

**Blood pressure variability over time:
statistical implications for risk assessment
and screening policies**

MEDICAL LIBRARY,
ROYAL FREE HOSPITAL
HAMPSTEAD.

THESIS
presented for the
DEGREE
of
DOCTOR OF PHILOSOPHY
in the Faculty of Medicine
(Field of Study: Statistics)

by

Michael David Hughes

Department of Epidemiology and General Practice
Royal Free Hospital School of Medicine
University of London

ProQuest Number: U554302

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

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



ProQuest U554302

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Screening programmes based on quantitative factors are designed to identify individuals at high risk of disease. When these factors vary within individuals over time, observations on repeat occasions should aid risk classification. However, there have been few statistical developments for assessing accumulated evidence during screening to determine whether intervention is indicated. This study addresses this issue in the context of screening diastolic blood pressure for cardiovascular risk.

Data on a cohort of 11299 middle-aged men is used to develop models describing variability during a period of four years. From annual diastolic pressure measurements, a model based on normal variation about an underlying subject mean level, with standard deviation dependent on level, fitted well. No evidence for a risk relationship with trends or variability about the mean level was found: increased risk appears to be established only through a raised underlying mean level. As this cannot be measured directly, a survival model based on observed level is fitted and adjustment for the effect of “regression dilution” made to determine the magnitude of this association.

Two alternative statistical models for screening strategies are proposed. The first emphasizes precision in determining an individual’s underlying level and hence also their risk. Substantial numbers of measurements over several months may be required particularly when the level is borderline for intervention. The second approach takes a public health perspective and aims to identify that proportion of the population (of a given size) in which expected risk is maximized subject also to a screening cost constraint (the average number of visits per person). Using this rule, most gains in identified risk can be achieved by averaging only a little more than one visit per person. These rules should enhance the value of screening providing an informed assessment of risk before commencing intervention.

Acknowledgements

I am grateful for permission to use data for this study from the World Health Organization sponsored trial on the primary prevention of ischaemic heart disease with clofibrate, and, in particular, Austin Heady for his support with understanding the data. I appreciate the guidance of Stuart Pocock and also Simon Thompson for stimulating discussion of the issues raised by this work. Finally, special thanks to Helen Teece for encouragement through to completion and her efforts with the preparation of the manuscript.

Contents

	<u>page</u>
1. Introduction	13
2. Study cohort	
2.1 Introduction	18
2.2 The clofibrate trial	18
2.3 Derivation of study cohort from the trial population	22
2.4 Limitations of the cohort	27
3. Describing blood pressure variability	
3.1 Introduction	30
3.2 Previous studies of blood pressure variability	
3.2.1 Between individuals	31
3.2.2 Within individuals	33
3.3 Method of blood pressure determination in this study	39
3.4 Digit preference	39
3.5 Blood pressure levels by treatment group	48
3.6 Within-person variability	50
3.7 Within-person trends	57
3.8 Blood pressure variability dependence on age, body mass index and smoking	64
3.8.1 Age	65
3.8.2 Smoking status	67
3.8.3 Body mass index	69
3.9 Modelling blood pressure variability	71
3.10 Summary of main results with discussion	83
4. Blood pressure and risk of cardiovascular mortality	
4.1 Introduction	88
4.2 Observed blood pressure level and risk	
4.2.1 Model definition and assessment of assumptions	90
4.2.2 Estimates of risk association: diastolic blood pressure	98
4.2.3 Estimation of risk association: other factors	107

	<u>page</u>
4.3 Underlying blood pressure level and risk: correcting for regression dilution	
4.3.1 Background to regression dilution	111
4.3.2 Regression dilution in the proportional hazards model: background	112
4.3.3 Regression dilution in the proportional hazards model with no censorship	113
4.3.4 Regression dilution in the proportional hazards model with censorship present	122
4.3.5 Blood pressure and risk: adjusting for regression dilution	127
4.4 Summary of main results with discussion	132
5. Screening blood pressure	
5.1 Introduction	138
5.2 Review of statistical developments for screening blood pressure	141
5.3 Screening the individual for treatment	
5.3.1 The general model	144
5.3.2 The normal model for screening the individual for treatment	146
5.4 Screening the individual for expected risk	162
5.5 Screening a population	172
5.5.1 Constrained optimization model for population screening	173
5.5.2 Maximizing expected risk: population screening of blood pressures in Edinburgh	178
5.6 Summary of main results with discussion	190
6. Further perspectives	198
6.1 Risk assessment	198
6.2 When is treatment effective?	202
6.3 Screening rules	204
6.4 Monitoring the effects of intervention	206
6.5 Conclusion	208
References	209

Tables

	<u>page</u>
2.1 Time intervals between visits in the study cohort	26
2.2 Numbers in study cohort by town and treatment group	27
2.3 Characteristics of subjects included in the study cohort	28
3.1 Summary of extremes in p_m when the standard deviation of the assumed blood pressure distribution is 5 or 6	44
3.2 Extreme cases of rounding to the nearest ten when the measured variance is not equal to $(\sigma^2 + \tau^2)/3$	47
3.3 Mean (s.d.) of diastolic blood pressures at each visit by treatment group	49
3.4 Means, standard deviations and correlation structure of diastolic blood pressure (mmHg) by town	51
3.5 Mean (s.d.) of linear regression parameters fitted to each person's sequence of diastolic blood pressures from visits V_{-3} through to V_0	53
3.6 Mean (s.d.) of linear regression parameters fitted to each person's sequence of diastolic blood pressure from visits V_{-3} through to V_0 , by age group	66
3.7 Mean (s.d.) of linear regression parameters fitted to each person's sequence of diastolic blood pressure from visits V_{-3} through to V_0 , by smoking status	68
3.8 Mean (s.d.) of linear regression parameters fitted to each person's sequence of diastolic blood pressure from visits V_{-3} through to V_0 , by body mass index	70
3.9 Parameter estimates (approximate standard errors) for the diastolic blood pressure model	74
3.10 Parameter estimates for the diastolic blood pressure model with μ and σ estimated within age-group	80
3.11 Parameter estimates for the diastolic blood pressure model with all parameters estimated within age-group	81
3.12 Parameter estimates for the diastolic blood pressure model with trends	83

	<u>page</u>
4.1 Cardiovascular mortality and length of follow-up by town and across all towns	89
4.2 Risk coefficients for diastolic pressure measured at a single visit when included separately in a model also with age, cholesterol, smoking status, town and treatment group	98
4.3 Risk coefficients for observed diastolic pressures when included simulataneously in a model also with age, cholesterol, smoking status, town and treatment group	100
4.4 Risk coefficients for functions of the four observed diastolic pressures when included separately in a model also with age, cholesterol, smoking status, town and treatment group	104
4.5 Risk coefficients for the mean of four observed diastolic blood pressures, by town, when included simultaneously in a model also with age, cholesterol, smoking status, town and treatment group	105
4.6 Risk coefficients and the associated estimates of relative risk for all variables included in the model with mean diastolic blood pressure over the four visits	108
4.7 Derivation of coefficients for associating underlying diastolic blood pressure with risk of dying from cardiovascular causes, by town	130
4.8 Estimated risk gradients associated with the means of different numbers of diastolic blood pressure measurements	134
5.1 Screening the individual for treatment in Edinburgh: critical values for L_i and U_i for visit i by age-group when $\xi_T=95\text{mmHg}$, $p_U=0.9$ and $p_L=0.1$	150
5.2 Screening the individual for treatment: critical values for L_i and U_i for visit i for all ages by town when $\xi_T=95\text{mmHg}$	153
5.3 Screening the individual for treatment when within-person variability is assumed not to depend on ξ : critical values for L_i and U_i for visit i for all ages in Edinburgh when $\xi_T=95\text{mmHg}$	154

	<u>page</u>
5.4 Simulation of 100000 Edinburgh men being screened for treatment using $\xi_T=95\text{mmHg}$, $p_U=0.9$ and $p_L=0.1$ according to whether their underlying level ξ is greater or less than 95mmHg	160
5.5 Screening the individual for expected risk: critical values for L_i and U_i for visit i for all ages by town	169
5.6 Comparison of screening rules applied to a population of Edinburgh men when the intervention group has a fixed size =5.27% of the population= $p(\xi>95)$	181
5.7 Results of a simulation of 100000 Edinburgh men entering screening programmes with the sequential rules given in table 5.6, when their blood pressures are measured exactly or are rounded to the nearest even digit	183

Figures

	<u>page</u>
3.1 Digit preferences by town	40
3.2 Rounding units and intervals	42
3.3 Relationship between observed residual variance and mean level	55
3.4 Normal probability plots for the observed trends within-person	59
3.5 Relationship between observed trends and mean level	63
3.6 Observed and predicted distributions of blood pressure	78
4.1 Crude cardiovascular mortality rates by quartile of diastolic pressure within town	92
4.2 $-\log[-\log S(t)]$ plots by mean diastolic pressure group	94
4.3 Residual plot for the proportional hazards model including mean diastolic pressure and other covariates	96
4.4 Residual plot for the proportional hazards model, stratified by mean diastolic pressure group	97
4.5 Crude cardiovascular mortality rates by quartile of annual trend in diastolic pressure, within town	102
4.6 Crude cardiovascular mortality rates by quartile of residual standard deviation of diastolic pressure, within town	103
4.7 Regression dilution: the relationship between β/β^* and the variance ratio, λ , for the proportional hazards model when the covariate is measured with a normally distributed error and there is no censorship	119
4.8 Regression dilution: relationship between the correction factor β^*/β and the expected risk coefficient relating to the observed covariate, β , in the proportional hazards model with no censorship present and normally distributed measurement errors	121

	<u>page</u>
4.9 Regression dilution in the proportional hazards model with censorship present: relationship between β/β^* and λ when $\beta=0.5$ for different levels of censorship	125
4.10 Regression dilution in the proportional hazards model with censorship present: relationship between β/β^* and λ when $\beta=1.0$ for different levels of censorship	126
5.1 Approach to screening for elevated diastolic blood pressure suggested by the World Health Organization guidelines	139
5.2 Screening the individual for treatment: critical values by visit giving defined probabilities $p(\xi > \xi_T = 95 \bar{x})$ for all ages in Edinburgh	156
5.3 Screening the individual for treatment: critical values by visit giving defined probabilities $p(\xi > \xi_T = 90 \bar{x})$ for all ages in Edinburgh	159
5.4 Screening the individual for risk: critical values by visit giving defined ratios of individual risk to the average in the population for all ages in Edinburgh	167
5.5 Conditional distribution for underlying diastolic pressure, ξ , for an Edinburgh man after four pressure measurements of 106, 90, 98 and 104mmHg, showing also the associated risk scales if the man has other risk factors as low risk or high risk	171
5.6 Optimal screening rules for an average of M visits per person with 5.27% of the population considered for intervention and a maximum of 4 visits	185
5.7 Conditional probability density function for ξ and the associated risk relative to the average in the population given an observed mean diastolic pressure of 95.1mmHg after 4 visits	188

5.8 Conditional probability density functions for ξ given
observed mean diastolic pressures equal to the critical
values for considering intervention (U) or stopping screening
(L) for the optimal rule with a maximum of 4 visits and with
M=1.25 and P=5.27%

1. Introduction

Primary prevention of chronic diseases is now a major component of health care policies in the developed world. In the United Kingdom, government policy has emphasized this with the introduction of the new contractual arrangements for general practitioners which promotes regular health checks (for a review, see Chisholm, 1990). For men, both in the United Kingdom and more generally in western countries, this will mainly concern the prevention of cardiovascular disease which is the dominant cause of mortality.

Rose (1981) focuses on two strategies for prevention. The first concentrates on the identification of those at high-risk of developing the disease, with the aim of applying intervention that will lower their risk. The second approach is a mass strategy where the aim is to shift the whole risk profile of the population downwards by altering lifestyles through heightening awareness of factors that increase risk. Although Rose suggests that it is the latter strategy which will, in the long-term, produce the greatest reduction in risk in the country as a whole, it is the former high-risk approach which is generally adopted by the medical profession if only because of their fundamental concern with the health of individual patients. This is the focus of screening policies for this study.

Identification of high-risk individuals requires screening of risk factors. In many cases, these factors are continuous variables which vary within individuals and so a single measurement may be a poor reflection of an individual's usual or average level. Thus repeated measurements can be expected to improve the assessment of level and hence the classification of risk and so determine the appropriateness of intervention. This

study focuses on one such risk factor for cardiovascular disease, that of diastolic blood pressure. However, the conceptual framework presented can be applied more generally to other screening situations involving factors which vary within subjects.

Two “expert” groups have formulated guidelines defining screening strategies based upon repeated diastolic blood pressure measurements (Participants of the W.H.O./I.S.H. meeting, 1986; British Hypertension Society working party, 1989). Both recommend levels for intervention based on averages of a series of measurements: typically above 100mmHg for pharmacological intervention and between 95 and 100mmHg for behavioural intervention. However, there is little attention paid, particularly in the latter report, to more formally defined rules that take account of the increased precision obtained in estimating the underlying level as the number of measurements is increased. In addition, both reports suggest no follow-up of patients whose initial blood pressure measurements (one or two) are less than 90mmHg or 95mmHg, respectively. Thus patients with truly raised blood pressures may not be screened further due to their initial measurements being, because of natural variation, below their usual levels. It is the development of rules that address these issues that is the focus of chapter 5 in this study.

The first type of rule considered concerns precision in estimating an individual’s underlying blood pressure or risk level. This then enables statements about the extent of elevated pressure or risk to be assessed either in absolute terms or in terms of probabilities that it is above (or below) some defined threshold. These can then form the basis of treatment decisions or of the desirability of further screening. An alternative approach put forward takes a public health perspective where the emphasis

is more on identifying a high risk group within the population and less on individual risk classification. By taking screening and treatment costs as constraints, a screening rule can be formulated which is optimal in the sense that it identifies that group of a given size in which the greatest morbidity is expected.

In order to develop these approaches to screening, it is first necessary to obtain the required building blocks. In general, these are models for describing variability in the factor of interest within individuals over time and for describing risk relationships with different characteristics of this variability, such as average level, trend and lability. In order to fit these models for diastolic blood pressure, data from a major trial looking at the primary prevention of cardiovascular disease using a cholesterol-lowering drug is used. This trial provides repeated measurements of diastolic pressure over several years and is used to give a study cohort of 11,299 middle-aged men from three European cities. The derivation of this cohort is described in detail in chapter 2 together with discussion of the implications of using such trial data.

Chapter 3 is concerned with the development of models of within-person variability for diastolic blood pressure. As the emphasis of the trial was cholesterol lowering, blood pressures were not subjected to as much experimental rigour as they might. Although this casual determination might parallel realistic general practice, it is useful to consider the magnitude of the component of within-person variance attributable to digit preferences, such as in recording blood pressures to the nearest five or ten mmHg. This enables the importance of such measurement practices to be assessed in the context of "total" within-person variability. It forms the first part of the chapter. The second part is primarily descriptive, illustrating the relationships between average

levels, trends over time and the remaining residual variance both with each other and with other factors such as age, within the three cities. Finally, based on these observed relationships, formal statistical models are fitted describing diastolic blood pressure changes within individuals during a period of three years.

Relating risk of cardiovascular mortality to characteristics of blood pressure variability is considered in chapter 4 using the same data set. This proceeds in two stages. The first is to develop survival models using the observed blood pressures. By considering just a single measurement, the models derived can be compared with those presented in the medical literature from other studies. These models are then developed further using repeated blood pressure measurements so that the prognostic value of trends over time and of lability can be assessed. Following this, a survival model is developed to relate risk to an individual's underlying pressure. This requires consideration of the issue of "regression dilution" whereby the desired risk relationship is estimated from the risk coefficients associated with observed blood pressures with adjustment to reflect their variability about the underlying level. This problem, also known as the "errors in variables" problem, is well-known in the usual linear regression modelling situation where a relatively simple solution arises under certain conditions. The literature on the problem in the context of survival models is more scarce. Thus part of chapter 4 considers methodological aspects of the magnitude of the adjustment required to allow for regression dilution in the context of the proportional hazards model. In particular, its dependence upon the underlying hazard function, the risk coefficient related to underlying level and the degree of censorship are considered. These results are then applied to the diastolic pressure/risk of cardiovascular disease setting of interest in the screening application considered in this study.

As each chapter forms a distinct piece of work of interest in its own right, the relevant literature is reviewed therein. In addition each chapter is concluded with a summary of the main results presented with discussion of the issues raised. Finally in chapter 6, the implications of the work are considered briefly particularly in the context of issues that would be useful for further research.

2. The study cohort

2.1 Introduction

In many clinical trials considerable amounts of follow-up data are accumulated on parameters not directly related to the primary objectives of the trial. Often this is necessary so that the safety of treatments under test can be monitored. This data can be a valuable source of information for other observational studies provided that both the selection of the trial population and any possible interaction of the treatments studied with either the parameters or the outcomes of interest are not overlooked. This study uses a cohort of men derived from the World Health Organization cooperative trial on primary prevention of ischaemic heart disease with clofibrate (henceforth, termed the clofibrate trial) (Heady, 1973; Committee of Principal Investigators, 1978, 1980, 1984). The design of this trial incorporated a five year treatment period during which regular medical examinations were undertaken, and also had a long-term follow-up for mortality beyond this treatment period. Therefore it offers substantial data on blood pressures and allows variability in pressure to be related to survival. In this chapter, the key elements of the trial are described and details of the derivation of this study's cohort are given.

2.2 The clofibrate trial

Clofibrate is a cholesterol lowering drug. The primary objective of the trial was to assess whether the incidence of ischaemic heart disease could be lowered in men with raised serum cholesterol levels using clofibrate. Thus the trial aimed to identify 10000 men whose serum cholesterol levels were in the top third of the population distribution

and to randomise half of these to receive clofibrate and half to receive a placebo. In addition, a further 5000 men with cholesterol levels in the bottom third of the population distribution were to be used as a second control group of men at low risk of ischaemic heart disease.

Between 1964 and 1972 over 52000 male volunteers, aged 30 to 59, from three European cities (Edinburgh, Budapest and Prague) were considered for entry into the trial. Recruitment was from a variety of sources: in Edinburgh the majority were volunteer blood donors although about a third were recruited through advertising, through contact with general practitioners or by encouraging donors to approach friends; in Budapest, about a third were blood donors and the remainder were recruited using population tuberculosis screening registers; in Prague, all recruitment was based on electoral rolls. Clearly then, the trial population is not based upon a random sample of men from the population of the chosen age-group, though the sampling frame used in the latter two towns might be expected to lead to reasonably representative samples. On the other hand, the need to use volunteers in a primary prevention trial might be expected to give more health conscious individuals and possibly subjects of a higher social classification than in a random sample from the population. Thus, particularly in Edinburgh, the trial population might have, on average, different (possibly lower) blood pressures and risk profiles than such a random sample.

From each volunteer, a serum cholesterol level was obtained. In Edinburgh this was from a single measurement; in Budapest and Prague two measurements were taken and the mean of these considered. Using distribution curves of levels, by town, obtained from a preliminary screening, it was determined whether a man fell into the top,

middle or bottom third of the distribution within their respective town. Those in the middle third were not considered further. Following a medical examination to confirm trial exclusion criteria and after obtaining consent, men in the top third were randomly allocated to receive either clofibrate (group I) or a placebo of olive oil (group II); also a random sample of half of the men in the bottom third were allocated to receive the same placebo (group III); the remainder were not considered further. Essentially, the exclusion criteria removed men from being entered into the trial with pre-existing heart disease:

- (i) history of treated myocardial infarction;
- (ii) ECG evidence of heart disease;
- (iii) clinical evidence of rheumatic heart disease;
- (iv) congenital heart disease;
- (v) pulmonary heart disease;
- (vi) other heart disease associated with cardiomegaly or heart failure.

