	0.096	5.3	-0.186	0.097	3.6
Oral	**137**	**-0.388**	**0.135**	**8.3**	**-0.355**	**0.137**	**6.7**
Larynx	43	-0.598	0.251	5 7	-0.472	0.251	3.5
Liver	**172**	**-0.357**	**0.120**	**8.8**	**-0.397**	**0.121**	**10.8**
COPD	**892**	**-0.253**	**0.051**	**24.8**	**-0.160**	**0.051**	**9.8**
Suicide & trauma	379	-0.049	0.073	0.4	-0.052	0.074	0.5
Pneumonia	627	-0.070	0.060	1.4	-0.051	0.061	0.7
Liver	**400**	**-0.231**	**0.077**	**9.0**	**-0.305**	**0.078**	**15.2**
Renal	297	0.118	0.081	2.1	0.071	0.082	0.8
Diabetes	**301**	**0.251**	**0.076**	**11.0**	**0.222**	**0.078**	**8.1**
Peptic ulcer	101	0.304	0.132	5 3	0.323	0.138	5.4
Dementia	147	-0.164	0.128	1.7	-0.166	0.130	1 6
TB	**56**	**-0.818**	**0.227**	**13.0**	**-0.628**	**0.229**	**7.5**
Septicaemia	86	-0.020	0.155	0.0	-0.060	0.153	0.2
Parkinsons	105	-0.177	0.149	1.4	-0.189	0.151	1.6
All except vascular, renal, liver	**16,149**	**-0.075**	**0.012**	**41.4**	**-0.072**	**0.012**	**37.7**
All causes	**29,725**	**0.060**	**0.008**	**52.6**	**0.053**	**0.008**	**40.5**

* Mean time to death=11 years. Therefore, RDR=0.63

Results with p<0 01 (chi-square>6 6) are in bold

All analyses are adjusted for cohort, sex and age

Table 8: Strength of the associations of *usual* non-HDL cholesterol with various causes of death, after removing deaths within 5 years of the cholesterol measurement

For those individuals with duration, sex, status, SBP, DBP, total cholesterol AND smoking (Y/N), alcohol (Y/N), height and weight

log hazard ratios per 1 mmol/l higher *usual non-HDL cholesterol**

		Adjusted for cohort, age and sex			Adjusted for cohort, age, sex, smoking (Y/N), alcohol (Y/N), weight (relative to height)		
Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square	log(HR)	SE(log(HR))	chi-square
All vascular	**2,132**	**0.322**	**0.028**	**131.8**	**0.291**	**0.029**	**102.2**
IHD	**1,307**	**0.410**	**0.032**	**162.9**	**0.381**	**0.034**	**127.1**
Acute MI	**809**	**0.388**	**0.041**	**89.8**	**0.345**	**0.042**	**67.5**
Old MI	**476**	**0.451**	**0.053**	**73.2**	**0.464**	**0.060**	**60.6**
All CVA	339	0.061	0.081	0.6	0.035	0.082	0.2
Subarachnoid haem	19	0.076	0.336	0.1	-0.016	0.340	0.0
Haem stroke	81	-0.019	0.167	0.0	-0.060	0.174	0.1
Isch stroke	72	0.077	0.179	0.2	0.026	0.185	0.0
Unclassified stroke	167	0.091	0.113	0.6	0.086	0.115	0.6
Other vascular	**486**	**0.196**	**0.064**	**9.5**	0.158	0.066	5.8
Atherosclerosis	14	-0.012	0.441	0.0	0.046	0.470	0.0
Heart Failure	62	-0.314	0 203	2.4	-0.440	0.220	4.0
Aortic aneurysm	**69**	**0.389**	**0.135**	**8.3**	0.429	0.168	6.5
Chronic rheum. heart disea	21	0.223	0.269	0.7	0.207	0.275	0.6
All cancer	1,711	-0.036	0.036	1.0	-0.042	0.037	1.3
Lung	492	-0.050	0.067	0.6	-0.027	0.068	0.2
Colorectal	196	-0.025	0.104	0.1	-0.059	0.107	0.3
Stomach	87	0.127	0.158	0.7	0.132	0.161	0.7
Prostate	90	-0.244	0.162	2.3	-0 223	0.170	1.7
Breast	81	0.084	0 160	0.3	0.072	0.164	0.2
Pancreatic	92	-0 220	0 162	1.8	-0.219	0.164	1.8
Leukaemia	40	-0 024	0.234	0.0	-0.001	0.245	0.0
Brain	50	0.053	0.201	0.1	0.115	0.205	0.3
Bladder	50	-0.190	0.216	0.8	-0.253	0.226	1.3
Renal	49	-0 211	0.215	1.0	-0.226	0.214	1.1
Ovarian	22	0.181	0 298	0.4	0.207	0.314	0.4
Melanoma	28	0.196	0.248	0.6	0.269	0.266	1.0
"Alcohol related" cancers							
Upper aerodigestive	86	0.044	0.158	0.1	0.075	0.160	0.2
Oesophagus	58	0.269	0.164	2.7	0.285	0.188	2.3
Oral	17	-0.098	0.370	0.1	0.001	0.359	0.0
Larynx	11	-1.313	0.515	6.5	-1.048	0.535	3 8
Liver	32	0.048	0.251	0.0	-0.068	0.251	0.1
COPD	**166**	**-0.483**	**0.117**	**17.1**	**-0.363**	**0.125**	**8.4**
Suicide & trauma	72	-0.099	0.180	0.3	-0.132	0.182	0.5
Pneumonia	86	-0.216	0.178	1.5	-0.239	0 182	1.7
Liver	**81**	**-0.448**	**0.151**	**8.8**	**-0.482**	**0.126**	**14.7**
Renal	41	0 221	0.205	1.2	0.163	0.222	0.5
Diabetes	47	0.354	0.174	4.1	0.238	0.192	1 5
Peptic ulcer	12	0.555	0.358	2.4	0.359	0 363	1 0
Dementia	18	-0.284	0.374	0.6	-0.321	0.388	0.7
TB	6	-1.420	0.767	3.4	-0.814	0 759	1.1
Septicaemia	11	-0.049	0.456	0.0	-0.108	0.465	0 1
Parkinsons	10	-0.399	0.475	0.7	-0.228	0 480	0.2
All except vascular, renal, liver	2,869	-0.065	0.028	5.2	-0.072	0.029	6.2
All causes	**5,957**	**0.123**	**0.018**	**44.4**	**0.101**	**0.019**	**28.5**