Also, men with diabetes mellitus requiring drug treatment or men with co-existing disease with an unfavourable prognosis that reduced the likelihood of completing the trial were excluded. In addition, there were hypertension criteria for exclusion:

- (i) a diastolic pressure of 120mmHg or greater on any one occasion;
- (ii) diastolic pressure of 110-119mmHg on any two occasions;
- (iii) a diastolic pressure of 110-119mmHg on any one occasion if accompanied by ECG signs of left ventricular hypertrophy or strain;
- (iv) diastolic pressure is within the accepted limits but only on account of treatment with antihypertensive drugs and the ECG shows signs of left ventricular hypertrophy or strain.

These are particularly relevant to this study as it implies that the distribution of blood

pressures and hence of risk of cardiovascular disease in the trial population will be under-representative of higher levels in the population.

In all, 15744 men were entered into the trial, approximately equally distributed both between the three groups and between the three towns. Treatment within the trial lasted for five years. During the first two years, the men were meant to be seen at six-monthly intervals and, for the next three years at annual intervals. At these visits, a medical examination was conducted including the measurement of blood pressure. However, men could be withdrawn from the trial in which case no further examinations were performed. Besides death, reasons for withdrawal were:

- (i) myocardial infarction according to defined symptom state, ECG criteria and serum enzyme levels;
- (ii) heart disease not previously recognised: congenital or pulmonary or other (associated with cardiomegaly or heart failure);
- (iii) diabetes mellitus requiring drug treatment;
- (iv) contra-indications for taking trial medication;
- (v) hypertension as defined in the exclusion criteria above.

Note, however, that the hypertension criterion was not universally applied and, in 1973 (two years before trial treatment was phased out), was withdrawn. Men developing hypertension were then allowed to remain in the trial with appropriate treatment, as necessary. 220 men (1.4% of those entering the trial) were withdrawn due to the hypertension criterion though 470 (3.0%) had, at some time during the trial, one blood pressure measurement of 120mmHg or more, or two measurements between 110 and 119mmHg. These figures are not directly comparable as there is insufficient information available to determine, retrospectively, which men satisfied other aspects of

the hypertension exclusion criteria and when. However they give an approximate idea of the extent to which men in the trial population developed hypertension (as defined) and the extent to which men with these levels of blood pressure might be under-represented in the cohort to be selected. During the treatment period, men were followed up for both morbidity and all-cause mortality. Thereafter only mortality follow-up was completed. This continued until the end of 1982 when an intensive check to establish survival status for all trial participants was carried out; this was successful in over 99% of cases.

In 1975, the trial was stopped and during the following year treatment was withdrawn as the men attended for their next scheduled visit. The average treatment period was 5.3 years with a mean reduction in serum cholesterol levels of 9% observed in the clofibrate-treated group compared with the high cholesterol control group. In comparing groups I and II (clofibrate versus placebo in men with raised serum cholesterol levels at screening), a 25% reduction in non-fatal ischaemic heart disease was observed but with no reduction in fatal myocardial infarctions. However, there was a significant 47% increase in all-cause mortality during the treatment period in those treated with clofibrate (adjusted χ^2_1 -statistic for the crude rates=7.33, $p<0.01$). This effect did not seem to carry over into the post-treatment follow-up period where the excess was 5%.

2.3 Derivation of the study cohort from the trial population

The essential and ideal feature of the cohort to be studied here is that it should have a series of visits at fixed intervals at which blood pressure is measured. Using the data

from the clofibrate study potentially gives six visits per subject spaced at annual intervals. This would cover the full five-year period of treatment. In practice, visits were not always timed as planned. Few visits were completed more than a couple of weeks before the scheduled time, but many were weeks or months late. Thus, to avoid an excessive loss of subjects from the cohort, only five visits, at approximately annual intervals, per subject will be considered. In addition, it is desirable to pick only those subjects that satisfy common criteria for adequate spacing of visits. This is achieved by the following algorithm for each subject:

1. Select a reference visit, V_0 , not previously considered: if no such visit exists, subject not eligible for the cohort.
2. Can a visit, V_{-1} , be found such that:
 - (i) V_{-1} is before V_0and (ii) the interval between V_{-1} and V_0 is between 183 and 547 days (i.e. 6 months and 1.5 years)?

If not: return to step 1.

3. Can a visit, V_{-2} , be found such that:
 - (i) V_{-2} is before V_0and (ii) the interval between V_{-2} and V_0 is between 548 and 913 days (i.e. 1.5 to 2.5 years)?

If not: return to step 1.

4. Can a visit, V_{-3} , be found such that:
 - (i) V_{-3} is before V_0and (ii) the interval between V_{-3} and V_0 is between 914 and 1278 days (i.e. 2.5 to 3.5 years)?

If not: return to step 1.

5. Can a visit, V_{-4} , be found such that:

(i) V_{-4} is before V_0

and (ii) the interval between V_{-4} and V_0 is between 1279 and 1643 days (i.e. 3.5 to 4.5 years)?

If not: return to step 1.

6. Add subject to cohort.

The labelling adopted for the visits reflects their use in later survival analyses where all models developed describe survival from a time-origin at V_0 . Note that if, at any step 2 through to 5, two or more visits could be found within the defined time window, then the one closest to the midpoint of the window was selected. The outcome of this algorithm is a cohort of subjects all with five visits V_{-4} , V_{-3} , V_{-2} , V_{-1} and V_0 , ordered in time, about a year apart and spanning no more than 4.5 years. From this set are further excluded subjects without a recorded diastolic blood pressure at each of the visits. This gives 11299 subjects in total in the study cohort. Of the 4445 subjects in the trial population not included, 2908 lapsed from the trial before accumulating sufficient visits, 461 did not have a satisfactory pattern of visits with complete blood pressure records and the remainder (1076) were withdrawals from the trial due to death, heart disease or hypertension, or side effects.

Alternative cohorts for analysis could be derived from the trial population. In particular, greater statistical precision might be obtained by including subjects with fewer than five visits or with greater variability in the relative timing of visits. However the approach adopted has two main advantages which aid in the understanding of the results obtained. First, the cohort is well-defined being middle-aged men free of major diagnosed cardiovascular disease before and during the period

of blood pressure measurement. As one aim of this study is to consider methods for blood pressure screening in the context of primary prevention of cardiovascular disease, this definition is useful: diagnosis of cardiovascular disease during screening will probably be the dominant determinant of any intervention strategy and so the value of blood pressure screening will be different from that when no pre-existing disease has been identified. Second, from a statistical perspective, analyses using repeated measurements, and hence also their interpretation, are simplified when the data for each subject has the same structure: that is the same number of visits at equally spaced intervals with no missing values. Similarly, the comparison of models developed for risk prediction employing different characteristics of a sequence of blood pressure measurements (e.g. a single observation versus the mean of several) is not complicated by varying cohort sizes.

Table 2.1 shows the timing of visits actually achieved for the cohort. For the majority (93%) of subjects, V_{-4} corresponds to the first visit of the trial. What is clear is the skewness of the distribution of visit times particularly for the interval between V_{-4} and V_0 . For visits V_{-3} , V_{-2} and V_{-1} , the majority of subjects are within about one month of the time point aimed for though there is slightly greater drift for V_{-4} . However, to select a cohort with time intervals closer to those aimed for would restrict considerably the number of patients available for analysis.

Table 2.2 shows the breakdown of the cohort by treatment group and by town. Note, particularly, that the distribution between treatment groups is reasonably equitable both overall and within town suggesting no substantial differential losses from the trial population due to treatment. Table 2.3 describes the basic characteristics of the

Visit	Interval aimed for (range)	interval achieved				
		minimum	lower quartile	median	upper quartile	maximum
V ₋₁	365 (183 to 547)	188	363	370	392	546
V ₋₂	730 (548 to 913)	548	725	736	768	913
V ₋₃	1096 (914 to 1278)	914	1086	1104	1134	1276
V ₋₄	1461 (1279 to 1643)	1279	1461	1481	1554	1643

Table 2.1: Time intervals between visits (days before visit V_0) in the study cohort.

cohort. These are presented for visit V_0 as this corresponds to the origin of time used in developing survival models in chapter 4. The age structure in Edinburgh and Budapest are similar with about three-quarters of the sample in each in their 40's or 50's, though on average a little older in Budapest. In Prague, the structure is more peaked being centred upon the 50-54 age-band with only a handful of men more than five years either side of this. In each town, the majority of men were current or ex-smokers at visit V_0 (ranging from 68% in Prague to 73% in Edinburgh) though the proportion of current cigarette smokers varied more so between towns at that time (from 34% in Edinburgh to 47% in Prague). There was little change in smoking status during the time spanned by the visits. Cholesterol levels in the cohort are clearly not

Town	Treatment group			All groups
	I: clofibrate high cholesterol	II: placebo high cholesterol	III:placebo low cholesterol	
Edinburgh	1138 (33.2%)	1141 (33.3%)	1145 (33.4%)	3424 (100%)
Budapest	1271 (32.9%)	1311 (33.9%)	1282 (33.2%)	3864 (100%)
Prague	1399 (34.9%)	1336 (33.3%)	1276 (31.8%)	4011 (100%)

Table 2.2: Numbers in study cohort by town and treatment group

representative of any general population: the strata determined at randomisation have distinctly different levels at visit V_0 with the cholesterol-lowering effect of clofibrate shown by the difference between groups I and II. This difference is established early in the trial and is found at each of the visits V_{-3} to V_0 .

2.4 Limitations of the cohort

The study cohort derived is not a random sample of a general male middle-aged population. At its simplest, it is a cohort of men all of whom are in good health throughout a period of about four years: free of diagnosed heart disease and other

	Edinburgh		Budapest		Prague	
<u>Age at V₀ (years)</u>						
Mean (s.d.)	47.4	(7.4)	50.1	(7.8)	51.6	(3.3)
no.(%)						
<40	117	(16.8%)	445	(11.5%)	0	(0%)
40-44	702	(20.5%)	539	(13.9%)	10	(0.2%)
45-49	814	(23.8%)	828	(21.4%)	998	(24.9%)
50-54	704	(20.6%)	869	(22.5%)	2315	(57.7%)
55-59	408	(11.9%)	574	(14.9%)	674	(16.8%)
≥60	220	(6.4%)	609	(15.8%)	13	(0.3%)
missing	0		0		1	
<u>Smoking status at V₀: no.(%)</u>						
Never smoked	915	(26.8%)	1145	(29.7%)	1260	(31.5%)
Never cig., ex-cigar/ tobacco	74	(2.2%)	3	(0.1%)	9	(0.2%)
Never cig., current cigar/tobacco	232	(6.8%)	16	(0.4%)	40	(1.0%)
Ex-cig., not current cigar/tobacco	849	(24.9%)	967	(25.0%)	786	(19.7%)
Ex-cig., current cigar/ tobacco	166	(4.9%)	31	(0.8%)	25	(0.6%)
Current cigarettes(per day)						
1-10	342	(10.0%)	338	(8.8%)	490	(12.3%)
11-19	276	(8.1%)	305	(7.9%)	357	(8.9%)
20	327	(9.6%)	400	(10.4%)	562	(14.1%)
21-39	192	(5.6%)	533	(13.8%)	387	(9.7%)
40+	42	(1.2%)	123	(3.2%)	80	(2.0%)
missing	9		3		15	
<u>Body mass index at trial entry (kg/m²)</u>						
mean (s.d.)	25.3	(3.0)	26.9	(3.4)	26.9	(3.2)
missing	34		8		5	
<u>Cholesterol at V₀ (mg/dl)</u>						
mean (s.d.):						
group I	225	(36)	226	(37)	263	(36)
group II	252	(38)	246	(38)	280	(34)
group III	200	(31)	197	(30)	218	(28)
missing	252		25		6	

Table 2.3: Characteristics of subjects included in the study cohort.

major disease. With respect to studying blood pressure, three further restrictions are placed on it as a representative sample. Firstly, the sample is determined on the basis of serum cholesterol levels observed at the beginning of the trial. In particular, half of men with low observed cholesterol levels and all of those with medium levels are excluded. This is a weak restriction as the correlation between blood pressure and cholesterol is known to be small: this will be considered in the next chapter. Secondly, some potential trial participants were excluded due to hypertension. This will have caused a truncation of the distribution of observed blood pressures found in the cohort. In addition, some of the men in the cohort may have lowered blood pressures on account of treatment for hypertension: it is not possible to quantify the extent of this though the timing and setting of the trial should mean that this applies only to a tiny proportion of the men involved. Thirdly, clofibrate is known to have a small anti-hypertensive effect so that men in the cohort in that treatment group are likely to have artificially slightly low blood pressures. This point is returned to in the next chapter.

Thus, these restrictions limit the immediate generalization of any results concerning blood pressure variation between individuals. The effect should be less restrictive when considering within-person variability and does not restrict investigation of its relationship with other covariates measured on the individuals.

3. Describing blood pressure variability

3.1 Introduction

The intrinsic variability of blood pressure has long been recognized. Despite this, there have been few studies that have attempted to quantify this variability. However, without such information, the statement that an individual has a diastolic blood pressure of, say, 95mmHg has little direct interpretation. In particular, this is so if it is to be compared with, for example, previous pressures on the same individual or the pressures of people in a clinical trial found to benefit from anti-hypertensive treatment. Of course, it is important to know what state the patient was in at the time(s) of measurement but even if this is known, then whether this figure summarizes one or more measurements on one or more occasions is important. It is this latter point with which we are concerned: a doctor rightly places less emphasis on a single measurement of 95mmHg than on a sequence obtained over several weeks which have an average of 95mmHg. Thus, whilst this doctor recognizes that there is variability in blood pressure, he cannot be certain within what range his patient's underlying level of blood pressure lies unless this variability has been quantified.

This chapter uses the data of the clofibrate trial to quantify components of variability of blood pressure, in particular those within the individual between visits. Following a review of the results of studies that have considered variability over time (section 3.2), the method of blood pressure determination used in the clofibrate trial is described (section 3.3) and the issue of digit preferences in its reporting is addressed (section 3.4). Sections 3.5 to 3.8 then describe characteristics of the sequences of observed diastolic pressure and relate these to other covariates. Using these results, statistical models

are developed in section 3.9. Finally, in section 3.10, the implications for practice of the models derived are discussed.

3.2 Previous studies of blood pressure variability

3.2.1 Between individuals

The primary focus of blood pressure studies has been the description of the distribution of blood pressures in the population and factors associated with variability between individuals. The findings are well established and so less attention will be paid to them in this chapter. However they are worth reviewing so that some comparison of this study's cohort can be made with other studies. Here we limit discussion to describing diastolic blood pressure in adult men.

Very few studies have considered the exact shape of the diastolic blood pressure distribution in the population. Any that do comment (for example: Armitage, Fox et al; 1966, Shaper et al; 1988), describe it as approximately normal generally with a small degree of skewness toward higher values. One notable study (Boe et al, 1957), carried out in the adult population in Bergen, Norway in association with a compulsory tuberculosis screening exercise, assessed the fit of a normal distribution within five-year age bands. Using normal probability plots, an extremely good fit was found for younger age bands (up to 35-39 years) with increasing departures thereafter with increasing age. For example, fitting a normal distribution to the 50 to 60 year age group gave a predicted 0.5% with blood pressures greater than 120mmHg compared with 2 to 3% actually observed. A log normal distribution (found to fit systolic blood pressure reasonably well) did not give an improved fit, causing skewness in the opposite

direction.

Factors associated with between-person variability have been more thoroughly studied. The dominant findings are the increases in blood pressure associated with age, body mass and alcohol intake. Most studies have reported cross-sectional relationships. For age, they suggest increases of about 1mmHg in diastolic pressure for each five year increase in age during middle age (for example: Kesteloot and van Houte, 1974; Gordon and Shurtleff, 1973; Mann et al, 1988; Shaper et al, 1988). In elderly people the effect of age is less clear with some studies reporting lower blood pressures though this may reflect a “healthy survivor” effect whereby those studied are the survivors with lower blood pressures. This can be confirmed by longitudinal studies. In particular the Framingham Study (Gordon and Shurtleff, 1973) suggests, on the basis of 20 years of a cohort follow-up, that diastolic pressure does rise with age at least until 50-54 though the relationship is less clear at older ages. Some studies have also shown increasing variability of blood pressures in the population with age (Boe et al, 1957; Master et al, 1950; Hypertension Detection and Follow-up Program Cooperative Group, 1978) approximately in proportion to the rise in mean pressure. This finding is not universal: for example it was not found in the Framingham Study (Gordon et al, 1976).

The relationship with weight is also widely reported (Boe et al, 1957; Kesteloot and van Houte, 1974; Master et al, 1950; Shaper et al 1988). No simple summary of the relationship is easy to make because of the different methods used to allow for the confounding of height. However Master et al (1950) show that blood pressure rises by weight within height categories and this is consistent with the findings of authors using indices of build such as the body mass index (weight/square of height) as, for example,

Shaper et al (1988). The only other factor that can be related to blood pressure in this study is cigarette smoking. There is little difference between smokers and non-smokers though the former seem to have slightly lower average levels (Friedman et al, 1982). This is illustrated by the British Regional Heart Study data (Shaper et al, 1988) where the difference found is between 1 and 2mmHg for diastolic blood pressure with no apparent dose-response relationship with amount smoked.

3.2.2 Within individuals

Within-person variability is a complex phenomenon. If blood pressure showed no trends with time or if all individuals in a population had underlying blood pressure levels changing at the same rate, then the correlation of two measurements taken on the same individual would be constant over time and would correspond to the ratio of between-person variance to total variance (between- plus within-person variances) as shown by Gardner and Heady (1973). This correlation has been termed the “tracking correlation” by Rosner (1977).

With varying trends between individuals this is not true. In general this will mean that an individual’s relative position in the population blood pressure distribution will change with time. Thus the greater the time-interval between measurements, the greater the change in relative position will be and consequently the correlation of blood pressures measured on the same individual will decrease with time. In addition, the act of measuring blood pressure may affect the level of blood pressure, itself. This effect is most marked as illustrated by the well-known “pressor” effect where stress associated with measurement raises blood pressure which then falls as the subject becomes

familiar with the procedure. This is most noticeable between measurements in the same session but can also be seen between measurement sessions (Gordon et al, 1976).

Few studies have been designed to look at within-person variability, per se. One of the earliest does show, however, the effects of the method of measurement used, or more particularly the state of the patient. Using a graphical display, Kilpatrick (1948), shows clearly that variability between measurements taken over a week can be substantially reduced if a basal method (whereby an individual is settled into an advanced state of rest prior to measurement) is used rather than a casual method. Although, for all practical purposes, casual blood pressures will be used, it serves to illustrate the component of variability attributable to measurement technique, be it extremes of method as in this example or less dramatic sources such as those caused by observer variation, rounding of measurements and digit preference, knowledge of previous measurements, time of measurement, environment of measurement and so on.

Quantification of variability was illustrated by Glock et al (1956) in a study of 21 subjects measured at the same time on working days for three weeks. Finding the range from minimum to maximum measurement for each individual, they found that the median value, across all individuals, of these ranges was 22mmHg (varying from 12 to 36mmHg). Also shown in this study was that the range of blood pressures between days was greater than that within a measuring session. Despite the dependence of these ranges on the number of measurements taken, these results are indicative of later research.

Three studies during the 1960's considered the tracking correlations of blood pressures

separated by time intervals of several years. The first (Harlan et al, 1962) considered a cohort of 1056 men selected in 1940 for Naval Flight Training. They were considered normotensive, of optimal weight and were of 24 years average age. Correlation of diastolic blood pressures between 1940 and 1952 was 0.13, between 1940 and 1958 was 0.16 and between 1952 and 1958, 0.36. The second study (McKeown et al, 1963) gave correlations based on two population samples of men not on anti-hypertensive treatment. In 833 Birmingham men in their 60's at the first measurement, they found a tracking correlation of 0.50 between measurements 3 years apart. In South Wales (Rhondda Fach and the Vale of Glamorgan) the correlation was 0.63 in 500 men (aged 40+) with measurements four years or so apart. The latter group was broken down by age and an increasing correlation with age was found: 0.57 for men in their 40's at the first visit, 0.61 for men in their 50's and 0.72 for men in their 60's or 70's. However this study's data and further follow-up were later analysed by Rosner et al (1977), including also the females studied, and they found no clear trend during middle and old age. In the third study, 208 male college graduates were studied retrospectively and a tracking correlation of 0.16 found for measurements taken when they were about 20 and 40 years old (Julius et al, 1964).