* Mean time to death=11 years Therefore, RDR=0.59

Results with p<0.01 (chi-square>6 6) are in bold

All analyses are adjusted for cohort, sex and age

Table 9: Strength of the associations of *usual* HDL cholesterol with various causes of death, after removing deaths within 5 years of the cholesterol measurement

For those individuals with duration, sex, status, SBP, DBP, total cholesterol AND smoking (Y/N), alcohol (Y/N), height and weight

log hazard ratios per 0.1 mmol/l higher *usual HDL cholesterol**

		Adjusted for cohort, age and sex			Adjusted for cohort, age, sex, smoking (Y/N), alcohol (Y/N), weight (relative to height)		
Cause of death	No of deaths	log(HR)	SE(log(HR))	chi-square	log(HR)	SE(log(HR))	chi-square
All vascular	**2,132**	**-0.106**	**0.010**	**105.7**	**-0.077**	**0.011**	**51.5**
IHD	**1,307**	**-0.135**	**0.014**	**97.7**	**-0.106**	**0.014**	**55.7**
Acute MI	**809**	**-0.144**	**0.018**	**67.9**	**-0.111**	**0.018**	**36.8**
Old MI	**476**	**-0.122**	**0.022**	**29.8**	**-0.102**	**0.023**	**19.4**
All CVA	339	-0.061	0.025	6.1	-0.035	0.025	1.9
Subarachnoid haem	19	-0.080	0.102	0.6	-0.053	0.104	0.3
Haem stroke	81	-0.056	0.051	1.2	-0.028	0.053	0.3
Isch stroke	72	-0.118	0.055	4.7	-0.087	0.056	2.4
Unclassified stroke	167	-0.037	0.035	1.1	-0.014	0.035	0.1
Other vascular	**486**	**-0.069**	**0.021**	**11.1**	-0.037	0.021	3.0
Atherosclerosis	14	0.096	0.116	0.7	0.051	0.130	0.2
Heart Failure	62	-0.036	0.055	0.4	0.021	0.059	0.1
Aortic aneurysm	**69**	**-0.286**	**0.065**	**19.2**	**-0.274**	**0.071**	**15.0**
Chronic rheum. heart dise	**21**	**-0.308**	**0.119**	**6.7**	-0 285	0.125	5.2
All cancer	**1,711**	**-0.029**	**0.011**	**7.1**	-0.025	0.011	5.2
Lung	492	-0.004	0.020	0.0	-0.004	0.020	0.0
Colorectal	196	-0.011	0 031	0 1	-0.004	0.033	0.0
Stomach	87	-0.131	0 052	6.3	-0.122	0.053	5.3
Prostate	90	0.038	0 031	1 5	0.065	0.048	1.9
Breast	81	-0.083	0.045	3 3	-0.074	0.047	2.5
Pancreatic	92	-0.050	0.048	1.1	-0.050	0.050	1.0
Leukaemia	40	-0.152	0.077	3 9	-0.194	0.082	5.6
Brain	50	-0.067	0.065	1.1	-0.121	0.071	2.9
Bladder	50	-0.030	0.067	0.2	-0.018	0.069	0.1
Renal	49	-0.042	0.066	0 4	-0.040	0.068	0.4
Ovarian	22	0.015	0.078	0 0	0 017	0.085	0.0
Melanoma	28	-0.043	0.089	0.2	-0.088	0.094	0.9
"Alcohol related" cancers							
Upper aerodigestive	**86**	**0.140**	**0.040**	**12.4**	**0.122**	**0.042**	**8.5**
Oesophagus	58	0.086	0.053	2.6	0.093	0.054	3.0
Oral	17	0.134	0 085	2.5	0.082	0.097	0.7
Larynx	**11**	**0.311**	**0.074**	**17.7**	**0.285**	**0.089**	**10.3**
Liver	32	-0.026	0.078	0.1	0.026	0.078	0.1
COPD	**166**	**0.080**	**0.018**	**19.2**	**0.114**	**0.028**	**16.5**
Suicide & trauma	72	0.004	0.053	0.0	0.022	0.056	0.2
Pneumonia	86	0.001	0 048	0.0	0.002	0.050	0.0
Liver	**81**	**0.056**	**0.014**	**16.3**	**0.051**	**0.013**	**15.4**
Renal	41	-0.098	0.070	2.0	-0.059	0.073	0.7
Diabetes	**47**	**-0.200**	**0.069**	**8.4**	**-0.102**	**0.073**	**2.0**
Peptic ulcer	12	0.062	0.125	0.2	0.114	0.129	0.8
Dementia	18	0.082	0.095	0.7	0 082	0.104	0.6
TB	6	-0.467	0.204	5.3	-0.360	0.174	4.3
Septicaemia	11	0.164	0.093	3.1	0.218	0.101	4.7
Parkinsons	10	0.008	0.120	0.0	-0.040	0.128	0.1
All except vascular, renal, liver	2,869	-0.010	0.008	1.4	-0.006	0.009	0.5
All causes	**5,957**	**-0.047**	**0.006**	**66.6**	**-0.031**	**0.006**	**26.8**