These studies suggest that the tracking correlation does, indeed, decay with time between measurements. Typically correlations for measurements three or four years apart are of the order of 0.6 reducing to about 0.1 to 0.2 for an interval of 20 years. This summary though may be confounded by age as the longer interval studies are based on young people in whom the rate of change of pressure might be expected to be greatest and probably more variable thus reducing the correlation further: indeed, in children, the correlations are markedly smaller even over intervals of a year or less

(Szklo, 1979).

The most important and complete picture of the decay of the tracking correlation comes from the Framingham Study (Gordon and Shurtleff, 1973). Correlations are available between all pairs of blood pressures taken at ten biennial visits on a population cohort of both sexes aged 30-62, when identified in 1948-1952. Casual blood pressures rounded, in practice, to the nearest even digit give correlations of about 0.62 for measurements two years apart and 0.59, when four years apart. At eighteen years apart this is reduced to 0.38. There is no clear effect on the correlations of the aging cohort or of losses from it suggesting reasonable stability across the age range.

The tracking correlation is a transformation of the ratio of within- to between-person variances and so its major drawback is in its dependence on the between-person variance in the population sampled. Thus it can be more useful to consider the size of the within-person variance, itself. Two studies set out to measure this, one in a highly controlled situation and the other in a survey environment (Armitage and Rose, 1966; Armitage, Fox et al, 1966). The first measured blood pressure on five men and five women (mean ages 35 and 31 years respectively) taking two observations on each of 20 occasions over a six week period. In a relaxed environment, a single trained observer using a random-zero sphygmomanometer measured blood pressures. Mean diastolic pressure (phase IV) was 65.4mmHg: low by population standards. Within-person standard deviations of 6.2mmHg between-visits and 3.3mmHg within-visits were estimated. In the survey 50 working men (mean age 48 years in the larger group from which they were derived) had their blood pressures measured at two visits a year apart. The within-person standard deviations were estimated as 7.2mmHg between-

visits and 5.9mmHg within-visits. The difference between the two studies for the within-visit standard deviation emphasizes the potential impact of a carefully controlled measurement technique. Less can be said about the difference in between-visit standard deviations because of the different time spans between measurements.

Since these two studies, several authors have produced estimates for these two components of variance. Gordon et al (1976) obtained estimates from the Framingham Study for the between-visit and within-visit standard deviations of 7.6 and 4.3mmHg, respectively. Hebel et al (1980) measured blood pressures, at home, in a random sample of black inner-city Baltimore residents and remeasured the pressures in those found to be normotensive at baseline about three years later. They obtained estimates of 7.4 and 2.0mmHg, respectively (after ignoring the first of the three measurements at each visit) based on samples of 100 people. Rosner and Polk (1983) used data from both worksite and community blood pressure programs involving weekly measurements over three/four weeks or daily measurements over one week, respectively. They obtained estimates of 5.3 and 2.8mmHg for the between-visit and within-visit standard deviations, respectively, in white men aged 30-49. They also found higher levels of variability in black men and slightly lower levels at older ages. Finally Wilson and Hebel (1988) obtained estimates of 7.3mmHg for the between-visit standard deviation in a sample of black hypertensives, after correcting for the selection as hypertensive on the basis of high blood pressures. Here visits were between 7 days and one year apart: the serial correlation between measurements on the same individual having been found to be negligible over such periods. They also found greater variability in older men compared with younger men.

Clearly these studies have produced varying estimates. It is useful to note, though, that within-visit variability is small compared with that between- visits. This suggests that repeat measures within a visit are of much less value compared with those obtained from several visits for estimating an individual's underlying blood pressure level. Estimates of between-visit standard deviation are about 7.5mmHg for visits a few months to a couple of years apart but may be smaller for shorter intervals. Further comparison is complicated by the differing age, race and sex mixes reported.

Despite these studies to assess variability, none present anything more than summary measures. In particular no attempt has been made to describe the observed distribution of within-person variability and hence whether there is heterogeneity in the underlying variance component or whether differences in observed values between individuals can be explained just by sampling variation. In this context Kannel, Sorlie and Gordon (1980) suggest, for systolic pressures, that there is not an identifiable group in the population with unusually labile pressures when measured at hourly clinic visits. In contrast, Hawthorn et al (1974) suggest that a normal model for differences between visits might be appropriate but pursue it no further. An equally important consideration is the possibility of increasing variability with level. This has been suggested by both Miall and Lovell (1967) and Kannel, Dawber and McGee (1980). If true, it would help explain the greater within-person variability with age, found in some studies, given the increase in level with age. These particular points are addressed in this chapter.

3.3 Method of blood pressure determination in this study

Blood pressure was measured in the manner described in the World Health Organization's recommendations (World Health Organization, 1959 and 1962). These are summarized as follows. Either a mercury manometer or an aneroid instrument can be used with a blood-pressure cuff of width at least 14cm and long enough to completely wrap around the arm. A casual blood pressure is measured by a physician with the subject in the sitting position. The right arm is used. Phase IV diastolic blood pressure was used in this study. To obtain this, the cuff is rapidly inflated to a point 20-30mmHg above the pressure at which the radial pulse is obliterated. Allowing the cuff pressure to fall at a rate of not more than 2-3 mmHg per pulse-beat, the diastolic pressure is the point at which the pulse-beat sound becomes muffled. In general, a single measurement was used. However, in Prague, at the first trial visit (V₄ for most subjects in this study's cohort) the mean of two measurements was recorded.

3.4 Digit preference

The advantage, for our purposes, of the method of measurement used in the trial is its closeness to the method that might be used in casual screening of any individual by a general practitioner. Thus within-person variability studied using this data will also include a component of variance associated with the degree of accuracy employed in taking the measurements. This includes the well-known problem of digit preference, reported many years ago by Janeway (1913), whereby some measurements are rounded to a particular digit, for example to the nearest ten or five. To illustrate this practice in the data being studied, the last digit of the recorded blood pressure for each of the

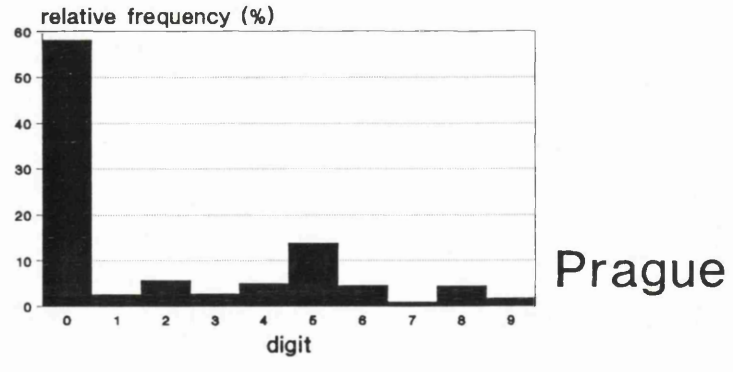
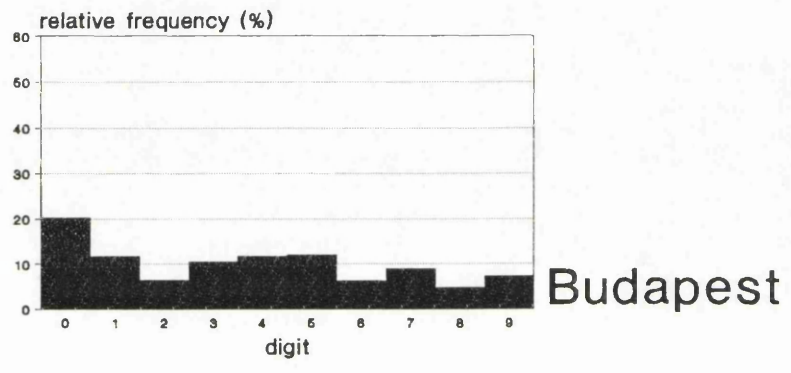
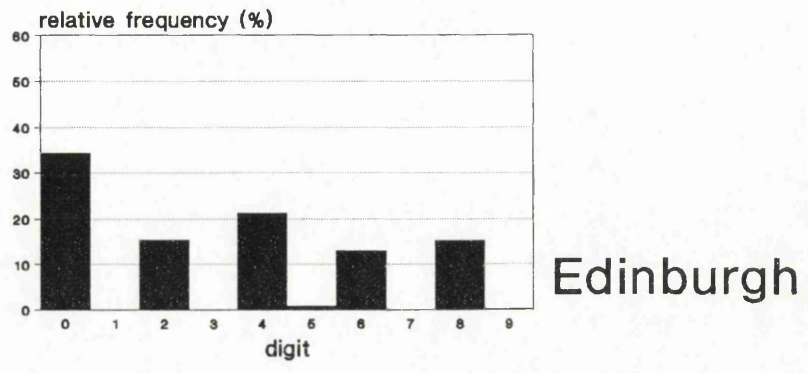


Figure 3.1: Digit preferences by town

five visits for each subject was found and the relative frequency distribution for digit determined. This is shown in figure 3.1 by town.

As the distribution of blood pressures in a population is not uniform, the distribution of last digit should not be expected to be perfectly uniform either. However the preference for certain digits is notable as is the manner in which this preference varies between town. In Edinburgh, a policy of rounding to an even digit is clear with less than 1% of all measurements having an odd final digit. On top of this, there is some additional preference for zero at the expense of most other even values. In Budapest, there is a similar preference for zero though all digits were used. In distinct contrast, in Prague there is a dramatic preference for zero at the expense of all other digits though five finds some favour. It is worth noting that the patterns of digit preference described did not vary systematically with calendar year of measurement though there was some small variation in the emphasis on the various digits.

To assess the impact that digit preference might have on the distribution of observed blood pressures, consider a simple situation where this distribution, if it could be observed exactly, is normal with mean μ and variance σ^2 . Four different rounding practices will be considered, where rounding is to the nearest whole multiple of

- (i) one
- (ii) two
- (iii) five
- (iv) ten.

The normal distribution can be sliced up into mutually exclusive intervals such that each interval is centred on a rounding unit. For example, when rounding is to the

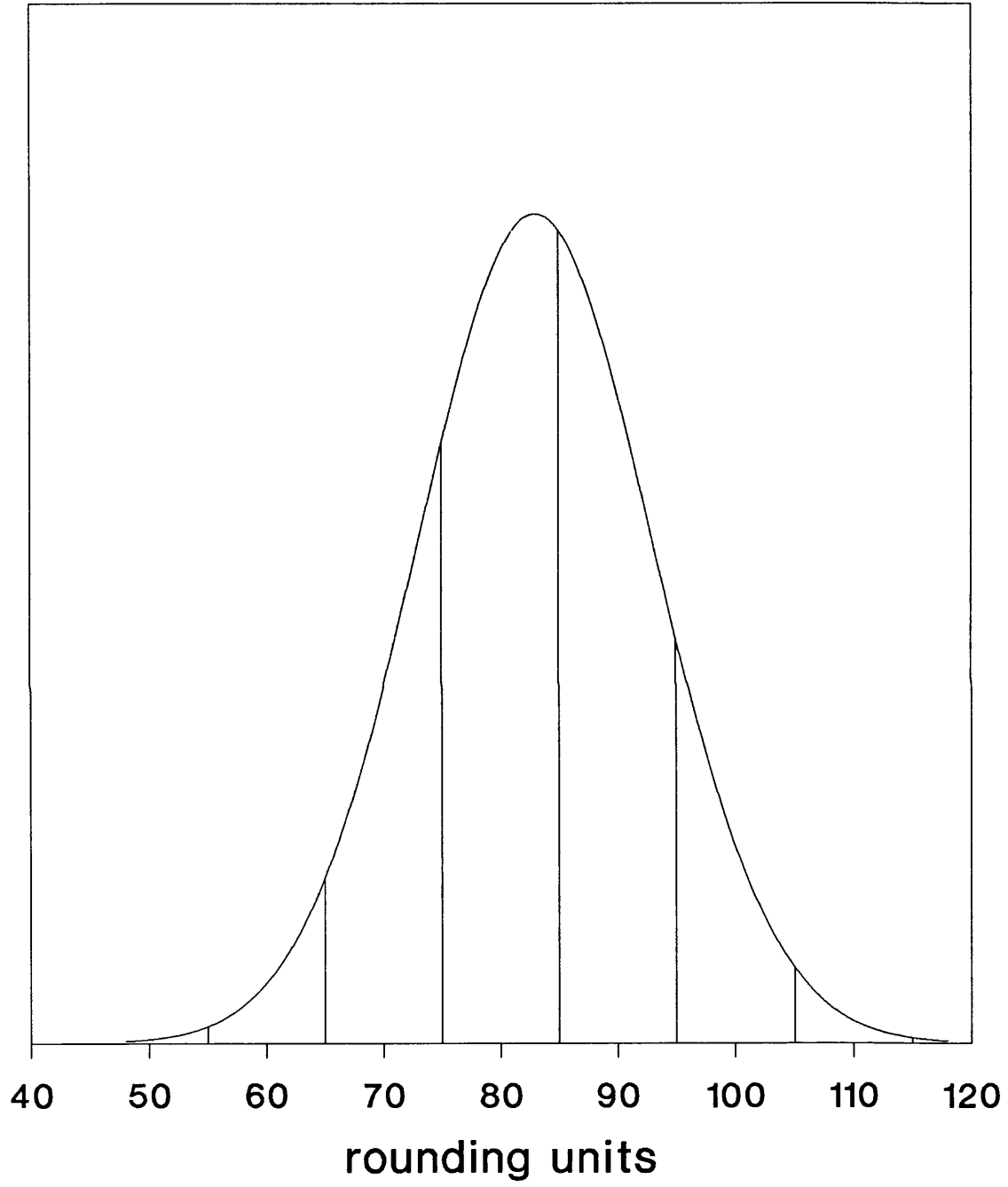


Figure 3.2: Rounding units and intervals

nearest ten, and the normal distribution has mean 83 and standard deviation 10, figure 3.2 shows the sliced normal distribution with each interval centred on a multiple of ten. Any interval can be described in terms of its associated rounding unit $(10n+m)$, where n is any integer and m is any allowable digit for the relevant practice:

- (i) m is any digit 0 through to 9
- (ii) m is any even digit 0 through to 8
- (iii) m is 0 or 5
- (iv) m is 0.

Taking the width of the intervals as 2τ , then any blood pressure in the interval $(10n+m-\tau, 10n+m+\tau)$ will be rounded to $(10n+m)$ and hence the probability of recording a blood pressure with last digit m will be the sum over n of probabilities across all such intervals:

$$\begin{aligned}
 P_m &= \sum_{n=-\infty}^{n=+\infty} \left[\Phi \left[\frac{10n+m+\tau-\mu}{\sigma} \right] - \Phi \left[\frac{10n+m-\tau-\mu}{\sigma} \right] \right] \\
 &= \sum_{n=-\infty}^{n=+\infty} P_{nm}, \text{ say,}
 \end{aligned}$$

where Φ is the standard normal cumulative distribution function.

For each rounding practice considered, these probabilities can be calculated for the set of allowable final digits. The interesting question is then the extent to which the normal distributional form assumed for blood pressures produces a departure from the distribution of final digits in which all are equally favoured (that is, the p_m 's are all equal). The greatest departures are found when the mean, μ , is equal either to a

2τ	σ	μ^*	0	1	2	3	4	digit, m	5	6	7	8	9
1	5	(i)	1014	1011	1004	996	989	986	989	996	1004	1011	
		(ii)	1013	1013	1008	1000	992	987	987	992	1000	1008	
1	6	(i)	1002	1001	1001	1000	999	998	999	1000	1001	1001	
		(ii)	1002	1002	1001	1000	999	998	998	999	1000	1001	
2	5	(i)	2027	-	2008	-	1978	-	1978	-	2008	-	
		(ii)	2022	-	2022	-	1992	-	1973	-	1992	-	
2	6	(i)	2003	-	2001	-	1998	-	1998	-	2001	-	
		(ii)	2002	-	2002	-	1999	-	1997	-	1999	-	
5	5	(i)	5046	-	-	-	-	4954	-	-	-	-	
		(ii)	5000	-	-	-	-	5000	-	-	-	-	
5	6	(i)	5005	-	-	-	-	4995	-	-	-	-	
		(ii)	5000	-	-	-	-	5000	-	-	-	-	

- * (i) μ equal to a rounding unit
(ii) μ equal to a boundary value of a rounding interval

Table 3.1: Summary of extremes of variation in p_m ($\times 10^4$) when the standard deviation of the assumed blood pressure distribution is 5 or 6.

rounding unit or to a boundary value of a rounding interval. Thus, in table 3.1, the extremes in variation in the values of p_m are shown for three rounding practices when the standard deviation, σ , is assumed to be 5 or 6. For example, if it is practice to round to the nearest even digit and $\sigma=5$, then when μ is equal to a rounding unit (that is, some multiple of 2), 20.27% of final digits should be zeros, 20.08% should be twos, and so on. These figures show only small departures from 20% whereby all allowable

digits are equally favoured. Similar comments apply to the other rounding practices. For larger values of σ than 6, the values of p_m are always equal, at least to four significant figures. If σ is smaller than 5, then the effect would be larger but this is not of practical relevance in the blood pressure setting. Thus it can be concluded that practices of rounding blood pressures should lead to distributions of final digits in which each allowable digit is equally likely to be reported. Hence the patterns observed in the three towns are not entirely consistent with any one formal rounding practice but must reflect a mixture of different practices.

It is important also to assess how rounding practices affect estimates of the blood pressure distribution parameters, μ and σ^2 . The expected value, μ_R , and the variance, σ_R^2 , of the recorded blood pressure distribution can be found:

$$\mu_R = \sum_{n=-\infty}^{n=+\infty} \sum_m [10n+m] p_{nm}$$

$$\sigma_R^2 = \sum_{n=-\infty}^{n=+\infty} \sum_m [10n+m]^2 p_{nm} - \mu_R^2$$

These can be compared with the mean, μ , and variance, σ^2 , of the assumed distribution of blood pressures.

Consider first the relationship between μ_R and μ . In general, for each of the four rounding practices and for most realistic values of the blood pressure standard deviation ($\sigma \geq 5$), there is no bias observable to four significant figures: that is μ_R and μ are equal to this degree of accuracy. The only exception occurs when rounding is to the nearest ten and σ is less than 7. Then, the departure of μ_R from μ is, for example,

maximally 0.0002 when $\sigma=7$, 0.0026 when $\sigma=6$ and 0.0229 when $\sigma=5$: these values occurring when μ equals a rounding unit plus $\tau/2$ (when μ_R underestimates μ) and when μ equals a rounding unit minus $\tau/2$ (when μ_R overestimates μ). Even in these circumstances this bias might be considered negligible when dealing with a blood pressure distribution with a mean of about 80.

Given, in most cases, the uniformity of p_m as m varies, it is not surprising that the relationship between σ_R^2 and σ^2 is generally simple. In effect, rounding error when σ is sufficiently large compared with τ is akin to adding on to the normal distribution an error term that is uniformly distributed on the interval $(-\tau, +\tau)$. The variance of this additional error is $\tau^2/3$ and, to four decimal places, it can be shown, numerically, that

$$\sigma_R^2 = \sigma^2 + \tau^2/3$$

for σ/τ sufficiently large. Departures from this relationship can be found and, in extreme and unrealistic cases, could produce values of σ_R^2 close to zero (if μ is on a rounding unit) or tending to infinity (if μ is on the boundary of a rounding interval). For all practical values of σ and τ studied, the only notable departures found are when $\tau=5$ (rounding to the nearest ten) and σ was 8 or less. This situation is illustrated for values of σ from 5 to 8 in table 3.2. Even then the departure from the variance expected by assuming the addition of a uniformly distributed error term is under 2.5% for $\sigma=5$ and almost negligible for larger values of σ . Thus for all practical purposes, the rounding error can be considered as a uniformly distributed error.

σ	μ^*	σ^2	σ_R^2	$\sigma^2 + \tau^2/3$
5	(i)	25	32.5413	33.3333
	(ii)	25	34.1254	33.3333
6	(i)	36	44.2069	44.3333
	(ii)	36	44.4597	44.3333
7	(i)	49	57.3203	57.3333
	(ii)	49	57.3463	57.3333
8	(i)	64	72.3325	72.3333
	(ii)	64	72.3342	72.3333

- * (i) μ equal to a rounding unit
(ii) μ equal to a boundary value of a rounding interval

Table 3.2: Extreme cases of rounding to the nearest ten ($2\tau=10$) when the measured variance, σ_R^2 , is not equal to $(\sigma^2 + \tau^2)/3$ (to four decimal places).

Returning to the blood pressure data from the clofibrate trial, it is clear that the varying preferences for different digits are not entirely consistent with any of the rounding patterns studied. Note that the small degree of skewness suggested by some authors for the distribution of diastolic blood pressure (for example, Armitage, Fox et al, 1966) would not account for the departures from uniformity observed. In effect what is practised can be considered as mixtures of the different rounding practices. For example, in Edinburgh the underlying practice is rounding to the even digit but with additional preference for 4, and more particularly, for 0 imposed on top. In effect the preference for 4 might be because doctors are rounding to the nearest 5 or 0 and are then being forced to round the 5's to an alternative even digit. Clearly such a

model is speculative and difficult to assess. However, the numerical results imply that any mixture of the rounding practices studied should give approximately unbiased estimates of μ and overestimates of σ^2 by an amount which is a weighted average of the variances of the uniform distributions associated with each practice.

So far the parameters μ and σ have been considered primarily as characteristics of the population blood pressure distribution. However, they can also be considered as parameters defining variability of blood pressure within an individual. In this case, greater attention must be placed upon smaller values of σ , then the within-person standard deviation for blood pressure. Even then, though, the studies described in section 3.2 suggest values of 6 or 7mmHg for σ and so the expected value μ_R for recorded blood pressures on a subject should be very close to μ , their underlying mean level, and σ_R^2 will be approximately $\sigma^2 + \tau^2/3$. However, the component of variance attributable to rounding, $\tau^2/3$, is in relative terms more important particularly if rounding is to the nearest ten as is predominantly the case in Prague.

3.5 Blood pressure levels by treatment group

In order to study within-person blood pressure variability using data from the clofibrate trial, it is necessary to consider both the effect that the cholesterol criterion for entry into the trial and the effect that treatment, itself, have on blood pressure levels. Table 3.3 shows the means and standard deviations for blood pressures measured at each visit by treatment group. As the majority of subjects have visit V₄ as the first visit in the trial completed before commencing treatment, comparison of the parameters by group gives an indication of the association between cholesterol and

Treatment group	V ₋₄	V ₋₃	Visit V ₋₂	V ₋₁	V ₀
I: clofibrate high cholesterol (n=3808)	86.93 (10.15)	83.48 (9.83)	83.31 (9.97)	83.80 (10.23)	83.45 (10.27)
II: placebo high cholesterol (n=3788)	86.88 (10.04)	84.91 (9.89)	83.77 (10.00)	84.85 (10.14)	84.61 (10.18)
III: placebo low cholesterol (n=3703)	85.04 (9.89)	83.15 (9.77)	82.38 (9.82)	82.66 (10.06)	82.62 (9.95)
all groups (n=11299)	86.29 (10.07)	83.85 (9.86)	83.16 (9.95)	83.78 (10.18)	83.57 (10.17)

Table 3.3: Mean (s.d.) of diastolic blood pressures at each visit by treatment group.

blood pressure levels in this population. It is clear that subjects with cholesterol levels in the upper third of the cholesterol distribution (groups I and II) have blood pressures raised on average by 1.87 mmHg above subjects in the lower third (group III) (95% confidence interval 1.47 to 2.26 mmHg). However, the difference in levels relative to variability in blood pressure suggests that the cholesterol selection criterion should not produce a major change in the blood pressure distribution from that obtainable had the whole population been sampled (that is, disregarding cholesterol levels).

Moving across the rows in table 3.3, mean changes in blood pressure with time by treatment group can be assessed. In both control groups (II and III), a fall of about 1.9 mmHg from visit V₋₄ to V₋₃ is seen. Similarly, there is a drop in group I, though a larger one, of 3.5 mmHg. The difference between the changes in groups I and II is significant ($P < 0.0001$) and reflects an established effect of clofibrate in causing a slight lowering of blood pressure. Although there are changes in level between visits V₋₃ and V₀, there are no further differences in behaviour between treatment groups nor so noticeable and consistent changes as seen between visits V₋₄ and V₋₃. This “placebo” effect might be explained by the “pressor” effect as the subjects become used to having their blood pressures measured. As such a measurement characteristic introduces a further component of variability which confounds the analysis of trends and other more natural residual variation, data from visit V₋₄ will not be considered further. Note though that this means that the cohort being analysed is one in which subjects are familiar with blood pressure measurement. However, the analysis will include all three treatment groups. This makes the assumption that there is no impact of clofibrate on blood pressure beyond its initial lowering effect. Again, this introduces an artificial element of variability into the between-person blood pressure distribution making it a little less representative of a true population sample.

3.6 Within-person variability

Because of the differing degrees of digit preference and as there is no reason to believe that blood pressure behaviour will be the same in each town, analysis of trends and variability will be described by town. Table 3.4 gives the means and standard deviations by visit together with the tracking correlation structure. Clearly, the three

	V ₋₃	V ₋₂	Visit V ₋₁	V ₀	all visits
Edinburgh (n=3424)					
mean	82.53	81.80	82.53	82.82	82.42
s.d.	10.21	10.31	10.44	10.43	

correlation	V ₋₃ 1.000	0.596	0.560	0.538	
	V ₋₂	1.000	0.600	0.560	
	V ₋₁		1.000	0.623	
	V ₀			1.000	

Budapest (n=3864)					
mean	84.95	83.58	84.02	82.97	83.88
s.d.	10.02	10.18	10.03	10.01	

correlation	V ₋₃ 1.000	0.450	0.455	0.467	
	V ₋₂	1.000	0.478	0.488	
	V ₋₁		1.000	0.480	
	V ₀			1.000	

Prague (n=4011)					
mean	83.92	83.92	84.61	84.78	84.31
s.d.	9.26	9.27	9.75	9.98	

correlation	V ₋₃ 1.000	0.426	0.402	0.377	
	V ₋₂	1.000	0.449	0.398	
	V ₋₁		1.000	0.454	
	V ₀			1.000	

Table 3.4: Means, standard deviations and correlation structure of diastolic blood pressure (mmHg) by town.

towns exhibit small differences in mean level and in the degree of variability: these shall be returned to later. Note the relative stability in the mean levels and, particularly, in the variability between visits within towns. Also, although the correlations between visits varies in size between towns, the nature of the correlation structure is similar. The differences between towns can be attributed, in part, to different levels of within-person variability caused by measurement digit preferences and also to the different age distributions resulting in different levels of between-person variability. Correlations between pairs of visits a year apart are reasonably constant though there is a suggestion that this is slightly higher between later visits. This may reflect the fact that the time interval between later visits is less variable than intervals between earlier visits. A similar stability is obtained between pairs of visits two years apart. Note that the correlations for Edinburgh are very close to those found in the Framingham Study (Gordon and Shurtleff, 1973) where rounding to the nearest even digit was also practised. Looking at the structure across rows or up columns, a decreasing correlation is clear as the interval of time between pairs of visits increases. This suggests varying trends in blood pressure within-subjects which would alter the relative positions of individuals within the distribution of blood pressures in the population as time passes by.

To study the existence and extent of such trends a simple linear regression can be fitted, by the method of least squares, to each individual's sequence of four blood pressures from visit V_{-3} through to visit V_0 . Three parameters can then be used to summarize blood pressure changes over this period of time for each individual: the mean level, trend and the residual variance about the fitted regression line. These are given in table 3.5 including a breakdown by treatment group. For this and all

	Treatment group			
	I	II	III	all groups
<hr/>				
<u>Edinburgh</u>				
mean level (mmHg)	82.20 (8.36)	83.57 (8.73)	81.49 (8.48)	82.42 (8.56)
trend (mmHg/year)	0.08 (3.18)	0.28 (3.29)	0.13 (3.18)	0.16 (3.22)
residual variance (mmHg ²)	43.97 (51.00)	41.53 (43.71)	40.01 (43.38)	41.83 (46.17)
number	1138	1141	1145	3424
<hr/>				
<u>Budapest</u>				
mean level (mmHg)	84.58 (7.92)	84.68 (7.85)	82.37 (7.59)	83.88 (7.86)
trend (mmHg/year)	-0.61 (3.22)	-0.56 (3.37)	-0.48 (3.19)	-0.55 (3.26)
residual variance (mmHg ²)	56.75 (59.99)	55.79 (59.14)	53.36 (58.25)	55.30 (59.13)
number	1271	1311	1282	3864
<hr/>				
<u>Prague</u>				
mean level (mmHg)	83.60 (7.31)	85.23 (7.05)	84.13 (7.07)	84.31 (7.18)
trend (mmHg/year)	0.61 (3.43)	0.36 (3.46)	-0.02 (3.42)	0.33 (3.44)
residual variance (mmHg ²)	50.89 (51.48)	49.58 (48.78)	50.69 (53.99)	50.39 (51.41)
number	1399	1336	1276	4011
<hr/>				

Table 3.5: Mean (s.d.) of linear regression parameters fitted to each person's sequence of diastolic blood pressures from visits V_{-3} through to V_0 .

subsequent analyses the explanatory variable, time, took values $-3/2$, $-1/2$, $+1/2$ and $+3/2$ years at visits V_{-3} to V_0 , respectively. This greatly simplifies and aids understanding of some of the later analyses looking at components of variance and is maintained throughout for consistency.

Considering first the residual variance, it is clear that variability does not appear to be related to treatment group: all three towns have mean residual variances that are very similar between the three groups. The only oddity is the larger standard deviation of the residual variance for the clofibrate group (I) in Edinburgh. Between the three towns the variation in mean residual variance is larger than would be expected if it were to be accounted for by rounding errors. Recall that rounding, in extreme, to the nearest ten would impose an additional variance component of 8.33. As the rounding practices are more complex, but less extreme, the expected differences between towns would be less marked. Thus it might be reasonable to assume different levels of natural within-subject variability between towns.

Grouping individuals into 5mmHg bands according to their mean blood pressure level allows an assessment of the relationship between the observed residual standard deviation, s , and mean level, \bar{x} . Figure 3.3 illustrates this, suggesting increasing variability with level. Also shown on each plot is the least squares regression of s on \bar{x} derived using the ungrouped data. The apparent departures of the observed data from the fitted lines at the extremes is attributable to small numbers of subjects with such mean pressure levels. The fitted lines are:

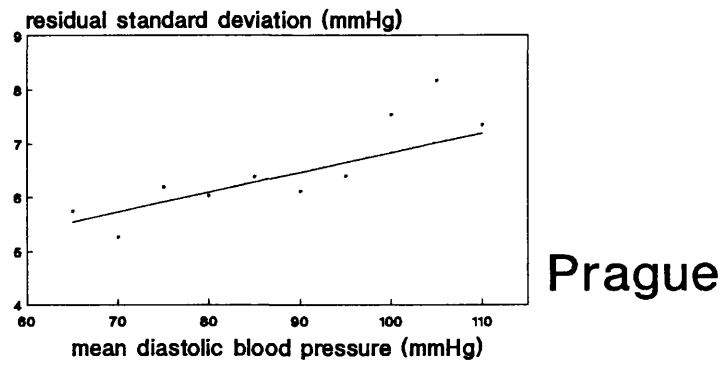
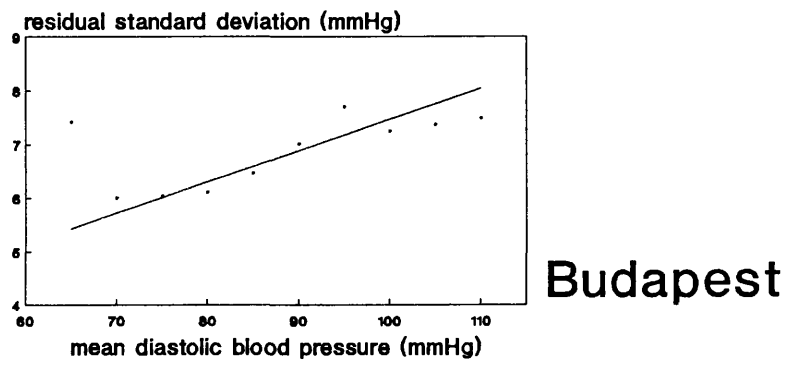
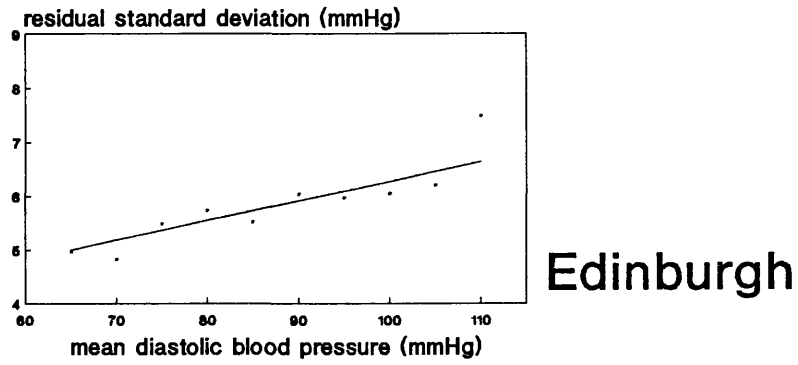


Figure 3.3: Relationship between observed residual variance and mean level

$$\begin{aligned}
& 2.6241 + 0.0365 \bar{x} && \text{(Edinburgh)} \\
s = & 1.6634 + 0.0580 \bar{x} && \text{(Budapest)} && (3.1) \\
& 3.1607 + 0.0367 \bar{x} && \text{(Prague)}
\end{aligned}$$

Thus the standard deviation increases by about 0.5 mmHg for each 10 mmHg rise in mean pressure. As the distribution of observed residual standard deviations about the least squares regressions are certainly non-normal, the usual significance tests for the gradients are not appropriate: however, formal testing is considered later in section 3.9. In addition, the mean blood pressure refers to the mean over four observations and this, itself, is measured with error so that, in general, increasing the number of observations on which the mean is based might be expected to increase the associated regression coefficient.

Although a formal model for blood pressure variability within an individual will be fitted later (section 3.9), it is convenient to consider here whether the variability in observed within-person variances, after allowing for blood pressure level, can be explained by sampling variation. To achieve this, consider the simple model whereby diastolic blood pressures within an individual, i , are normally distributed about some underlying level, then

$$\frac{2s_i^2}{\sigma_i^2} = \frac{\sum_{j=-3}^0 [x_{ij} - \hat{x}_{ij}]^2}{\sigma_i^2} \sim \chi_2^2$$

where \hat{x}_{ij} is the predicted diastolic pressure at visit j (obtained from the within-person

regression through the four observations) and σ_i^2 is the predicted within-person variance for subject i which depends upon their underlying diastolic pressure level. Given the large cohort size being studied, σ_i^2 could be replaced by the values predicted from the models for s_i described by (3.1). However, it is well known that the observed standard deviation, s , is a biased estimator of σ and that the bias when it is based upon a small sample size is substantial: here, two degrees of freedom are available for estimating σ_i and so s_i needs to be increased by a factor of 1.1284 to give an unbiased estimate of σ_i (Pearson and Hartley, 1954). Applying this adjustment to the coefficients in (3.1) gives

$$\begin{aligned} & 2.6241 + 0.0365 \bar{x} && \text{(Edinburgh)} \\ \sigma = & 1.6634 + 0.0580 \bar{x} && \text{(Budapest)} && (3.2) \\ & 3.1607 + 0.0367 \bar{x} && \text{(Prague)} \end{aligned}$$

where \bar{x} is the mean of four observations. Calculating the ratio $2s_i^2/\sigma_i^2$ for each individual, comparison of their distribution with a chi-squared distribution on two degrees of freedom can be made using a Kolmogorov-Smirnov test. This gives test statistics, D , of 0.021 ($0.1 > P > 0.05$), 0.012 ($P > 0.2$) and 0.016 ($P > 0.2$) for Edinburgh, Budapest and Prague, respectively. This good fit suggests that a model with observed diastolic pressures normally distributed about an underlying mean level is appropriate.

3.7 Within-person trends

Returning to table 3.5, it is clear that average trends do vary between the three towns. Both Edinburgh and Prague show increasing trends whereas there is a mean decline of

over 0.5mmHg per year in Budapest. This latter result is unusual in that diastolic blood pressure generally rises slowly with age in developed countries (Gordon and Shurtleff, 1973; Shaper et al, 1988). In both Edinburgh and Budapest, analysis of variance shows that there is no significant variation between the mean trends by treatment group ($P=0.32$ and $P=0.60$, respectively). In contrast the variation in Prague is highly significant ($P<0.0001$): there being no apparent mean trend in group III (low cholesterol, placebo), but increasing trends in both of the high cholesterol groups particularly in the one receiving clofibrate (group I). Explaining this difference between treatment groups is difficult particularly as the ordering between groups differs from town to town and there is no apparent imbalance in other factors that might explain it. Therefore it might be reasonable to assume that there is no inherent variation in average trends between treatment groups.

The standard deviations of the observed trends shown in table 3.5 by treatment group within-town are very similar in magnitude and also show little difference between towns. It is useful, then, to assess whether this is attributable to sampling variation or whether there is evidence of heterogeneity, between subjects, in underlying diastolic pressure trends. First, though, consider the distributional form of the observed trends. Figure 3.4 shows normal probability plots, by town, for the trends (ignoring treatment group). For clarity these are shown in a cumulative form stepping at intervals of 0.2mmHg/year on the trend scale. Thus there is a small exaggeration of departures from the assumed normal model which is shown by the dashed line. Clear, though is the closeness of the observed distribution to normality with a handful of subjects in each town showing, in extreme, larger trends than expected if the normal distribution is a correct description. Goodness of fit can be further assessed using Kolmogorov-

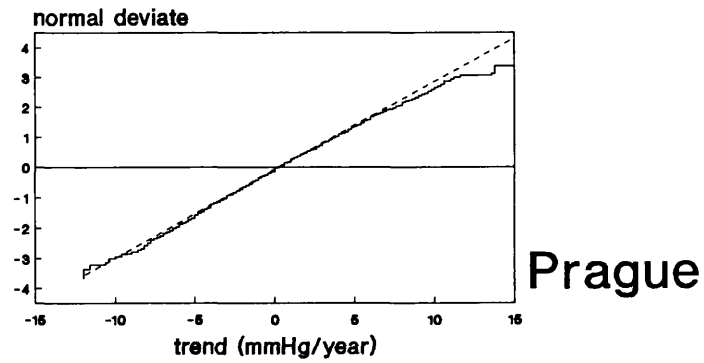
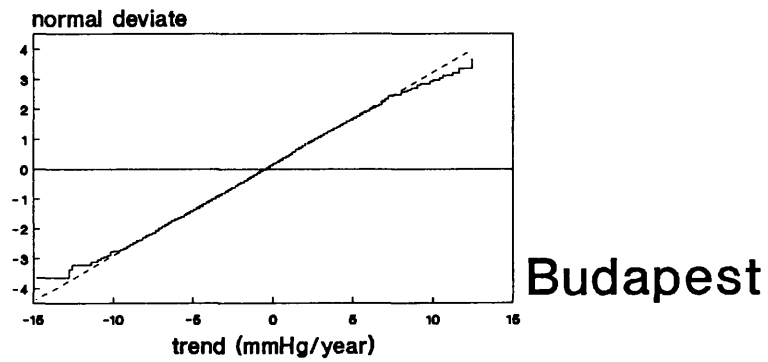
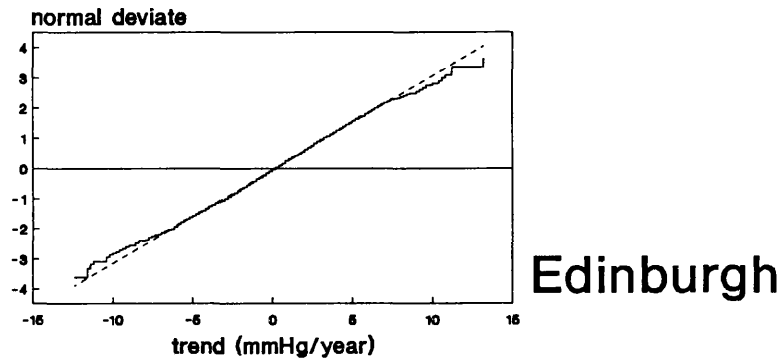


Figure 3.4: Normal probability plots for the observed trends within-person

Smirnov tests. These show mildly significant departures from normality (Edinburgh: test-statistic, $D=0.025$, $0.05 > P > 0.01$; Budapest: $D=0.021$, $0.1 > P > 0.05$; Prague: $D=0.029$, $0.01 > P > 0.005$). Thus a model for normally distributed trends would seem a reasonable approximation.