* Mean time to death=11 years Therefore, RDR=0 65

Results with p<0 01 (chi-square>6 6) are in bold

All analyses are adjusted for cohort, sex and age

Chapter 9: Discussion

The basis of this thesis has been the worldwide collaboration between investigators of prospective observational studies (which I have co-ordinated since its inception) to establish a centralised collection of data on individuals in these studies. Analytically, the aim has been to develop and to use appropriate statistical techniques to assess the age- and sex-specific associations of usual blood pressure and of usual cholesterol with cause-specific mortality. Since the data set is uniquely large, and because appropriate methods of analysis (with fuller account taken of the time-dependent nature of the regression dilution bias) have been developed and used, the truer associations have been established reliably. An integral part of the methodological element of the thesis was to investigate the systematic underestimation of associations between risk factor and disease that are obtained when only a single baseline measurement is used to assess risk factor levels (the regression dilution bias). As part of this, the extent of the bias was investigated in those studies that had re-measured risk factor levels during follow-up. One particularly novel aspect has been the emphasis on, and methods developed to account for, the regression dilution bias in several studies simultaneously and in an appropriately time-dependent way.

1 The regression dilution bias

Unless corrections are made for regression dilution, the strength of any associations between *usual* blood pressure or *usual* cholesterol and disease risk obtained from prospective observational studies, will be underestimated by about one-third during the first decade of follow-up, one-half during the second decade and as much as two-

thirds during the third decade. The increasing magnitude of the underestimation with increasing follow-up means that such corrections must be appropriately time-dependent. In Chapter 5, the time-dependent magnitude of the regression dilution ratio (*RDR*) during prolonged follow-up was determined for blood pressure and total cholesterol using repeated measurements of these factors from all of the studies in the PSC that were able to provide such data. These analyses showed how, irrespective of quality, size or other study characteristics or circumstances, the analyses of all studies should include substantial correction for the regression dilution bias. Furthermore, the remarkable similarity in the magnitude of the *RDR* between cohorts for any given interval between measurements, and the general trend within cohorts for more extreme *RDR*s with increasing follow-up (and thus greater underestimation of the risk relationships if only baseline measurements are used) confirm the need for period-specific (i.e. time-dependent) correction factors. Due to the remarkable similarity in the results from different cohorts, simple period-specific correction procedures could be developed that were widely applicable to the analysis of all observational studies and their overviews (Chapter 6). These analyses suggested that for systolic blood pressure (SBP) and total cholesterol, the *RDR* was approximately 70% within 5 years of the baseline measurement, and decreasing by about 1% in each subsequent year of follow-up. The *RDR* for diastolic blood pressure (DBP) was rather more extreme, being approximately 60% within 5 years of the baseline measurement, and decreasing by about 2% per year thereafter. These non-parametric estimates were independent of any assumptions about the constancy of risks over time, about the biological mechanisms by which exposures affect disease, or about whether changes in the blood pressure and cholesterol were real or artefactual.

2 Blood pressure and premature death

The corrected analyses presented in Chapter 7 demonstrated approximately log-linear associations of usual blood pressure with risks of death from ischaemic heart disease (IHD), stroke (both ischaemic and haemorrhagic), heart failure, renal disease and diabetes. The relationships were continuous throughout the entire range of blood pressures studied, with no apparent levels above or below which blood pressure was no longer positively associated with risk. Although similar in men and women, the sizes of the associations were highly dependent on age at death, with the proportional relationships for both IHD and stroke being nearly twice as strong among younger people than among those who were older. At all ages, the risk associations were about 30% steeper for stroke than for IHD but, after appropriate stratification by age at death, were similar for haemorrhagic and ischaemic stroke. During an average 13 years of follow-up, each 10 mmHg lower usual SBP or about 5 mmHg lower usual DBP was associated with:

- 20-25% lower risk of death from stroke, and 15-20% lower risk of death from IHD at 70-79 years; and
- 35-40% lower risk of death from stroke, and 25-30% lower risk of death from IHD at 40-59 years.

3 Cholesterol and premature death

With data on 64,000 vascular deaths, 30,000 cancer deaths and 42,000 other deaths, it was possible in Chapter 8 to investigate more reliably than has previously been possible the relationships of cholesterol (total cholesterol, and where available, HDL and non-HDL cholesterol) with IHD, stroke and various non-vascular causes of

death. There was a strong, direct association of usual total cholesterol with risk of death from IHD that was approximately log-linear throughout the range studied. As with many risk factor-disease relationships, the proportional relevance of total cholesterol to IHD risk attenuated with age, but the association remained strong and positive even among those dying in their seventies. The associations with non-HDL cholesterol were somewhat stronger than those with total cholesterol. However, they did not attenuate with age to the same extent, being almost as important in old age as in early middle age. As a consequence, although non-HDL cholesterol was only about 10% more important than total cholesterol at ages 40-59, it became about twice as important for those dying in their seventies. HDL cholesterol was inversely associated IHD risk, with no apparent levels (within the range studied) above or below which higher HDL cholesterol was no longer associated with lower IHD risk. Moreover, the strength of this relationship was similar in all age groups studied. These relationships were all similar in men and women within each age group.

Thus, over 13 years of follow-up, 25% lower risk of IHD death was associated with:

- 0.6 mmol/l lower usual total cholesterol or 0.5 mmol/l lower usual non-HDL cholesterol for men and women in middle age (40-59);
- 1.3 mmol/l lower total cholesterol or 0.7 mmol/l lower usual non-HDL cholesterol for men and women in old age (70-79); and
- 0.2 mmol/l higher usual HDL cholesterol for men and women at any age.