To assess the evidence for heterogeneity of trends, the sampling variation of the observed trends needs to be considered. As the trends were calculated assuming visit times of $-3/2$, $-1/2$, $1/2$ and $3/2$ years (for visits V_{-3} to V_0 , respectively), the variance of the estimated trend, $\hat{\beta}_i$, for individual i is

$$\text{var}(\hat{\beta}_i) = \frac{\sigma_i^2}{5} = \sigma_{\beta}^2(\bar{x}), \text{ say.}$$

But σ_i^2 has already been seen to depend on the mean level of blood pressure for the individual concerned. Therefore the sampling distribution for the estimated trends, $\hat{\beta}$, over all individuals assuming no heterogeneity can be described in terms of the hierarchical model:

$$\begin{aligned} \hat{\beta} &\sim N[\beta, \sigma_{\beta}^2(\bar{x})] = f(\beta | \bar{x}) \\ \bar{x} &\sim f(\bar{x}) \end{aligned}$$

where \bar{x} is the observed mean level. Thus the sampling distribution for $\hat{\beta}$ is

$$f(\hat{\beta}) = \int_{\bar{x}} (2\pi)^{-1/2} \sigma_{\beta}^{-1} \exp\left[-\frac{(\hat{\beta} - \beta)^2}{2\sigma_{\beta}^2(\bar{x})}\right] f(\bar{x}) d\bar{x}.$$

If the dependence of σ_{β} on \bar{x} is not too strong, as can be seen from (3.2), then a first

order approximation gives $f(\hat{\beta})$ as a normal distribution which would agree with that observed. In addition

$$E(\hat{\beta}) = \int_{\hat{\beta}} \hat{\beta} \int_{\bar{x}} f(\hat{\beta} | \bar{x}) f(\bar{x}) d\bar{x} d\hat{\beta}$$

which, on reversing the order of integration, gives $E(\hat{\beta}) = \beta$; that is, $\hat{\beta}$ is unbiased. In a similar manner, it follows that

$$\text{var}(\hat{\beta}) = E[\sigma_{\hat{\beta}}^2(\bar{x})]$$

which can be approximated, given the large sample size, by the mean residual variance divided by 5. This gives theoretical sampling variances for the observed trends as 8.37, 11.06 and 10.08 for Edinburgh, Budapest and Prague, respectively.

Using the assumption of normality, under the hypothesis that the observed distribution is attributable to sampling error in estimating the trends, alone:

$$\frac{(n-1) \times \text{observed variance of trends}}{\text{var}(\hat{\beta})} \sim \chi_{(n-1)}^2 \approx N[(n-1), 2(n-1)]$$

where n is the number of individuals in the town and is sufficiently large to justify the normal approximation. For each town, this ratio can be obtained and hence a normal deviate to test the hypothesis. These are 9.86, -1.67 and 7.93 for Edinburgh, Budapest and Prague, respectively. As the alternative hypothesis being considered is of heterogeneity in the trends, a one-tailed test is appropriate. Thus it is clear that there is highly significant overdispersion in both Edinburgh and Prague ($P < 0.0001$) but

not in Budapest.

In the two towns with apparently real variation in trends, it is useful to attempt to quantify the level of variation. Taking the differences between the variance of the observed trends and the variance attributable to sampling, and square-rooting, gives estimated standard deviations of 1.41mmHg/year and 1.34mmHg/year for Edinburgh and Prague respectively. Thus the variability in the real underlying trends is much smaller than the variability in observed trends given in table 3.5. From a practical point of view this clearly implies that large observed trends based on a limited number of observations need to be interpreted with considerable care particularly when they may affect treatment decisions.

As with residual variance, a check of the relationship between mean levels and size of trend can be made. Figure 3.5 shows plots of the average observed trends against mean level grouped by 5mmHg bands. Also shown is the least squares regression. Clear from the latter is that there is a very weak positive correlation between trend and mean level. In the three towns, Edinburgh, Budapest and Prague, trends increase by 0.12 ($P=0.045$), 0.08 ($P=0.179$) and 0.40mmHg/year ($P<0.001$) for each 10mmHg change in mean level. Given that the standard deviation of mean levels is about 8mmHg in the three towns, the magnitude of increase in trends across the range of blood pressures observed is small, and possibly negligible, compared with the observed mean trends in the cohort.

As trends are so small, it is useful to consider the merit in ignoring them altogether. For the purposes of screening and monitoring patients where the timescale considered

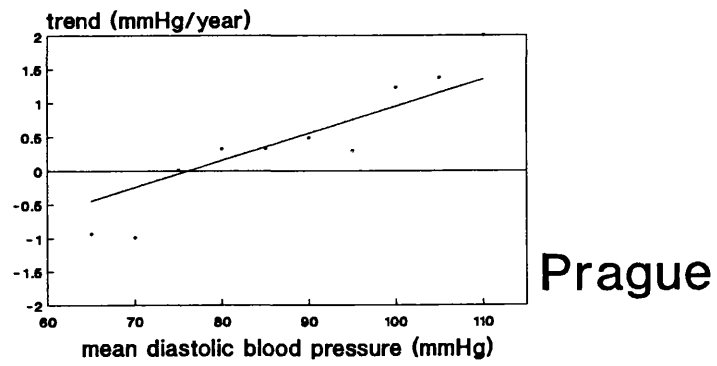
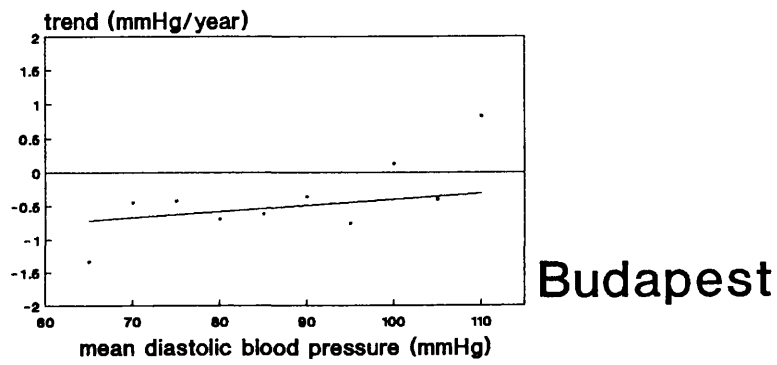
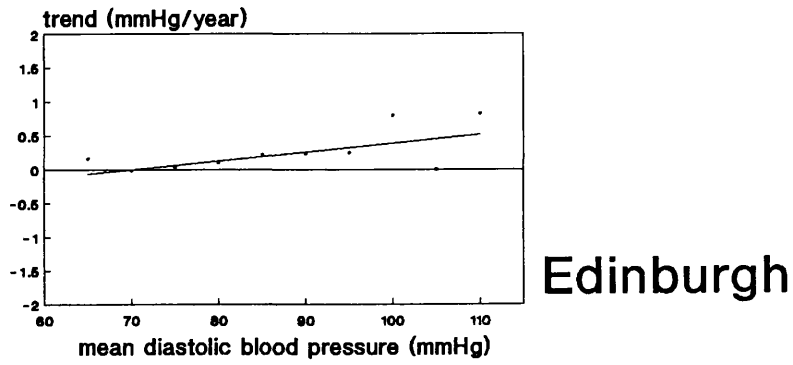


Figure 3.5: Relationship between observed trends and mean level

will generally be short, the impact of any real trend will be more or less negligible. From a statistical point of view, it also enables an extra degree of freedom in the estimation of the within-person residual variance. Repeating the analysis leading to the relationship (3.2) gives the following relationship between residual variance (ignoring trends) and observed mean pressure (from 4 measurements):

$$\begin{aligned} \sigma_w(\bar{x}) &= 3.0314+0.0434\bar{x} && \text{(Edinburgh)} \\ &= 1.6318+0.0681\bar{x} && \text{(Budapest)} \\ &= 2.3677+0.0580\bar{x} && \text{(Prague)} \end{aligned}$$

Comparing this with (3.2) shows that there is a small increase in the standard deviation and, particularly in Prague, its rate of rise with level which is attributable to trends. For comparison with the models derived in section (3.9), it is useful to rewrite this as a relationship between the variance $\sigma_w^2(\bar{x})$ and the departure of an individual's mean, \bar{x} , from the population mean:

$$\begin{aligned} \sigma_w^2(\bar{x}) &= 45.1940 [1+0.0066(\bar{x} - 82.4198)]^2 && \text{(Edinburgh)} \\ &= 55.1040 [1+0.0093(\bar{x} - 83.8780)]^2 && \text{(Budapest)} \\ &= 53.5391 [1+0.0080(\bar{x} - 84.3081)]^2 && \text{(Prague)} \end{aligned} \quad (3.3)$$

3.8 Blood pressure variability dependence on age, body mass index and smoking

The purpose of this section is to describe the relationships observed between the blood pressure variability parameters (mean level, trend and residual variance) and the three available descriptive variables: age, body mass index and smoking. Without the benefit

of a general population sample, it is not possible to draw definitive conclusions about these relationships, though it would be reasonable to expect the nature of any association found, if not its magnitude, to be generalizable to other cohorts of middle-aged and apparently healthy men.

3.8.1 Age

Table 3.6 shows the means and standard deviations for the variability parameters by the age-groups previously used. Recall that age refers to visit V_0 , the last of the four visits. Looking at the mean levels, each town shows the rising pattern of blood pressure with age that is well established. Edinburgh and Budapest suggest blood pressures higher in older age groups by about 1mmHg or just under for each additional five years. In Prague the effect is similar if the two outer age-groups, which have very small numbers of individuals, are ignored. Note that in both Edinburgh and Prague, the oldest age group (≥ 60 years) have lower mean blood pressures than the groups in their 50's. This is most likely a consequence of the exclusion criteria requiring good health for entry to and continuance in the trial so that more elderly men with raised blood pressures are more likely to have diagnosed major disease, and so be excluded from the trial, than younger men with similar levels. Also of interest and more apparent in Edinburgh and Prague is the increasing variance between subjects as age increases.

The picture for trends is interesting in that it does not correspond well with what might be expected from considering the mean levels. Both Budapest and Prague show mean within-person trends declining in value with increasing age though in Budapest

	Age at visit V_0						all ages
	≤ 39	40-44	45-49	50-54	55-59	≥ 60	
<u>Edinburgh</u>							
mean level (mmHg)	80.7 (7.9)	81.8 (8.3)	82.8 (8.4)	83.1 (9.0)	84.1 (8.8)	82.5 (9.2)	82.4 (8.6)
trend (mmHg/year)	0.00 (3.29)	0.16 (3.19)	0.28 (3.24)	0.06 (3.19)	0.47 (3.08)	-0.13 (3.34)	0.16 (3.22)
residual variance (mmHg ²)	38.7 (46.1)	41.7 (44.8)	43.3 (45.8)	41.6 (47.0)	44.0 (45.7)	42.1 (50.4)	41.8 (46.2)
number	576	702	814	704	408	220	3424
<u>Budapest</u>							
mean level (mmHg)	81.3 (7.7)	82.6 (7.4)	83.6 (7.7)	85.0 (7.9)	84.5 (8.0)	85.1 (7.8)	83.9 (7.9)
trend (mmHg/year)	-0.10 (3.22)	-0.36 (3.15)	-0.37 (3.31)	-0.76 (3.30)	-0.62 (3.27)	-0.92 (3.20)	-0.55 (3.26)
residual variance (mmHg ²)	52.8 (52.9)	53.3 (55.8)	51.0 (52.8)	56.4 (61.8)	59.6 (62.1)	59.1 (66.9)	55.3 (59.1)
number	445	539	828	869	574	609	3864
<u>Prague</u>							
mean level (mmHg)	-	84.4 (7.2)	83.1 (6.8)	84.6 (7.1)	85.3 (7.7)	81.3 (3.5)	84.3 (7.2)
trend (mmHg/year)	-	1.65 (2.24)	0.53 (3.36)	0.27 (3.50)	0.21 (3.36)	0.08 (3.54)	0.33 (3.44)
residual variance (mmHg ²)	-	40.6 (33.8)	48.2 (50.1)	51.3 (52.2)	50.9 (51.0)	37.1 (55.3)	50.4 (51.4)
number	0	10	998	2315	674	13	4011

Table 3.6: Mean (s.d.) of linear regression parameters fitted to each person's sequence of diastolic blood pressure from visits V_{-3} through to V_0 , by age group.

the mean trends are all negative whereas in Prague they are all positive. This suggests that the observed rises in mean levels of blood pressure with age are either a cohort effect, or the pattern of trends is particularly marred by the healthy "in-trial" individual effect mentioned above, so that individuals with increasing trends particularly at older ages are excluded. In Edinburgh, there is no such clear pattern of trends though the negative mean trend for those aged 60 or over ties up with the lower mean levels in that group. In all three towns the standard deviations of the observed trends do not vary notably between age groups. However, this is not surprising given that the variability is dominated by sampling error.

Residual variation appears to increase in varying degrees with age in all three towns. Most of this rise can be accounted for by the corresponding rise in mean level with age. From (3.2), this is between 0.5 and 1.0mmHg² for a unit increase in blood pressure at the levels observed. The only anomaly is in Budapest where the two older age groups are a little higher than might be expected.

3.8.2 Smoking status

As there are so few smokers of tobacco and/or cigars as distinct from cigarettes in Budapest and Prague, this analysis excludes those subjects who currently smoke or previously have smoked only tobacco and/or cigars. This enables a ready comparison between towns with easily recognizable categories of smoking status. Table 3.7 shows the means and standard deviations for each of the blood pressure variability parameters, by smoking status. Looking at both the trends and residual variance, there is no clear pattern of change between smoking status or by dose. With respect to

Smoking category at V_0

	Never*	Ex- cigarette*	Current cigarette smokers				
			1-10	11-19	20	21-39	>40
<u>Edinburgh</u>							
mean level (mmHg)	82.0 (8.4)	83.3 (8.7)	82.1 (9.0)	81.6 (8.3)	81.5 (7.9)	82.1 (7.8)	83.8 (7.7)
trend (mmHg/year)	0.18 (3.14)	0.19 (3.39)	0.11 (3.13)	0.12 (3.22)	0.07 (2.86)	0.11 (3.25)	-0.55 (3.33)
residual variance (mmHg ²)	41.9 (48.6)	42.5 (45.4)	45.2 (50.0)	37.8 (41.4)	37.4 (45.7)	42.8 (48.3)	52.6 (51.3)
number	915	849	342	276	327	192	42
<u>Budapest</u>							
mean level (mmHg)	84.6 (8.0)	84.7 (7.8)	83.9 (8.0)	82.2 (7.8)	83.1 (7.8)	82.8 (7.5)	82.7 (7.7)
trend (mmHg/year)	-0.54 (3.27)	-0.56 (3.43)	-0.22 (3.21)	-0.66 (3.12)	-0.69 (2.98)	-0.51 (3.26)	-0.61 (3.14)
residual variance (mmHg ²)	58.6 (61.6)	57.5 (62.4)	53.2 (63.6)	53.9 (57.2)	49.8 (54.2)	50.9 (51.7)	50.7 (47.7)
number	1145	967	338	305	400	533	123
<u>Prague</u>							
mean level (mmHg)	85.3 (7.2)	84.7 (7.3)	83.4 (6.8)	83.9 (7.4)	83.2 (7.1)	83.2 (6.6)	84.5 (7.4)
trend (mmHg/year)	0.34 (3.51)	0.17 (3.54)	0.34 (3.35)	0.35 (3.39)	0.50 (3.26)	0.27 (3.52)	0.05 (3.39)
residual variance (mmHg ²)	49.8 (51.7)	50.8 (51.2)	52.0 (55.4)	51.0 (51.5)	48.9 (47.9)	50.3 (48.9)	51.8 (61.1)
number	1260	786	490	357	562	387	80

Table 3.7: Mean (s.d.) of linear regression parameters fitted to each person's sequence of diastolic blood pressures from visits V_{-3} through to V_0 , by smoking status.

* excludes subjects currently or previously only cigar/tobacco smokers.

mean levels there appears, in Budapest and Prague, a small lowering of blood pressure (of between 1 and 2mmHg) in current smokers compared with both ex-smokers and those who have never smoked; the latter two groups being very similar. In Edinburgh, there is no apparent difference between smokers and the group who have never smoked. However the ex-smokers have a mean level slightly raised above that of those that have never smoked. In none of the towns is there any evidence of a relationship of blood pressure level with dose. Thus apart from a possible, and mild, reduction in blood pressure in those who smoke compared with those who don't, smoking has no other clear and consistent effect on blood pressure. This is compatible with the previous studies of smoking and blood pressure mentioned earlier.

3.8.3 Body mass index

Body mass index is only available at entry to the trial and so changes during the period in which the blood pressures were measured cannot be assessed. However, it might be expected to be reasonably constant over a three year period. In table 3.8, means and standard deviations for each of the blood pressure regression parameters by body mass index are given. The clearest relationship, visible in each of the three towns, is the increasing level of mean blood pressure with body mass index. Regressing mean level on the index using the method of least squares, the rate of increase is found to be 0.77, 0.83 and 0.50mmHg in Edinburgh, Budapest and Prague, respectively, for a unit increase in the index. This is of a similar order of increase to that described by Shaper et al (1988) in the British Regional Heart Study. Such an increasing level also explains the increasing residual variance with body mass index.