The relationships between cholesterol and stroke were weaker, and appeared to differ from one age group to another. Thus, only by studying the age-specific relationships between particular cholesterol fractions and stroke sub-types will the

real relationships emerge. There was no evidence within these data of any relationships between either HDL or non-HDL cholesterol and either of the stroke sub-types, but this lack of association may be due to small numbers. Thus, the real sex-, and more importantly, age-specific relationships between cholesterol fractions and stroke sub-types await the availability of yet more data.

Any inverse relationships with cancer mortality were due largely to negative associations with HDL cholesterol, which largely disappeared within 5 years of the cholesterol measurements (suggesting reverse-causality). The only relationship sustained beyond 5 years was that of HDL cholesterol with upper aero-digestive cancers. However, these cancers are strongly and positively related to alcohol consumption, which in turn is positively related to HDL cholesterol. Therefore, this relationship may be largely or wholly attributed to residual confounding with alcohol. There were no associations with deaths from suicide and violence, and the only other non-vascular relationships sustained beyond 5 years were those with chronic obstructive pulmonary disease (COPD). The risk of COPD-related mortality was positively associated with HDL cholesterol and negatively associated with non-HDL cholesterol. Although the strength of these relationships was somewhat reduced after adjusting for weight (relative to height), they remained substantial.

The large numbers of IHD deaths provided reliable evidence for the importance of HDL and non-HDL cholesterol to IHD mortality at different ages. Furthermore, large numbers of non-vascular deaths avoided spurious relationships being observed by the play of chance. However, despite this large data set, more evidence is needed to clarify the residual uncertainties about relationships between cholesterol fractions and stroke sub-types.

4 Conclusions and implications for epidemiology and public health

Worldwide, almost 20 million adult deaths will be caused by vascular diseases in the year 2000, with 12 million attributed to heart disease and 6 million to stroke. One of the fundamental principles of public health is that it is generally possible to save many lives by only moderate reductions in major causes of death. However, to achieve these moderate reductions, reliable quantitative information is needed about the important determinants of such causes. This thesis has illustrated the extent to which random error and inappropriate statistical analysis lead to misleading conclusions concerning the importance of blood pressure and blood cholesterol, particularly in premature death. Only by studying adequate numbers of deaths (136,000 deaths among 1 million adults during 13 million person-years of follow-up) and by using appropriate statistical techniques - taking proper account of (a) the regression dilution bias; (b) the full range of blood pressure and cholesterol; (c) the opposing effects of HDL and the remaining non-HDL cholesterol; and (d) age at death - did it become possible to investigate properly the real relationships between blood pressure, cholesterol fractions and vascular and other causes of death. These analyses have demonstrated reliably that, as causes of IHD death in early middle age, blood pressure and blood lipids are at least three times as important as many previous analyses had suggested, with no clinically relevant inverse associations with cancer or other non-vascular mortality (except, surprisingly, COPD).

REFERENCES

References

Anderson KM, Castelli WP, Levy D. Cholesterol and mortality. 30 years of follow-up from the Framingham study. JAMA 1987; 257(16):2176-80.

Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project--1981-86. 2. Incidence, case fatality rates and overall outcome at one year of cerebral infarction, primary intracerebral and subarachnoid haemorrhage. J Neurol Neurosurg Psychiatry 1990; 53(1):16-22.

Bashir SA, Duffy SW. The correction of risk estimates for measurement error. Ann Epidemiol 1997; 7(2):154-64.

Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses' health study. Am J Nurs 1978; 78(6):1039-40.

Breslow NE, Day NE. Statistical methods in cancer research. Volume II--The design and analysis of cohort studies. IARC Sci Publ 1987; (82):1-406.

Buring JE, Hebert P, Romero J *et al.* Migraine and subsequent risk of stroke in the Physicians' Health Study. Arch Neurol 1995; 52(2):129-34.

Carroll RJ, Stefanski LA. Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. Stat Med 1994; 13(12):1265-82.

Castelli WP, Doyle JT, Gordon T *et al.* Alcohol and blood lipids. The cooperative lipoprotein phenotyping study. Lancet 1977; 2(8030): 153-5.

References

Chen, Z., Peto R, Collins R *et al.* Serum cholesterol concentration and coronary heart disease in a population with low cholesterol concentrations. BMJ 1991; 303(6797): 276-82.

Clarke R, Shipley M, Lewington S *et al.* Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. Am Epidemiol [*in press*].

Cox DR. Regression models and life tables (with discussion). J R Statist Soc 1972; 34:187-220.

Daviglus ML, Orencia AJ, Dyer AR *et al.* Dietary vitamin C, beta-carotene and 30-year risk of stroke: results from the Western Electric Study. Neuroepidemiology 1997; 16(2):69-77.

Delahaye F, Bruckert E, Thomas D, Emmerich J, Richard JL. [Serum cholesterol and cancer. Is there a causal relationship?]. Arch Mal Coeur Vaiss 1992; 85 Spec No 3:37-45.
Original Title: Cholesterolemie et cancer: s'agit-il d'une relation de cause a effet?

Duffy SW, Maximovitch DM, Day NE. External validation, repeat determination, and precision of risk estimation in misclassified exposure data in epidemiology. J Epidemiol Community Health 1992; 46(6):620-4.

Duffy SW, Rohan TE, Day NE. Misclassification in more than one factor in a case-control study: a combination of Mantel-Haenszel and maximum likelihood approaches. Stat Med 1989; 8(12):1529-36.

Eastern Stroke and Coronary Heart Disease Collaborative Research Group. Blood pressure, cholesterol and stroke in Eastern Asia. Lancet 1998; 352:1801-07.

Easton DF, Peto J, Babiker A. Floating absolute risk: an alternative to relative risk in survival and case-control analysis avoiding a reference group. Stat Med 1991; 10:1025-35.