	Body mass index at entry to trial (kg/m ²)							all values
	≤21	>21 ≤23	>23 ≤25	>25 ≤27	>27 ≤29	>29 ≤31	>31	
Edinburgh								
mean level (mmHg)	77.5 (7.5)	80.2 (7.6)	81.2 (8.4)	83.3 (8.4)	84.5 (8.4)	86.0 (9.1)	87.3 (7.3)	82.4 (8.6)
trend (mmHg/year)	0.13 (3.12)	0.14 (3.03)	0.17 (3.13)	0.16 (3.15)	0.15 (3.47)	0.41 (3.43)	-0.23 (3.71)	0.16 (3.22)
residual variance (mmHg ²)	39.3 (44.4)	39.5 (47.6)	41.0 (42.1)	41.6 (43.3)	44.9 (50.4)	43.1 (51.8)	50.6 (59.4)	41.8 (46.2)
number	204	567	870	905	501	206	137	3424
Budapest								
mean level (mmHg)	77.9 (6.2)	79.5 (7.4)	81.5 (7.2)	83.3 (7.3)	84.9 (7.5)	86.5 (7.2)	88.8 (7.9)	83.9 (7.9)
trend (mmHg/year)	-0.17 (2.82)	-0.24 (2.98)	-0.45 (3.15)	-0.47 (3.17)	-0.64 (3.35)	-0.71 (3.37)	-0.89 (3.59)	-0.55 (3.26)
residual variance (mmHg ²)	59.9 (73.5)	44.7 (43.4)	49.6 (53.9)	54.2 (57.6)	59.3 (63.7)	58.6 (63.8)	62.2 (59.9)	55.3 (59.1)
number	112	324	735	850	847	568	420	3864
Prague								
mean level (mmHg)	81.6 (7.4)	81.4 (6.5)	83.0 (6.9)	83.7 (7.0)	84.9 (7.0)	86.2 (7.4)	87.2 (7.0)	84.3 (7.2)
trend (mmHg/year)	0.68 (3.57)	0.32 (3.24)	0.30 (3.32)	0.46 (3.49)	0.27 (3.41)	0.12 (3.52)	0.39 (3.65)	0.33 (3.44)
residual variance (mmHg ²)	43.6 (37.4)	50.6 (52.3)	48.7 (50.5)	48.9 (47.6)	53.2 (55.0)	47.8 (47.0)	56.0 (60.7)	50.4 (51.4)
number	75	288	772	1033	915	539	384	4011

Table 3.8 Mean (s.d.) of linear regression parameters fitted to each person's sequence of diastolic blood pressures from visits V₋₃ through to V₀, by body mass index.

The picture for trends is less clear. In Edinburgh, mean trends are remarkably stable for all but the highest two body mass index groups. However these do not show any consistent pattern. In Budapest, trends become increasingly negative with body mass index and also more variable. In Prague, there is again no clear relationship. This emphasizes further the unexpected pattern of declining trends in Budapest with the greatest declines in those who might be expected to be at highest risk of cardiovascular disease: that is the older, smokers and the overweight. There is no apparent reason for this finding.

3.9 Modelling blood pressure variability

The previous sections have described some of the features of blood pressure variability. The purpose of this section is to derive a simple model for variability. As it is difficult to assess how far the study cohort departs, in character, from a genuine population sample, the emphasis will be on modelling within-person variability and treating between-person variability of secondary interest.

From the observed blood pressure data, several important features of within-person variability have been determined. These can be summarized as follows:

- (i) there is no heterogeneity in within-person variability except that it increases approximately linearly with blood pressure level;
- (ii) trends are small relative to residual variability, approximately normally distributed in the population, and approximately independent of level;
- (iii) neither residual variability nor trend appears to depend on other factors.

Mathematically, then, an observed blood pressure, x at time t , might be modelled in the following form:

$$\begin{aligned}
 x &\sim N[\xi_t, \sigma_w^2(\xi_t)], \text{ with } \sigma_w(\xi_t) = \sigma_w[1+c(\xi_t-\mu)] \\
 \xi_t &= \xi + \beta t \\
 \beta &\sim N[\theta, \sigma_\beta^2] \\
 \xi &\sim f(\mu, \phi)
 \end{aligned}
 \tag{3.4}$$

where f is the distribution of underlying levels between individuals in the samples and is dependent on parameters μ ($=E(\xi)$) and ϕ (such as age etc.), and where the values of any of the model parameters may vary between towns. Using maximum likelihood methods to estimate the parameters in this random effects model is complicated by the dependence of the within-person variability on ξ and β . Thus the marginal distribution of the observed blood pressures, x , cannot be analytically derived.

To simplify the estimation, the inclusion in the model of a distribution for trends can be removed with the knowledge that these have been shown to be small. Note that the effect of this assumption is to inflate the apparent within-person variability, σ_w^2 , by a small amount: from the observed data, ignoring the possibility of trends produces an increase in residual variance of, at most, 10% in the three towns. In addition, the distribution f might be assumed to be the normal probability density function as suggested by Boe et al (1957). However, it is worth noting that the dependence of the within-person variance on ξ induces a degree of skewness into the distribution of observed blood pressures which, therefore, confounds direct inferences about the exact form of the distribution of underlying mean blood pressure levels from that observed.

Making these assumptions, a simple but reasonable model is then:

$$\begin{aligned} x &\sim N[\xi, \sigma_w^2(\xi)], \text{ with } \sigma_w(\xi) = \sigma_w[1+c(\xi-\mu)] \\ \xi &\sim N[\mu, \sigma^2]. \end{aligned} \quad (3.5)$$

Estimates of the model parameters (σ_w^2 , c , μ and σ^2) have been obtained by maximum likelihood methods. Given four observations \underline{x}_i for person i , the likelihood for that person is:

$$l_i = (2\pi)^{-5/2} \sigma_w^{-4} [1+c(\xi_i-\mu)]^{-4} \sigma^{-1} \exp \left[-\frac{1}{2} \left[\sum_{j=1}^4 \frac{[x_{ij}-\xi_i]^2}{\sigma_w^2 [1+c(\xi_i-\mu)]^2} + \frac{(\xi_i-\mu)^2}{\sigma^2} \right] \right]$$

Here the ξ_i 's are of little interest and so the marginal likelihood, l_i^m , can be defined:

$$l_i^m = \int l_i(\xi_i) d\xi_i.$$

Then the logged marginal likelihood across all individuals is:

$$L^m = \sum_i \ln(l_i^m)$$

and its maximum can be determined with appropriate numerical integration. In practice convergence using a quasi-Newton algorithm was fast taking fewer than ten iterations despite the use of finite difference methods to obtain derivative estimates. Only limited accuracy can be obtained by this method because of the need to sum numerically determined integrals, though the parameter estimates presented here are accurate to four significant figures.

	μ	σ^2	σ_w^2	c ($\times 10^3$)	$\ln(\text{likelihood})$
Edinburgh	82.41 (0.1)	60.44 (1.4)	45.24 (0.7)	10.43 (1.5)	-48680.9
Budapest	83.87 (0.1)	46.58 (1.5)	55.00 (0.8)	12.67 (1.3)	-55732.7
Prague	84.31 (0.1)	36.66 (1.1)	53.59 (0.7)	13.88 (1.7)	-57326.9

Table 3.9 Parameter estimates (approximate standard errors) for the diastolic pressure model:

$$x \sim N[\xi, \sigma_w^2(1+c(\xi-\mu))^2],$$

$$\xi \sim N[\mu, \sigma^2]$$

In table 3.9, the estimated parameter values for the model are given for each town. In each case, the within-person variance, σ_w^2 at the mean blood pressure level, μ , is about the same as the mean observed residual variance, ignoring trends. However, the gradient of increase in residual variation is substantially bigger when related to true underlying pressures rather than the observed mean of four measurements. Here the within-person standard deviation of diastolic blood pressure increases by about 1% or so for each unit increase in level of underlying pressure.

By fitting the models with the parameter c assumed equal to zero, the statistical significance of the increasing within-person variance with level can be assessed using a

likelihood ratio test. This gives chi-squared test statistics (on one degree of freedom) of 107, 161 and 144 for Edinburgh, Budapest and Prague, respectively, which are all very highly significant.

Exact standard errors for the parameters estimated are difficult to obtain because of computational limitations in deriving the necessary matrix of second derivatives of the likelihood function. Therefore, also given in table 3.9 are approximate “standard errors” for each of the parameters estimated obtained by drawing on the parallel between confidence intervals and significance tests. For example, an approximate standard error for $\hat{\mu}$ is obtained by finding those two values of μ (μ_1 and μ_2) such that the hypothesis that $\mu = \mu_i$ ($i=1, 2$) is rejected in two-sided likelihood ratio test (for μ with the other parameters constrained to their values at the maximum) at exactly the 5% level of significance. Then, the standard error can be obtained by using the result that the likelihood is approximately normally distributed in large samples about the maximum so that the interval (μ_1, μ_2) has width two times 1.96 times the standard error of $\hat{\mu}$. This derivation, though somewhat heuristic, does enable an assessment of whether there is significant difference between the three towns in the parameter c which, due to computational limitations cannot be assessed more directly by fitting a model with a common value in the three towns. A pooled estimate, c_p , can be obtained:

$$c_p = \frac{\sum_k w_k c_k}{\sum_k w_k}$$

where c_k is the parameter estimate from town k and w_k is a weight equal to the precision of its estimation (that is, the inverse of the square of its standard error).

This gives a value of 12.25×10^{-3} with standard error 0.8×10^{-3} . A test of homogeneity of the parameter c across the three towns is then obtained using the test-statistic:

$$X_2^2 = \sum_k w_k (c_k - c_p)^2$$

which is distributed as a chi-squared on two degrees of freedom (DerSimonian and Laird, 1986). This gives $X_2^2 = 2.50$ which is not statistically significant at any conventional level. Thus the hypothesis that the rate of increase in within-person standard deviation is the same in the three towns cannot be rejected.

In contrast, there is no reason to expect the parameter σ_w^2 to be constant between the three towns because this parameter includes a component of variance attributable to the digit preferences operating. However, in section 3.4, this component of variance was found to equal $\tau^2/3$ where 2τ is equal to the width of a rounding interval. In Budapest, where digit preferences are least pronounced, this component is unlikely to be more than 1mmHg^2 and even in Edinburgh, with rounding to even digits and some additional preference for zero, it is unlikely to be greater than 2mmHg^2 . Thus a difference of nearly 10mmHg^2 between the values of σ_w^2 for Edinburgh and Budapest cannot be explained by digit preferences nor, by considering the standard errors, by sampling variation. In Prague, the large digit preference shown in figure 3.1 for the digit zero would imply a component of variance attributable to this rounding practice of the order of 5mmHg^2 . However, this still means that the Prague estimate for σ_w^2 is quite distinct in magnitude from the values in either of Edinburgh and Budapest. Thus it seems reasonable to conclude that there may be real differences, not explained by digit preferences, in the average levels of within-person variability between the three

towns. However, these differences need not necessarily represent purely physiological differences between the populations of the three towns. They could be attributable to other differences in measurement technique in the three centres possibly related to environmental factors in the clinics or to the state of rest of the subjects at the time of measurement.

It is also worth noting that the model does give a good representation of the observed correlation structure presented in table 3.4. To see this, the variance/covariance structure for observed blood pressures needs to be obtained. Using the model, the variance of the observed blood pressures is $(\sigma^2 + \sigma_w^2 + c^2 \sigma^2 \sigma_w^2)$ where the last term is, here, negligible. In addition the covariance for repeat measures on the same individual is σ^2 so that the correlation is

$$\sigma^2 (\sigma^2 + \sigma_w^2 + c^2 \sigma^2 \sigma_w^2)^{-1} \approx (1 + \sigma_w^2 / \sigma^2)^{-1}.$$

Evaluating this using the estimates obtained gives values of 0.57, 0.46 and 0.41 for Edinburgh, Budapest and Prague, respectively.

To assess the more general fit of the model, the predicted number of individuals with given blood pressures can be compared with that observed. However, to reduce the sensitivity of this comparison to digit preference, it is better to compare the distribution of observed mean blood pressures with that predicted. For illustration, figure 3.6 shows the fit, by town, with the mean levels rounded to integer values. Clear in each case is the small degree of skewness not picked up by the model. These departures are all significant ($P < 0.001$, Kolmogorov-Smirnov test). This suggests an

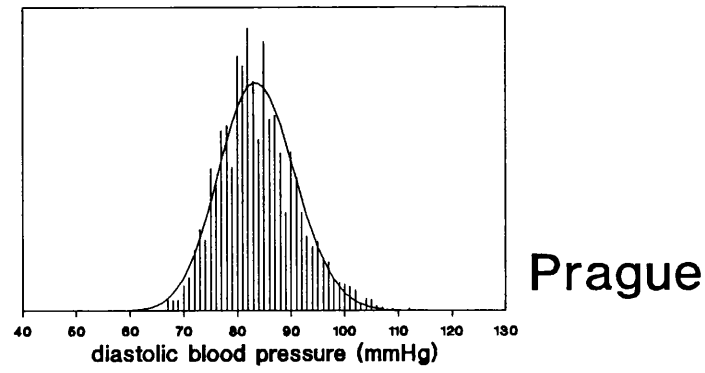
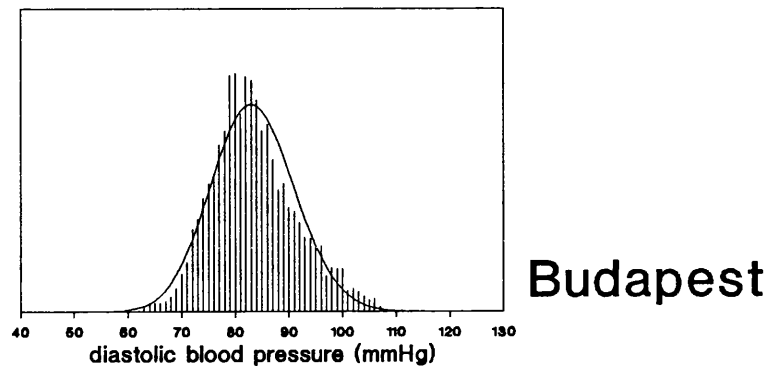
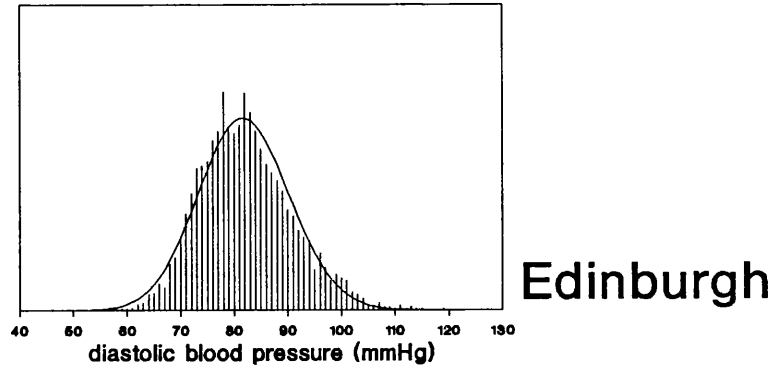


Figure 3.6: Observed and predicted distributions of blood pressure

element of skewness in the distribution of underlying mean blood pressures not incorporated in the model here.

It is difficult when using trial-based data to justify an alternative model as the cohorts studied are specially selected: in particular further skewness might be expected if the hypertensives excluded from the trial were to be included. However, one potential source of variation, that due to age, does deserve some attention. Even if blood pressures are normally distributed in the population for any given age, then a skew distribution of ages in the study cohort may introduce skewness into the distribution of blood pressures. To assess the impact of this possibility upon the estimated parameters, σ_w^2 and c , the model (3.5) can be fitted with the between-person parameters estimated within age-group: here, in five-year intervals. The results of this analysis can be found in table 3.10; for Prague, the lower age-group (≤ 49) includes the eleven men aged 44 or under. The improvement in fit of the model can be assessed by comparing the logged likelihoods achieved with those in table 3.9 using a likelihood ratio test. For each town, this is statistically significant ($P < 0.0001$ in each town). Note that both the mean level, μ and, particularly, the variability between individuals increase with age. Despite this, the goodness of fit, in practical terms, is not noticeably improved and there is only negligible change in the within-person variability parameters σ_w^2 and c .

In table 3.11, this stream of analysis is taken one step further by stratifying σ_w^2 and c by age as well. However, this does not give a statistically significant improvement in fit on the model with these parameters fixed across age-groups ($P > 0.1$, likelihood ratio test, in each town). In addition, there is no distinct pattern of change in either σ_w^2 or c

Age group	μ	σ^2	σ_w^2	c ($\times 10^3$)	$\ln(\text{likelihood})$
Edinburgh					
≤ 39 (n=576)	80.71	49.66			
40-44 (n=702)	81.80	55.74			
45-49 (n=814)	82.71	57.40			
50-54 (n=704)	83.09	68.37	45.24	10.45	-48649.5
55-59 (n=408)	84.03	61.71			
≥ 60 (n=220)	82.53	71.89			
Budapest					
≤ 39 (n=445)	81.40	43.08			
40-44 (n=539)	82.65	40.41			
45-49 (n=828)	83.69	44.37			
50-54 (n=869)	84.88	46.62	55.02	12.69	-55684.5
55-59 (n=574)	84.44	47.81			
≥ 60 (n=609)	85.03	46.50			
Prague					
≤ 49 (n=1009)	83.16	32.35			
50-54 (n=2315)	84.52	35.37	53.58	13.82	-57301.0
≥ 55 (n=687)	84.26	44.54			

Table 3.10: Parameter estimates for the diastolic blood pressure model:

$$x \sim N\left[\xi, \sigma_w^2(1+c(\xi-\mu))^2\right],$$

$$\xi \sim N\left[\mu, \sigma^2\right]$$

with μ and σ estimated within age-group.

Age group	μ	σ^2	σ_w^2	c (x10 ³)	ln(likelihood)
Edinburgh					
≤39 (n=576)	80.67	50.01	43.83	10.38	-8115.4
40-44 (n=702)	81.81	56.37	44.84	7.34	-9952.4
45-49 (n=814)	82.77	57.13	46.47	10.27	-11588.6
50-54 (n=704)	83.06	68.28	44.51	11.77	-10025.1
55-59 (n=408)	84.10	62.30	45.93	7.89	-5817.9
≥60 (n=220)	82.53	68.94	46.56	21.84	<u>-3143.4</u>
					<u>-48642.8</u>
Budapest					
≤39 (n=445)	81.33	43.79	52.39	12.44	-6373.2
40-44 (n=539)	82.58	41.24	52.10	11.60	-7704.0
45-49 (n=828)	83.60	44.39	52.45	15.39	-11861.7
50-54 (n=869)	84.95	46.66	56.69	10.28	-12579.2
55-59 (n=574)	84.54	46.62	58.34	14.44	-8332.9
≥60 (n=609)	85.11	45.90	57.61	12.62	<u>-8826.4</u>
					<u>-55677.5</u>
Prague					
≤49 (n=1009)	83.11	32.77	51.49	14.19	-14314.4
50-54 (n=2315)	84.55	34.98	54.80	14.52	-33132.2
≥55 (n=687)	84.25	45.14	52.59	11.70	<u>-9851.8</u>
					<u>-57298.4</u>

Table 3.11: Parameter estimates for the diastolic blood pressure model:

$$x \sim N\left[\xi, \sigma_w^2(1+c(\xi-\mu))^2\right],$$

$$\xi \sim N\left[\mu, \sigma^2\right]$$

with all parameters estimated within age-group.

with age in the three towns. Therefore it is reasonable to assume that within-person variability is constant with age after adjusting for changes in the underlying level with age.

Finally, and briefly, it is useful to consider the model (3.4) including allowance for trends, but without the age stratification for the parameters μ and σ^2 . Fitting this gives the parameter estimates in table 3.12. The inclusion of the trend parameters produces a highly significant improvement in fit ($P < 0.0001$ in each town using a likelihood ratio test). The estimated mean trends given by θ are essentially the same as the mean observed trends. The variability of the trends, as found earlier, is estimated as zero in Budapest and so suggests a fixed and statistically significant downward trend in all individuals. This is a surprising result and is difficult to explain. In Edinburgh and Prague trends are variable between individuals and σ_β is estimated as 1.36 and 1.22mmHg/year, respectively, so that in these towns about 99% of the population would have underlying trends within the range ± 3 mmHg/year. By finding the difference in the estimates for σ_w^2 from this model and the model presented in table 3.9, the effect on the within-person variance of omitting the trends from the model can be assessed. This difference varies from 0.52mmHg² in Budapest to 2.70mmHg² in Prague and 3.17mmHg² in Edinburgh. Thus it is very small in each town and is comparable to the component of variance that arises from a practice of rounding blood pressures to the nearest 5mmHg.

	μ	σ^2	θ	σ_β^2	σ_w^2	$c (x10^3)$	$\ln(\text{likelihood})$
Edinburgh	82.40	61.27	0.15	1.86	42.07	10.62	-48652.5
Budapest	83.85	46.71	-0.56	0	54.48	12.81	-55676.5
Prague	84.31	37.42	0.33	1.51	50.89	13.33	-57326.9

Table 3.12: Parameter estimates for the diastolic blood pressure model with trends:

$$x \sim N[\xi + \beta t, \sigma_w^2(1 + c(\xi - \mu)^2)],$$

$$\beta \sim N[\theta, \sigma_\beta^2],$$

$$\xi \sim N[\mu, \sigma^2]$$

3.10 Summary of main results with discussion

This chapter has described within-person diastolic pressure variability in a cohort of middle-aged men who are essentially healthy in that they are free of diagnosed major disease during a period of four to five years. Blood pressure measurement in the trial from which the cohort is taken was casual and would reflect well measurement methods used by general practitioners in screening. Thus valuable information can be drawn from this study that can be used to aid the interpretation of diastolic blood pressure monitoring during screening.

The main result of the chapter is that observed diastolic pressures within an individual can be considered as normally distributed about some underlying average level, and that the standard deviation of this distribution is, conditional on the level, apparently constant across individuals. For an individual with an underlying level of 80mmHg,

this standard deviation is similar in the three towns studied being 6.5mmHg in Edinburgh, 6.9mmHg in Prague and 7.1mmHg in Budapest. These figures are similar to the findings of other studies. The differences between the three towns are not solely attributable to the different rounding practices observed but are probably small enough to be related to other facets of blood pressure determination: for example, the physical environment, the observer characteristics or the state of rest of the subject at examination. There was no evidence that the rate of increase in variability was different in the three towns giving within-person standard deviations of about 5mmHg at an underlying level of 60mmHg and rising to about 8.5mmHg at a level of 100mmHg. The dependence of the variability on age found by Wilson and Hebel (1988) was also observed in this study but can be explained by allowing for increasing diastolic pressure levels with age.

Although the blood pressure measurements used in this study were taken at approximately annual intervals, the model derived is likely to be applicable for measurements separated by other intervals of time. For long time periods, underlying trends found in this study are small (within ± 3 mmHg/year for nearly all individuals) and failure to recognize them in a model produces only a comparatively small change in the within-person variance. At the other extreme, the result of Wilson and Hebel (1988) that measurements seven days apart show negligible autocorrelation suggests that the model would be appropriate for weekly or monthly measurements. Thus, where screening is opportunistic (for example, whenever a subject visits his general practitioner), the combination of blood pressure measurements from visits separated by as little as a week or as much as three or four years should enable a reliable assessment of underlying pressure during that time using the model developed.

Important practical issues arise from this model. First, consider the distribution of diastolic blood pressures within an individual, $f(x|\xi)$. If the underlying level, ξ , is high, say 90mmHg, and so of particular interest because of its association with increased risk of cardiovascular morbidity, then the standard deviation of this distribution is about 7.5mmHg. Thus a single observed diastolic pressure will be a very poor measure of an individual's underlying level: with $\xi=90$ mmHg, 95% of observations would be expected between 75 and 105mmHg. Large changes between measurements over time should therefore be expected being attributable just to random variation. Even with ten measurements the standard error of their mean is 2.4mmHg and so a 95% confidence interval for an individual's underlying level will span a range of 10mmHg. Often, this would be large enough to hamper treatment decisions. Of course, in practice, the true underlying level is never known and so the exact standard deviation of the measured pressures is never known. However, it is clear that substantial numbers of measurements are required to give precise estimates of underlying level and, therefore, that there is advantage in accumulating this information whenever a subject visits his general practitioner.

A second important issue concerns the monitoring of diastolic pressure for change. Underlying trends are small and so will require extensive numbers of measurements to establish their existence given the large natural variability of diastolic pressure. In contrast, a strategy for checking the success of intervention to lower blood pressure would require accurate determination of level prior to treatment and then again when the effect is expected to be established. This may require substantial numbers of measurements at each time (possibly greater than ten) particularly when relatively small reductions of the order of 10mmHg or so are expected. Such a number of tests

would seem justified given the undesirability of long-term intervention which is costly and carries the possibility of adverse effects.

A third issue that arises from the model for blood pressure variability concerns what is meant by "hypertension". Traditionally, this is defined in terms of ranges for diastolic pressures though it is not always clear what is meant by a blood pressure: a single observation, the mean of several observations or the long-term underlying level. For instance, guidelines proposed by the Participants of the W.H.O./I.S.H. meeting (1986) for the treatment of mild hypertension used a definition of a long-term average between 90 and 104mmHg whereas the Medical Research Council's Mild Hypertension Trial (Medical Research Council Working Party on Mild to Moderate Hypertension, 1985) use a screening criteria based upon the mean of four measurements taken during two visits being between 90 and 110mmHg to determine entry into the trial. Clearly the variability of diastolic pressure means that the two definitions do not describe the same set of individuals within the population. For instance, the model given in table 3.9 for Edinburgh implies that 23% of the population would have a single observed diastolic pressure above 90mmHg but only 16% have an underlying pressure above 90mmHg. Such a difference in percentages has a clear implication for planning and costing public health strategies.

This definitional issue is also important in determining blood pressure levels at which intervention, particularly pharmacological, is appropriate: that is, when the benefits in terms of risk reduction are likely to outweigh the possible adverse effects of treatment. Clinical trials of antihypertensive treatments should aim to describe the patients randomized in terms of their underlying blood pressure levels as definitions in terms of

observed blood pressures will be affected by the source of the trial population. For instance, the distribution of underlying blood pressures in a trial population selected from a general population because their observed pressures fulfill some defined criteria will be quite different from a population selected from patients attending a hypertensive clinic on the basis of the same criteria. In order to achieve this definition, trial investigators need to know the distribution of underlying levels in the population sampled and then apply the results for within-person variability obtained in this study. Alternatively, more precise determination of the distribution of underlying levels of the subjects entered into the trial must be undertaken by repeated measurement after the decision to randomize but prior to commencing treatment. Although this will increase the cost of trials, it would aid the doctor who, faced with a patient's blood pressure measurements, is trying to relate these to complex trial entry criteria and so make the decision whether to treat that patient or not.

4. Blood pressure and risk of cardiovascular mortality

4.1 Introduction

The relationship between blood pressure level and risk of dying from cardiovascular causes has been well established by many observational studies. However there is presently interest in establishing more precisely the magnitude of this relationship. Simple though this may seem, it is in fact complicated by the variability within individuals in blood pressure. The strength of the relationship appears weaker if related to a single measurement on each member of a cohort than if related to their underlying pressure (that is, the long-term mean level). This reflects the “regression to the mean” phenomenon and its effect on estimates of the risk association has been termed “regression dilution” by MacMahon et al (1990).

In this chapter, the proportional hazards model (Cox, 1972) is used to quantify the relationship between risk of cardiovascular mortality and blood pressure. In the next section, risk is related to observed blood pressure and the model assumptions are assessed. Also considered are the effects of trends in blood pressure and the lability of blood pressure in improving predictive power. Then in section 4.3, the regression dilution effect is considered. First the methodological aspects of regression dilution are considered in the context of survival models. Then the magnitude of the actual risk relationship with underlying pressure is estimated.

Before proceeding to the next section and the development of survival models, it is useful to summarize the extent of mortality data available for the study cohort taken from the clofibrate trial. Throughout the chapter, deaths from cardiovascular causes

	Edinburgh	Budapest	Prague	All towns
No. in study cohort	3424	3864	4011	11299
No. dying from cardiovascular causes	152	159	197	508
% dying from cardiovascular causes	4.44	4.11	4.91	4.50

<u>Follow-up in individuals not dying from cardiovascular causes (years)</u>				
maximum	13.92	11.58	11.22	13.92
upper quartile	12.16	9.62	9.76	10.25
median	10.51	8.73	9.30	9.30
lower quartile	9.14	7.96	8.80	8.63
minimum	0.42	0.08	0.13	0.08
<u>Follow-up in individuals dying from cardiovascular causes (years)</u>				
maximum	11.99	10.20	10.97	11.99
upper quartile	9.24	7.75	8.15	8.18
median	6.90	6.15	6.68	6.59
lower quartile	4.25	3.52	4.35	3.99
minimum	0.37	0.25	0.32	0.25

Table 4.1: Cardiovascular mortality and length of follow-up by town and across all towns

will be modelled with deaths from other causes being considered as observations censored at the time of death. Cardiovascular causes can be summarized as ischaemic heart disease, stroke and other circulatory diseases (Committee of Principal Investigators, 1978 and 1980). Follow-up is defined as the time from visit V_0 (the fourth visit in the sequence for which blood pressures are available) until death, or until censoring at the final follow-up for the trial. Table 4.1 summarizes the mortality and

follow-up figures both by town and overall. Cardiovascular mortality rates are similar in the three towns, averaging 4.5% during a median follow-up of 9.3 years from visit V_0 .

Throughout this chapter, all survival models developed have other risk and confounding factors included (age, cholesterol level and smoking status at visit V_0 , together with clofibrate/placebo treatment group and town). However, as these are of secondary interest and their estimates vary little between models, the effects associated with these variables will only be reported for a “final” model presented at the end of the section 4.2. Because some individuals have missing data on some of these variables, the development excludes 311 individuals including 16 that died, giving a mortality rate in this group (5.1%) slightly higher than that in the cohort as a whole; any bias arising from this missing data is likely to be small. Note also that cholesterol, in particular, and smoking status are also variables measured “with error” and so could be subjected to a similar form of analysis as here for blood pressure.

4.2 Observed blood pressure level and risk

4.2.1 Model definition and assessment of assumptions

Denoting by $h(t|x,\mathbf{z})$, the hazard function for cardiovascular mortality at time t given an observed blood pressure x at time 0 and other covariates \mathbf{z} , then the proportional hazards model is defined by

$$h(t|x,\mathbf{z}) = h(t|x=0, \mathbf{z}=\mathbf{0}) \exp(\beta x + \boldsymbol{\gamma}^T \mathbf{z}) \quad (4.1)$$

Thus $\exp(\beta)$ is the relative increase in the hazard (or instantaneous risk) of dying from cardiovascular causes associated with a unit increase in observed blood pressure. There are two fundamental assumptions underlying this model: first, that of a log-linear relationship between each covariate and risk; second, that of proportionality of hazards throughout time: that is, for given values of the covariates (x_1, z_1) and (x_2, z_2) , then

$$\frac{h(t_a|x_1, z_1)}{h(t_a|x_2, z_2)} = \frac{h(t_b|x_1, z_1)}{h(t_b|x_2, z_2)}$$

for any points in time t_a and t_b .

The log-linear relationship between risk and level of blood pressure has been studied by MacMahon et al (1990) in their overview of nine major observational studies including 420,000 individuals of whom nearly 5,000 suffered fatal cardiovascular events. They showed strong evidence for linearity in the relationship between the logarithm of risk of coronary heart disease and blood pressure level and also suggested that linearity in the relationship between the logarithm of risk of stroke and blood pressure was a good approximation. Given that deaths from coronary heart disease dominate stroke deaths by a factor of 3:1 in the clofibrate trial, the assumption of log-linearity between risk of cardiovascular death, overall, and blood pressure seems justified. The shape of this relationship can also be assessed in the cohort from the clofibrate trial being studied here. Dividing the cohort into quartiles within each town based on each individual's mean diastolic blood pressures across the four visits V_{-3} to V_0 , then figure 4.1 shows crude mortality rates during the follow-up period by the average pressure in each quartile. Though based upon comparatively small numbers of events, this provides

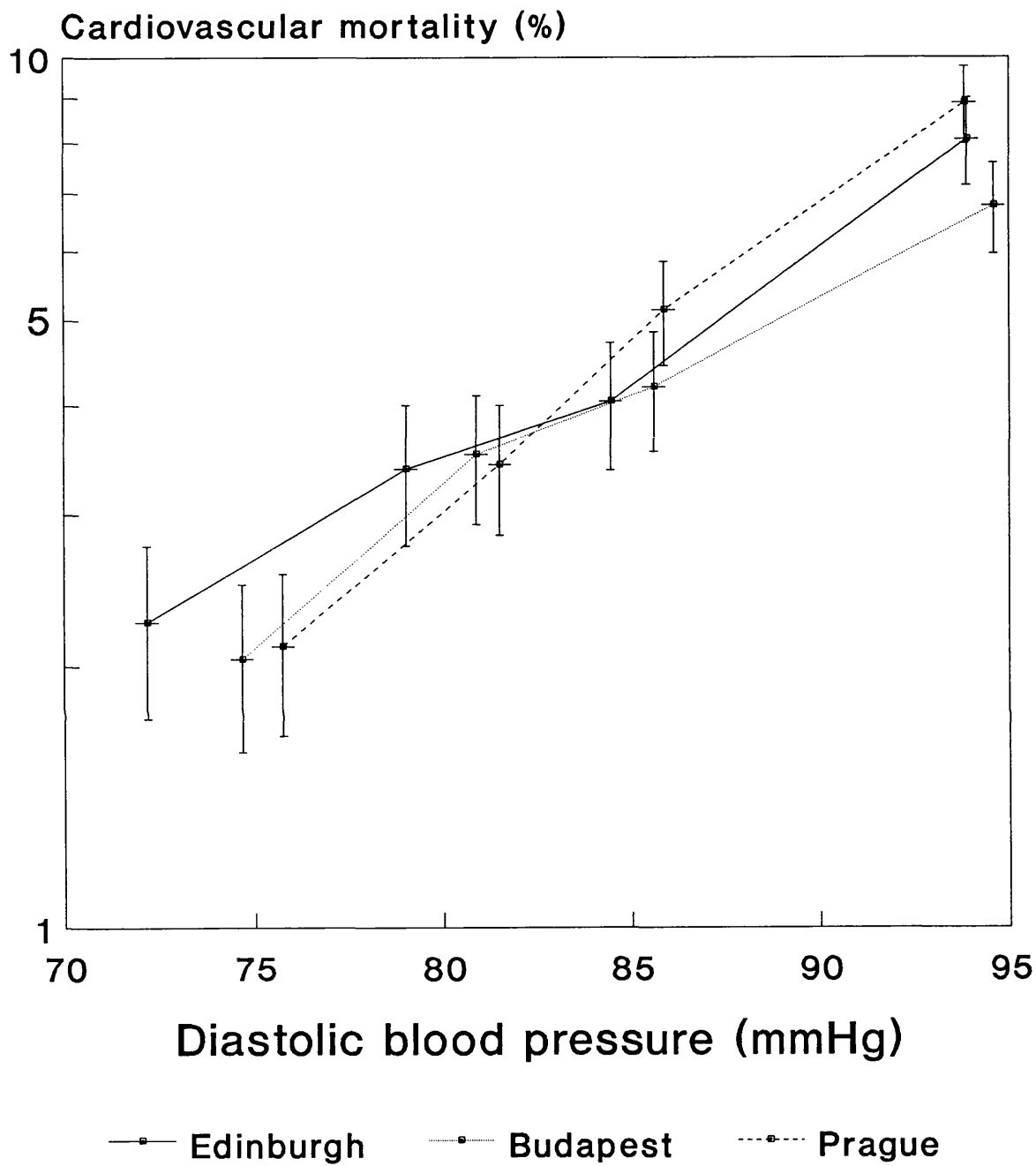


Figure 4.1: Crude cardiovascular mortality rates by quartile of diastolic pressure (mean of four measurements) within town (bars are \pm one standard error)

further evidence for the log-linear trend associating risk and blood pressure level. In addition, this was confirmed by including quadratic terms in diastolic pressure in the survival models: none reached any conventional level of statistical significance.

For assessing the assumption of proportional hazards the cohort can be subdivided into six groups on the basis of their mean diastolic blood pressure (over four visits). If the assumption is valid, then the hazards $h(t)$ within each group should be proportional to one another. Equivalently, the logarithm of the cumulative hazards in each group should differ by a constant amount throughout time. This leads to the so-called $\log(-\log)$ plots (Kalbfleisch and Prentice, 1980; chapter 4) where the survival function $S(t)$ is estimated using Kaplan-Meier methods within each group, so that $\log(-\log[S(t)])$ is the logarithm of the cumulative hazard. Figure 4.2 shows the plot of $-\log(-\log[S(t)])$ for the grouped diastolic pressures. For convenience, the plots are smoothed between the points marking the times of deaths: strictly, the estimated curves should be step functions showing decreases only at the times of these deaths. There are no clear departures from parallelism of the curves.

An alternative approach for assessing the assumption of proportional hazards is to use a statistical test suggested by Harrell (1986). This is based upon the evaluation of “partial residuals” (Schoenfeld, 1982) at each time at which a death occurred. Each residual is simply the difference between the observed diastolic pressure of the individual dying and the expected value of the pressure estimated from data on the set of individuals still being followed-up at that time point. Schoenfeld shows that these residuals are, under the assumption of proportional hazards, uncorrelated, but are correlated if this assumption fails. Harrell uses this fact to give a test of the hypothesis

$-\ln(\text{cumulative hazard})$

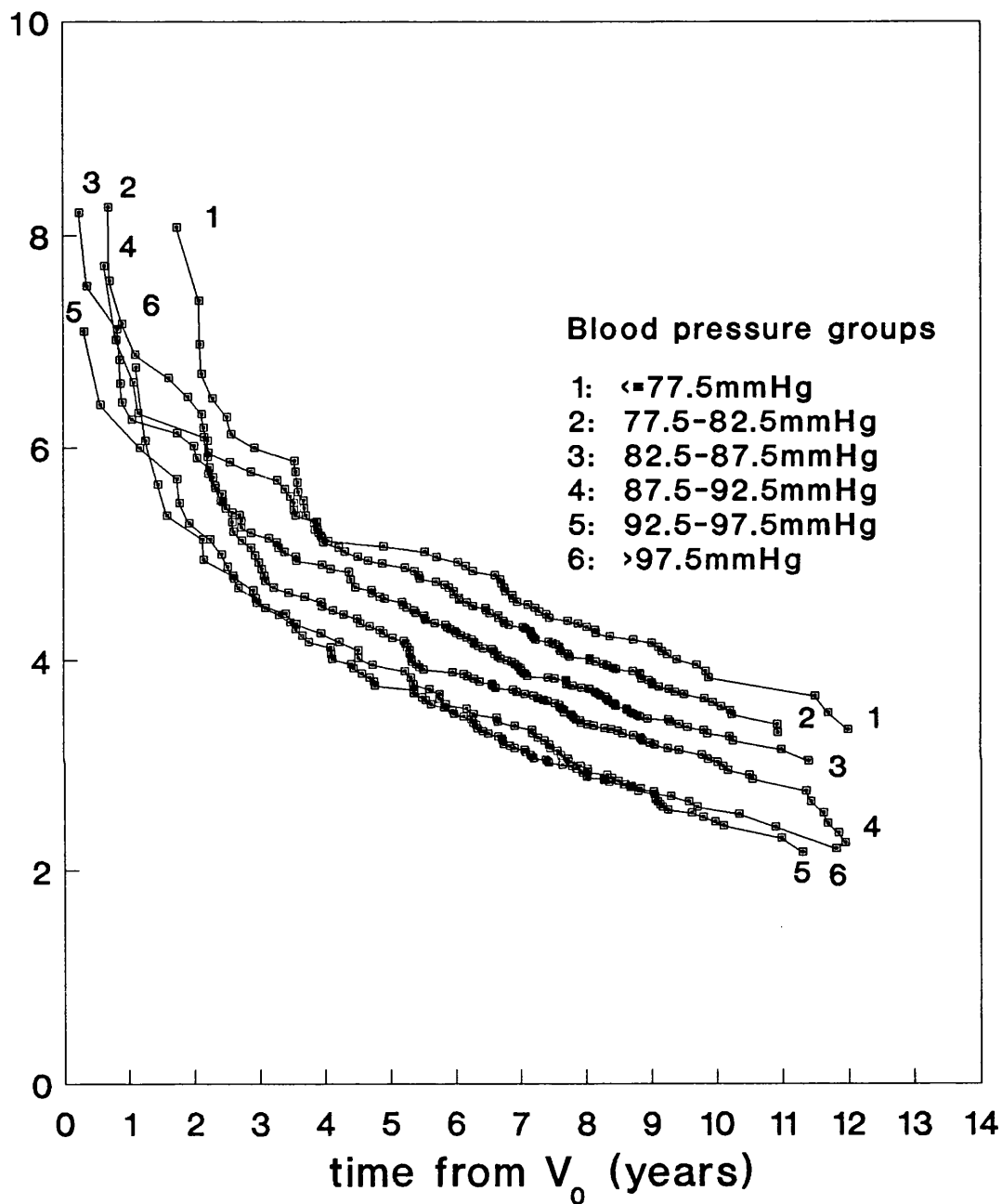


Figure 4.2: $-\log[-\log S(t)]$ plots by mean (of four observations) diastolic pressure group (points marked correspond to the times of deaths observed)

of zero correlation which is then a test of proportional hazards. Considering the mean of four diastolic pressures as the covariate of interest, using this test gives no significant evidence for departures from proportional hazards ($P=0.73$).