Elton RA, Duffy SW. Correcting for the effect of misclassification bias in a case-control study using data from two different questionnaires. Biometrics 1983; 39(3):659-63.

Frost C, Thompson S. Correcting for regression dilution bias: comparison of methods for a single predictor variable. Stat Med [*in press*].

Galton F. Regression towards mediocrity in hereditary stature. Journal of the Anthropological Institute 1886; 15:246-63.

Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. Am J Med 1977; 62(5):707-14.

Gordon T, Doyle JT. Drinking and mortality. The Albany Study. Am J Epidemiol 1987; 125(2):263-70.

Jacobs D, Blackburn H, Higgins M *et al.* Report of the Conference on Low Blood Cholesterol: Mortality Associations. Circulation 1992; 86(3):1046-60.

References

Kaplan JR, Muldoon MF, Manuck SB, Mann JJ. Assessing the observed relationship between low cholesterol and violence-related mortality. Implications for suicide risk. Ann N Y Acad Sci 1997; 836:57-80.

Kodama K, Sasaki H, Shimizu Y. Trend of coronary heart disease and its relationship to risk factors in a Japanese population: a 26-year follow-up, Hiroshima/Nagasaki study. Jpn Circ J 1990; 54(4):414-21.

Law MR, Wald NJ, Wu T *et al.*. Systematic underestimation of association between serum cholesterol concentration and ischaemic heart disease in observational studies: Data from the BUPA study. BMJ 1994; 308(6925): 363-66.

Lichtenstein MJ, Shipley MJ, Rose G. Systolic and diastolic blood pressures as predictors of coronary heart disease mortality in the Whitehall study. Br Med J Clin Res Ed 1985; 291(6490):243-5.

MacMahon S, Peto R, Cutler J *et al.* Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet 1990; 335(8692):765-74.

Miall WE, Lovell HG. Relation between change of blood pressure and age. BMJ 1967; 2(553):660-4.

Murray CJ, Lopez AD. Global and regional cause-of-death patterns in 1990. Bull World Health Organ 1994; 72(3):447-80.

Peto R. The horse-racing effect: Statistical peculiarities of rates of change of medical parameters observed in prospective studies, with special reference to lung disease. Proceedings of the 9th International Biometrics Conference. 1977.

Prospective Studies Collaboration. Cholesterol, diastolic blood pressure, and stroke: 13,000 strokes in 450,000 people in 45 prospective cohorts. Lancet 1995; 346(8991-8992):1647-53.

Qizilbash N, Duffy SW, Rohan TE. Repeat measurement of case-control data: correcting risk estimates for misclassification due to regression dilution of lipids in transient ischemic attacks and minor ischemic strokes. Am J Epidemiol 1991; 133(8):832-8.

Rimm EB, Stampfer MJ, Ascherio A, Giovannucci E, Colditz GA, Willett WC. Vitamin E consumption and the risk of coronary heart disease in men [comment]. N Engl J Med 1993; 328(20):1450-6.

Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol 1990; 132(4):734-45.

Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. Am J Epidemiol 1992; 136(11):1400-13.

Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 1989; 8(9):1051-69; discussion 1071-3.

Shaten B, Kuller L, Neaton J. Association between baseline risk factors, cigarette smoking, and CHD mortality after 10.5 years. Prev Med 1991; 20(5): 655-9

Stamler J, Dyer AR, Shekelle RB, Neaton J, Stamler R. Relationship of baseline major risk factors to coronary and all-cause mortality, and to longevity: findings from long-term follow-up of Chicago cohorts. Cardiology 1993; 82(2-3):191-222.

Stampfer MJ, Sacks FM, Salvini S, Willett WC, Hennekens CH. A prospective study of cholesterol, apolipoproteins, and the risk of myocardial infarction [see comments]. N Engl J Med 1991; 325(6):373-81.

Stegmayr B, Asplund K, Kuulasmaa K, Rajakangas AM, Thorvaldsen P, Tuomilehto J. Stroke incidence and mortality correlated to stroke risk factors in the WHO MONICA Project. An ecological study of 18 populations. Stroke 1997; 28(7):1367-74.

Thun MJ, Peto R, Lopez AD *et al.* Alcohol consumption and mortality among middle-aged and elderly U.S. adults [see comments]. N Engl J Med 1997; 337(24):1705-14.

WHO MONICA Project Principal Investigators. Geographical variation in the major risk factors of coronary heart disease in men and women aged 35-64 years. The WHO MONICA Project. World Health Stat Q 1988; 41(3-4):115-40.

APPENDICES

APPENDIX 1

The Prospective Studies Collaboration

The investigators of the studies comprising the Prospective Studies Collaboration are:

Akita: H Iso, H Sato; **ARIC**: L Chambless; **BIRNH**: G De Backer, D De Bacquer, M Kornitzer; **BRHS**: S Ebrahim, P Whincup, G Wannamethee; **BUPA**: N Wald, J Morris; **Busselton**: M Knuimann; **Caerphilly**: P Sweetnam, P Elwood; **Cardiovascular Health Study**: R Kronmal; **CB project**: D Kromhout; **Charleston**: S Sutherland, J Keil; **Copenhagen**: P Schnohr, G Jensen; **ERGO**: D Grobbee, J Witteman; **Evans County**: C Hames; **Finnish Mobile Clinic Survey**: A Aromaa, P Knekt, A Reunanen; **Finrisk**: J Tuomilehto, P Jousilahti, E Vartiainen; **Framingham**: D Levy, R D'Agostino, H Silbershatz; **Glostrup**: T Thomsen; **Gothenburg Women**: C Bengtsson; **Honolulu**: D Sharp; **Israeli IHD**: U Goldbourt, S Yaari; **Japan Railway (EJR)**: T Murayama, M Tomita, M Nishimoto; **Leuven**: J Staessen; **LRC**: M Criqui, C Davies; **MHHP**: D Jacobs, H Blackburn, R Luepker; **MHS**: D Jacobs, H Blackburn, R. Luepker; **MRFIT**: J Neaton; **NHEFS**: C Cox, M Ofstedal; **Normative Ageing Study**: S Weiss, P Cassano, D Sparrow, P Vokonas; **Norwegian**: A Tverdal, R Selmer; **NPHS**: T Meade, K Garrow, J Cooper; **OG Rome**: A Menotti, A Spagnolo; **Ohasama**: I Tsuji, Y Imai, T Ohkubo, S Hisamichi; **Oslo**: L Haheim, I Holme, I Hjermann, P Leren; **Paris**: P Ducimetiere, J Richard; **Perth**: K Jamrozik, R Broadhurst; **IPC:PARIS**: A Benetos, L Guize; **PROCAM**: G Assmann, H Schulte; **Puerto-Rico (HHP)**: P Sorlie, M Garcia-Palmeri; **Rancho Bernado**: R Langer, E Barrett-Conner; **Renfrew/Paisley**: C Gillis, D Hole; **Saitama**: K Nakachi; **Seven-Countries Croatia**: R Buzina; **Seven-Countries Finland**: P Kivinen, A Nissinen; **Seven-Countries Greece**: C Aravanis, A Dontas; **Seven-Countries Italy**: A Menotti; **Seven-Countries Japan**: H Toshima, H Adachi, T Imaizumi; **Seven-Countries Netherlands**: D Kromhout; **Seven-Countries Serbia**: S Nedeljkovic, N Vojvodic, M Ostojic; **Seven-Countries US Railraod**: D Jacobs, H Blackburn; **Seven cities: China**: X Fang, S Li; **Shanghai**: Z Chen; **SHHS**: H Tunstall-Pedoe; **Shibata**: T Nakayama; **Speedwell**: P Sweetnam, P Elwood; **Tecumseh**: J Keller; **Tromso**: K Bonaa, E Arnesen; **UKHDPP**: H Tunstall-Pedoe; **US Health Professionals**: E Rimm; **US Nurses**: F Speizer, M Stampfer; **US Physicians**: C Hennekens, J Buring; **Värmland**: S Törnberg, J Carstensen; **West of Scotland**: C Gillis; **Whitehall**: M Shipley, D Leon, M Marmot.

One relevant reference for each partcipating study, as indicated by the principle investigator.

Study name (acronym)	Authors	Title	Reference
Akita	Shimamoto T, Komachi Y, Inada H, Doi M, Iso H,	Trends for coronary heart disease and stroke and their risk factors in Japan.	Circulation 1989; 79: 503-15
Atherosclerosis Risk In Communities (ARIC)	The ARIC investigators	The Atherosclerosis Risk In Communities (ARIC) study: Design and objectives.	Am J of Epidemiol 1989; 129: 687-702
BIRNH	De Backer G on behalf of the BIRNH study	Regional differences in dietary habits, coronary risk factors and mortality rates in Belgium.	Acta Cardiologica 1984; 39: 285-92
British Regional Heart Study (BRHS)	Wannamethee G, Shaper AG, Whincup PH and Walker M.	Low serum total cholesterol concentrations and mortality in middle aged British men.	BMJ 1995; 311: 409-13
BUPA	Law MR, Wald NJ, Wu T, Hackshaw A and Bailey A.	Systematic underestimation of association between serum cholesterol concentration and ischaemic heart disease in observational studies: data from the BUPA study.	BMJ 1994; 308: 363-6
Busselton	Knuiman MW and Vu HT.	Risk factors for stroke mortality in men and women: The Busselton study.	J Cardiovasc Risk 1996; 3: 447-52
Caerphilly & Speedwell	Bainton D, Miller NE, Bolton CH, Yarnell JWG, Sweetnam PM, Baker IA et al.	Plasma triglyceride and high density lipoprotein cholesterol as predictors of ischaemic heart disease in British men. The Caerphilly and Speedwell Collaborative Heart Disease Studies.	Br Heart J 1992; 68: 60-6
Cardiovascular Health Study (CHS)	Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA et al.	The Cardiovascular Health Study: design and rationale.	Ann Epidemiol 1991; 1(3): 263-76
CB project	Verschuren WMM, Kromhout D	Total cholesterol and mortality at a relatively young age: Do men and women differ? The Netherlands Consultation Bureau Project on Cardiovascular Diseases.	BMJ 1995; 311: 779-83
Charleston	Keil JE, Sutherland SE, Knapp RG, Lackland DT, Gazes PC and Tyroler HA.	Mortality rate and risk factors for coronary disease in blacks as compared with white men and women.	N Engl J Med 1993; 329: 73-8
Copenhagen City Heart Study	Hansen AT, Schnohr P, Jensen G, Nyboe J, Appleyard M, Danielsen B, et al.	The Copenhagen City Heart Study. Osterbroundersogelsen. A book of tables with data from the first examination (1976-78) and a five-year follow-up (1981-83).	Scand J Soc Med 1989; 170 (Suppl 41): 11-160
Evans County	Heyman A, Karp HR, Heyden S, Bartel A, Cassel JC, Tyroler HA et al.	Heyman A, Karp HR, Heyden S, Bartel A, Cassel JC, Tyroler HA et al. Cerebrovascular disease in the bi-racial population of Evans County, Georgia.	Stroke 1971; 2:509-18
Finnish Mobile Clinic Survey (FMCS)	Reunanen A, Aromaa A, Pyorala K, Punsar S, Maatela J and Knekt P	The Social Insurance Institution's coronary heart disease study. Baseline data and 5-year mortality experience.	Acta Med Scand 1983; Suppl 673: 1-120
Finrisk	Jousilahti P, Vartiainen E, Pekkanen J, Tuomilehto J, Sundvall J and Puska P.	Serum cholesterol distribution and coronary heart disease risk: observation and predictions among middle-aged population in eastern Finland.	Circulation 1998; 97:1087-94
Framingham	Dawber TR, Meadors GF, Moore FE	Epidemiological approaches to heart disease: the Framingham Study	Am J Public Health 1951; 41: 279-86
Glostrup	Schroll M, Jorgensen T and Ingerslev J.	The Glostrup Population Studies 1964-1992.	Dan Med Bull 1992; 39: 204-7
Gothenburg Women	Bengtsson C, Ahlquist M, Andersson K, Bjorkelund C, Lissner L, Soderstrom M	The prospective population study of women in Gothenburg 1968-69 to 1992-93.A 24-years follow-up study with special reference to participation, representativeness, and mortality.	Scand J Prim Health Care 1997; 15: 214-9
Honolulu	Iribarren C, Sharp D, Burchfiel CM, Sun P and Dwyer JH.	Association of serum total cholesterol with coronary disease and all-cause mortality: multivariate correction for bias due to measurement error.	Am J Epidemiol 1996; 143: 463-71
IPC - Paris	Benetos A, Safar M, Rudnichi A, Smuylan H, Richard J-L, Ducimetiere P et al.	Pulse Pressure: A predictor of long-term cardiovascular mortality in a French male population.	Hypertension 1997; 30: 1410-15

Israel	Goldbourt U, Yaari S and Medalie JH.	Factors predictive of long-term coronary heart disease mortality among 10,059 male Israeli civil servants and municipal employees. A 23-year mortality follow-up in the Israeli Ischaemic Heart Disease Study.	Cardiology (Basel) 1993; 82: 100-21
Japanese Rail (EJR)	Tomita M, Yamaguchi M, Sakuma K, Gitsukawa H, Saito N, Nabutomo K et al.	Risk factors of ischemic heart disease in middle-aged Japanese male. (Japanese)	J Transportation Med 1987; 41: 30-33
Leuven	Staessen JA, Roels H, and Fagard R.	Lead exposure and conventional and ambulatory blood pressure: a prospective population study. PheeCad Investigators.	JAMA 1996; 275: 1563-70
Lipid Research Clinics (LRC)	Jacobs DR Jr, Mebane IL, Bangdiwala SI, Criqui MH, Tyroler HA, for the lipid research clinics program.	High-density lipoprotein cholesterol as a predictor of cardiovascular disease mortality in men and women: the follow-up study of the Lipid Research Clinics Prevalence Study.	Am J Epidemol 1990; 131: 32-7
Minnesota Heart Health Program (MHHP)	Luepker RV, Murray DM, Jacobs DR Jr, Mittelmark MB, Bracht N, Carlaw R, et al.	Community education for cardiovascular disease prevention: risk factor changes in the Minnesota Heart Health Program.	Am J Public Health 1994; 84: 1383-93
Minnesota Heart Survey (MHS)	Iso H, Jacobs DR Jr and Goldman L.	Accuracy of death certificate diagnosis of intracranial hemorrhage and nonhemorrhagic stroke. The Minnesota Heart Survey.	Am J Epidemiol 1990; 132: 993-8
Multiple Risk Factor Intervention Trial (MRFIT) screenees	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 Eng J Med 1989; 320: 904-10
Munster Heart Study (PROCAM)	Cullen P, Schulte H and Assmann G.	The Munster Heart Study (PROCAM): total mortality in middle-aged men is increased at low total and LDL-cholesterol concentrations in smokers but not in nonsmokers.	Circulation 1997; 96: 2128-36
NHANES I Epidemiologic Followup Study (NHEFS)	Cox CS, Mussolino ME, Rothwell ST, Lane MA, Golden CD, Madans JH et al.	Plan and operation of the NHANES I Epidemiologic Followup Study 1992. National Center for Health Statistics.	Vital Health Stat 1. 1997; 35: 1-231
Normative Ageing Study (NAS)	Cassano PA, Segal MR, Vokonas PS and Weiss ST.	Body fat distribution, blood pressure, and hypertension. A prospective cohort study of men in the normative aging study.	Ann Epidemiol 1990; 1: 33-48
Northwick Park Heart Study (NPHS)	Meade TW, Ruddock V, Stirling Y, Chakrabarti R and Miller GJ.	Fibrinolytic activity, clotting factors, and long-term incidence of ischaemic heart disease in the Northwick Park Heart Study.	Lancet 1993; 342: 1076-9
Norwegian	Bjartveit K, Foss OP Gervig T and Lund-Larsen PG.	The cardiovascular disease study in Norwegian counties. Background and organization.	Acta Med Scand Suppl 1979; 634: 1-70
Ohasama	Tsuji I, Imai Y, Nagai K, Ohkubo T, Watanabe N, Minami N et al.	Proposal of reference values for home blood pressure measurement: prognostic criteria based on a prospective observation of the general population in Ohasama, Japan.	Am J Hypertens 1997; 10: 409-18
Oslo	Haheim LL, Holme I, Hjermann I and Leren P.	The predictability of risk factors with respect to incidence and mortality of myocardial infarction and total mortality. A 12-year follow-up of the Oslo Study, Norway.	J Intern Med 1993; 234:17-24
Paris	Ducimetiere P, Richard JL, Cambien F, Rakotovao R and Claude JR.	Coronary heart disease in middle-aged Frenchmen. Comparisons between Paris Prospective Study, Seven Countries Study and Pooling Project.	Lancet 1980; 1: 1346-50
Perth	Australian Institute of Health and Welfare and National Heart Foundation of Australia.	Risk factor prevalence study.	Survey no. 3, 1989; 1991: Canberra NHF
Puerto-Rico Heart Health Program (PRHHP)	Garcia-Palmieri MR, Sorlie PD, Havlik RJ, Costas R Jr and Cruz-Vidal M.	Urban-rural differences in 12-year coronary heart disease mortality: the Puerto-Rico Heart Health Program.	J Clin Epidemiol 1988; 41: 285-92
Rancho-Bernado	Criqui MH, Barrett-Connor E, and Austin M.	Differences between respondents and non-respondents in a population-based cardiovascular disease study.	Am J Epidemiol 1978; 108: 367-72.
Renfrew/Paisley	Isles CG, Hole DJ, Gillis CR, Hawthorne VM and Lever AF	Plasma cholesterol coronary heart disease and cancer in Renfrew and Paisley survey.	BMJ 1989; 298: 920-24

Rotterdam (ERGO)	53 Bots ML, Hoes AW, Koudstaal PJ, Hofman A and Grobbee DE.	Common carotid intima-media thickness and risk of stroke and myocardial infarction: the Rotterdam Study.	Circulation 1997; 96: 1432-7
Saitama	Imai K and Nakachi K.	Cross sectional study of effects of drinking green tea on cardiovascular and liver diseases.	BMJ 1995; 10: 693-96.
Scottish Heart Health Study (SHHS)	Bolton-Smith C, Woodward M, Brown CA and Tunstall Pedoe H.	Nutrient intake by duration of ex-smoking in the Scottish Heart Health Study.	Br J Nutr 1993; 69: 315-32
Seven Cities: China	Cheng XM, Ziegler DK, Lai YH, Li SC, Jiang GX, Du XL et al.	Stroke in China, 1986 through 1990.	Stroke 1995; 26: 1990-4
Seven Countries	Ancel Keys (ed) et al.	Seven Countries: A multi-variate analysis of death and coronary heart disease.	Harvard University Press 1980
Seven countries: Greece	Aravanis C, Corcondilas A, Dontas A, Keys A.	The Greek islands of Crete and Corfu.	Am Heart Association Monograph No 29, 1970: 1-88.
Seven Countries: Italy	Menotti A, Seccareccia F, Lanti M, Giampaoli S and Dima F.	Time changes in predictability of coronary heart disease in an Italian aging population.	Cardiology 1993; 82: 172-80
Seven countries: Japan	Tsuruta M, Hashimoto R, Adachi H, Imaizumi T and Numora G.	Hyperinsulinemia as a predictor of hypertension: an 11-year follow-up study in Japan.	J Hypertension 1996; 14: 483-88.
Seven countries: Netherlands	Kromhout D, Bosschieter EB, Drijver M, de-Lezenne-Coulander C.	Serum cholesterol and 25-year incidence of and mortality from myocardial infarction and cancer. The Zutphen Study.	Arch Intern Med 1988; 148(5): 1051-5
Shanghai	Chen Z, Peto R, Collins R, MacMahon S, Lu J and Li W.	Serum cholesterol concentration and coronary heart disease in a population with low cholesterol concentration.	BMJ 1991; 303: 276-82
Shibata	Nakayama T, Date C, Tokoyama T, et al.	A 15.5 year follow-up study of stroke in a Japanese provincial city: the Shibata study.	Stroke 1997; 28: 45-52
Tecumseh	Higgins M and Keller JB.	Cholesterol, coronary heart disease and total mortality in middle-aged and elderly men and women in Tecumesh.	Ann Epidemiol 1992; 2: 69-76
Tromso	Bonaa KH and Thelle DS.	Association between blood pressure and serum lipids in a population. The Tromso Study.	Circulation 1991; 83: 1305-14
United Kingdom Heart Disease Prevention Project (UKHDPP)	Heller RF, Chinn S, Tunstall-Pedoe HD and Rose G.	How well can we predict coronary heart disease? Findings in the United Kingdom Heart Disease Prevention Project.	BMJ 1984; 288: 1409-11
US Health Professionals Follow-Up Study (USHPF)	Rimm EB, Stampfer MJ, Ascherio A, Giovannucci E, Colditz GA and Willett WC.	Vitamin E consumption and the risk of coronary heart disease in men.	N Engl J Med 1993; 328: 1450-5
US Nurses	Goldhaber SZ, Grodstein F, Stampfer MJ, Manson JE, Colditz GA, Speizer FE et al.	A prospective study of risk factors for pulmonary embolism in women.	JAMA 1997; 277: 642-5
US Physicians Health Study (PHS)	Stampfer MJ, Sacks FM, Salvini S, Willett WC and Hennekens CH.	A prospective study of cholesterol, apolipoproteins, and the risk of myocardial infarction.	N Engl J Med 1991; 325/6: 373-81
US Railroad	Slattery ML, Jacobs DR, Nichamen MZ.	Leisure time physical activity and coronary heart disease death: US railroad study.	Circulation 1989; 79: 304-11
Varmland	Tornberg SA, Holm LE, Carstensen JM and Eklund GA.	Risks of cancer of the colon and rectum in relation to serum cholesterol and beta-lipoprotein	N Engl J Med 1986; 315: 1629-33
West of Scotland	Hawthorne VM and Fry JS.	Smoking and health: the association between smoking behaviour, total mortality and cardiorespiratory disease in West Central Scotland.	J Epidemiol Community Health 1978; 32: 260-6
Whitehall	Smith GD, Shipley MJ, Marmot MG and Rose G.	Plasma cholesterol concentration and mortality. The Whitehall Study.	JAMA 1992; 267:70-6

* Used throughout the thesis