Overall the goodness of fit of a model can be assessed using generalized residual plots (Kalbfleisch and Prentice, 1980; chapter 4). The philosophy behind these is that the estimated survival probabilities at the time of observed death or censorship for each individual based on their observed covariate values form a (censored) sample from a uniform distribution. Equivalently, the corresponding generalized residuals for each individual, the estimated cumulative hazard based on their covariate values at the time of censorship or death from cardiovascular causes, form a (censored) sample from an exponential distribution with unit mean. Thus $-\log(\text{proportion of residuals} > r)$, where r is any real positive number, plotted against r should follow a line of unit gradient. Figure 4.3 shows the plot obtained for the model including the mean diastolic pressure: the fit is apparently good. Similar plots can be obtained for an analysis stratified by blood pressure group (figure 4.4). This permits assessment of the particular fit of blood pressure. Again the fit is good, the departures for larger values of r , which appear visually large, being less important statistically because they are based upon long follow-up times when few within the respective group are at risk.

Similar checks of the assumptions and goodness of fit were made for the other covariates in the model. For both age and cholesterol, the log-linear model seems reasonable. More generally, there was no evidence of departures from proportional hazards and the residual plots suggested a good fit.

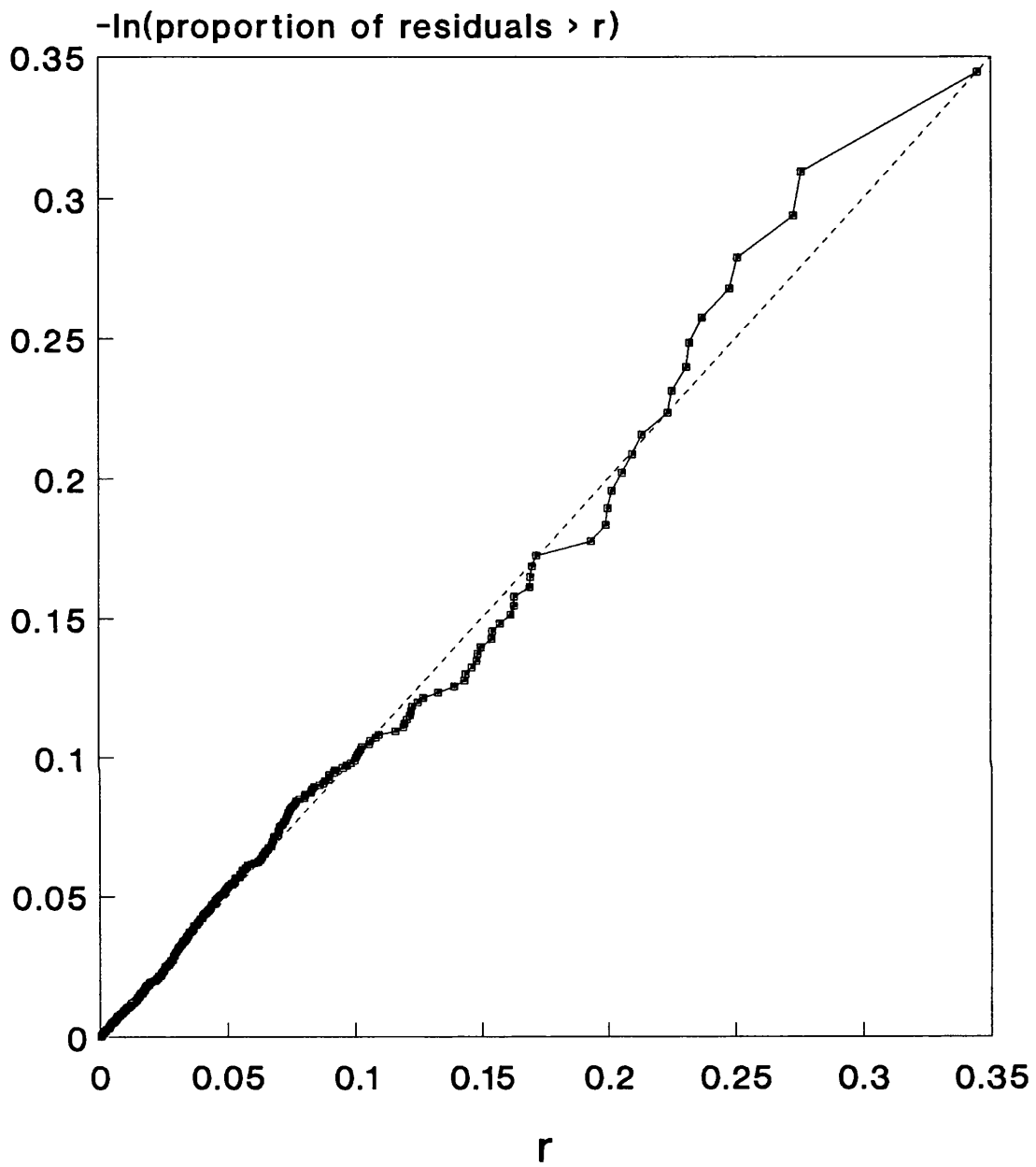


Figure 4.3: Residual plot for the proportional hazards model including mean diastolic pressure and other covariates

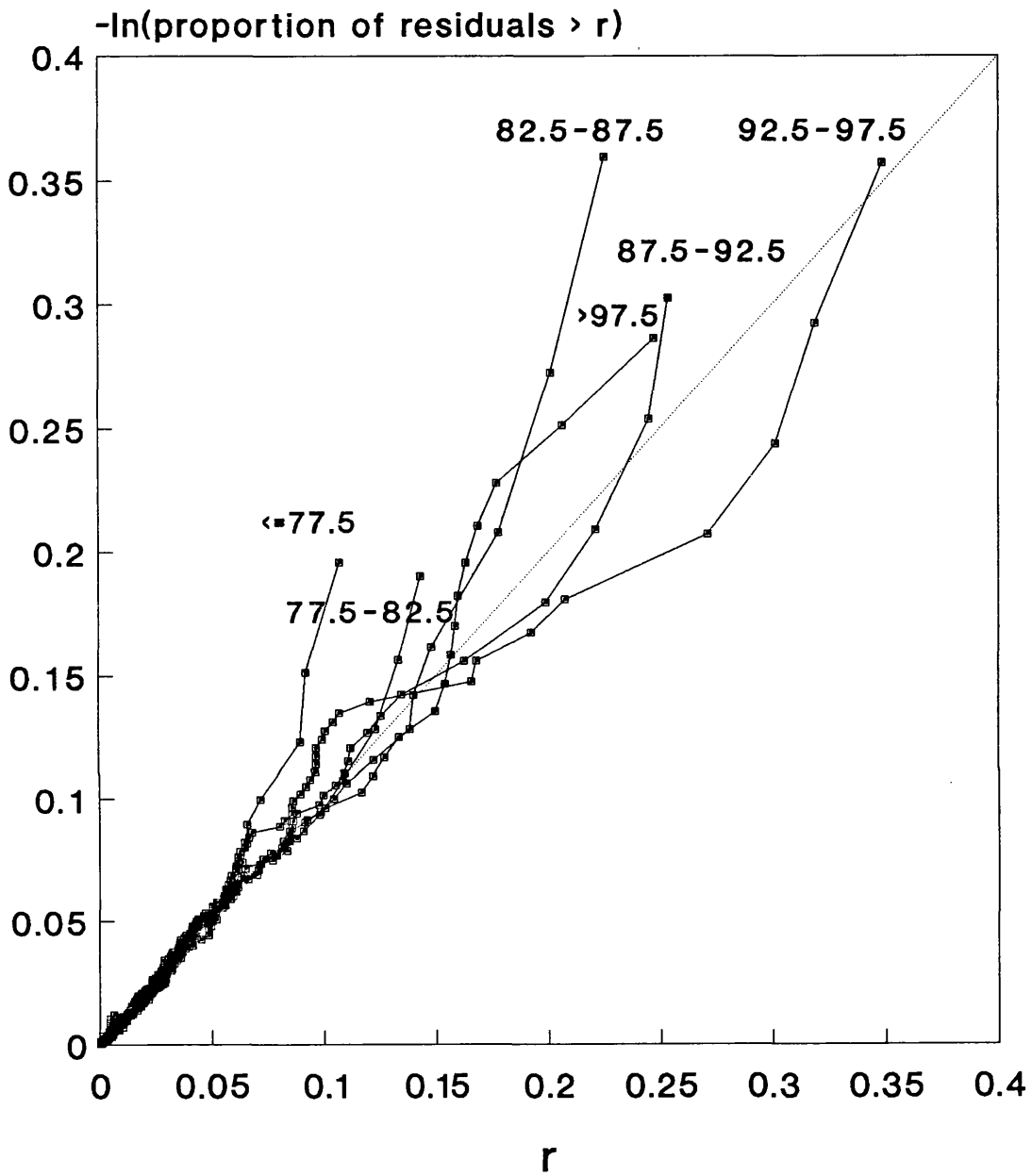


Figure 4.4: Residual plot for the proportional hazards model, stratified by mean diastolic pressure group

Diastolic b.p from visit	Risk coefficient, $\hat{\beta}$	Standard error of $\hat{\beta}$	-2 log likelihood at maximum
V_0	0.0335	0.0041	8549.53
V_{-1}	0.0330	0.0043	8553.76
V_{-2}	0.0320	0.0044	8559.90
V_{-3}	0.0325	0.0044	8559.99

Table 4.2: Risk coefficients for diastolic blood pressure measured at a single visit when included separately in a model also with age, cholesterol, smoking status, town and treatment group.

4.2.2 Estimates of risk association: diastolic blood pressure

The simplest model for predicting survival based on diastolic blood pressure level is to use just a single observed measurement at visit V_0 . Including this in a proportional hazards model indicates an increased risk of 3.40% associated with a 1mmHg difference in observed diastolic pressure (approximate 95% confidence interval: 2.57% to 4.25%). Thus a difference of 10mmHg in observed diastolic pressure measured at a single visit (a difference corresponding to approximately one standard deviation in the population sample studied), gives a 40% increase in relative risk.

Of course, the survival model could also be based on the single measurement recorded at visit V_{-3} , or at visit V_{-2} , or at visit V_{-1} . The estimated risk coefficients, $\hat{\beta}$, and their standard errors are given in table 4.2 for models using, in turn, each of these measurements. Very clear is the stability of the coefficients which for all practical

purposes are the same. Also shown is -2 times the log likelihood at the maximum. Although the models are not nested and so formal statistical comparison using a likelihood ratio test is not appropriate, these values do allow informal comparison suggesting that using the most recent measurement explains a little more variation than using earlier measurements.

The estimates obtained are remarkably similar to the logistic regression coefficients obtained for coronary heart disease mortality in white men aged 40-54 with no pre-existing heart disease reported by Kannel et al (1986) in their summary of results from several epidemiological studies. In particular they report coefficients for diastolic pressure measured at one visit of 0.0331 (standard error 0.0023) in the largest study, the Multiple Risk Factor Intervention Trial screening cohort of 221076 men, and of 0.0342 (standard error 0.0099) in the longer term but smaller Framingham study.

The likelihood values can be compared with the value from the saturated survival model where diastolic pressures from each of the four visits are included simultaneously. The estimates from this model are given in table 4.3. There is a highly significant improvement in statistical significance in moving from one to four diastolic pressures ($P < 0.0001$, likelihood ratio test). There are no statistically significant differences between the risk coefficients. Thus, although the coefficients vary considerably in relative terms (the estimate for visit V_0 is over 50% higher than that for visit V_{-2}), lack of power limits an ability to claim that the level of diastolic pressure closer to V_0 , the time origin used for predicting survival from, is more important for predicting future risk.

Diastolic b.p. from visit	Risk coefficient, $\hat{\beta}$	Standard error of $\hat{\beta}$	-2 log likelihood at maximum
V ₀	0.0161	0.0053	8521.25
V ₋₁	0.0135	0.0055	
V ₋₂	0.0102	0.0056	
V ₋₃	0.0122	0.0055	

Table 4.3: Risk coefficients for observed diastolic blood pressures when included simultaneously in a model also with age, cholesterol, smoking status, town and treatment group.

A further model can be developed by constraining the risk coefficients for diastolic pressures at visits V₋₃ to V₀ to be equal. This is equivalent to entering into the model just the mean of the four pressures as a predictor variable. In this case, the risk coefficient is 0.0522 with standard error 0.0054. This gives an increase in risk of 5.36% associated with a 1mmHg increase in the observed mean of four diastolic pressures measured approximately one year apart (95% confidence interval: 4.26% to 6.48%). More important, though, is the small change in the maximum value of the likelihood from the previous model ($-2 \log \text{likelihood} = 8521.77$, an increase of 0.52) which is not statistically significant. Thus all the information from such a sequence of four diastolic pressures would appear to be contained within their mean.

The important consequence of this result is that there is insufficient power obtainable in this dataset to determine whether any other function of the four observed blood pressures has greater predictive power than their mean. Also, no function will significantly improve the predictive power of a survival model including the mean.

This implies that neither trend nor lability (measured here by the residual variation about the fitted trend through the four measurements) improve predictive power. For completeness and to emphasize this point, in figures 4.5 and 4.6 are shown the crude mortality rates by quartile of observed trend (within individual) and observed residual standard deviation about the fitted trend, respectively, within each town: the lack of any relationships is clear. The result for the lack of predictive power of trends agrees well with findings from the Framingham study (Hofman et al, 1983). Considering trends over a longer period (12 years) with prediction over the following 14 years, they concluded that the trend did not add anything to prediction after the level at the end of the 12-year period was considered. In addition they assessed the predictive value of residual variability about the trend during the 12 year period and, as here, found no effect. Note that the lack of predictive power of residual variation might be expected given that a model describing within-person variability as homogeneous in the population (after allowing for level) was found, in the previous chapter, to fit well. Thus variation in this characteristic is, apart from sampling variation, just describing differences in underlying mean diastolic pressure.

Kannel, Sorlie and Gordon (1980) also considered as possible predictors of risk, the maximum, minimum and mean observed pressures within a visit. They found that they were each equally good predictors of risk and also that within-visit variability had no predictive power. Although this study has no similar within-visit information, it is useful to consider whether either the maximum or the minimum diastolic pressure across the four annual visits performs as well as the mean of the four. Table 4.4 shows the estimates of the risk coefficients obtained together with their standard errors and the model likelihood values. An informal comparison of the likelihood values for each

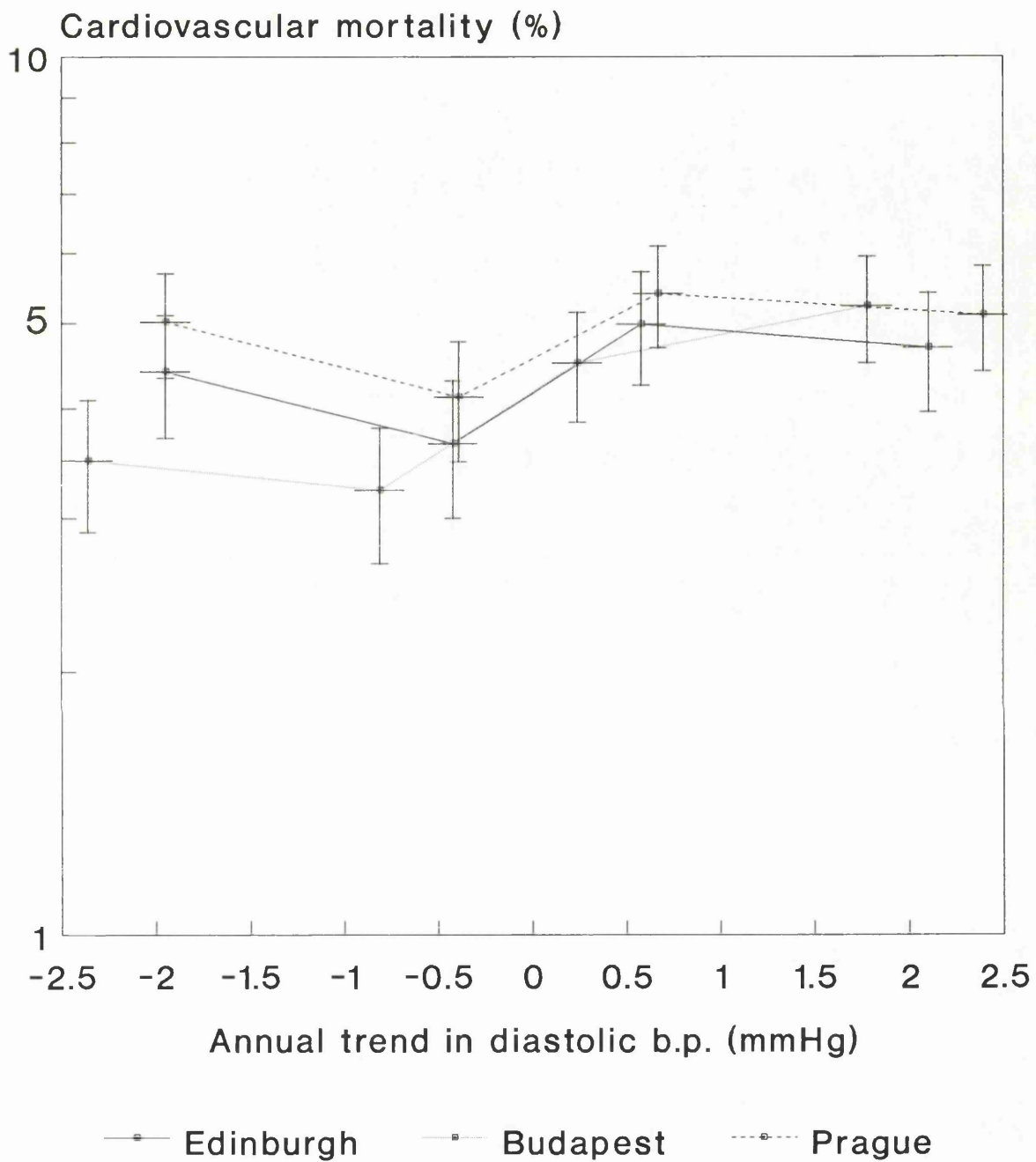


Figure 4.5: Crude cardiovascular mortality rates by quartile of annual trend in diastolic pressure, within town (bars are \pm one standard error)

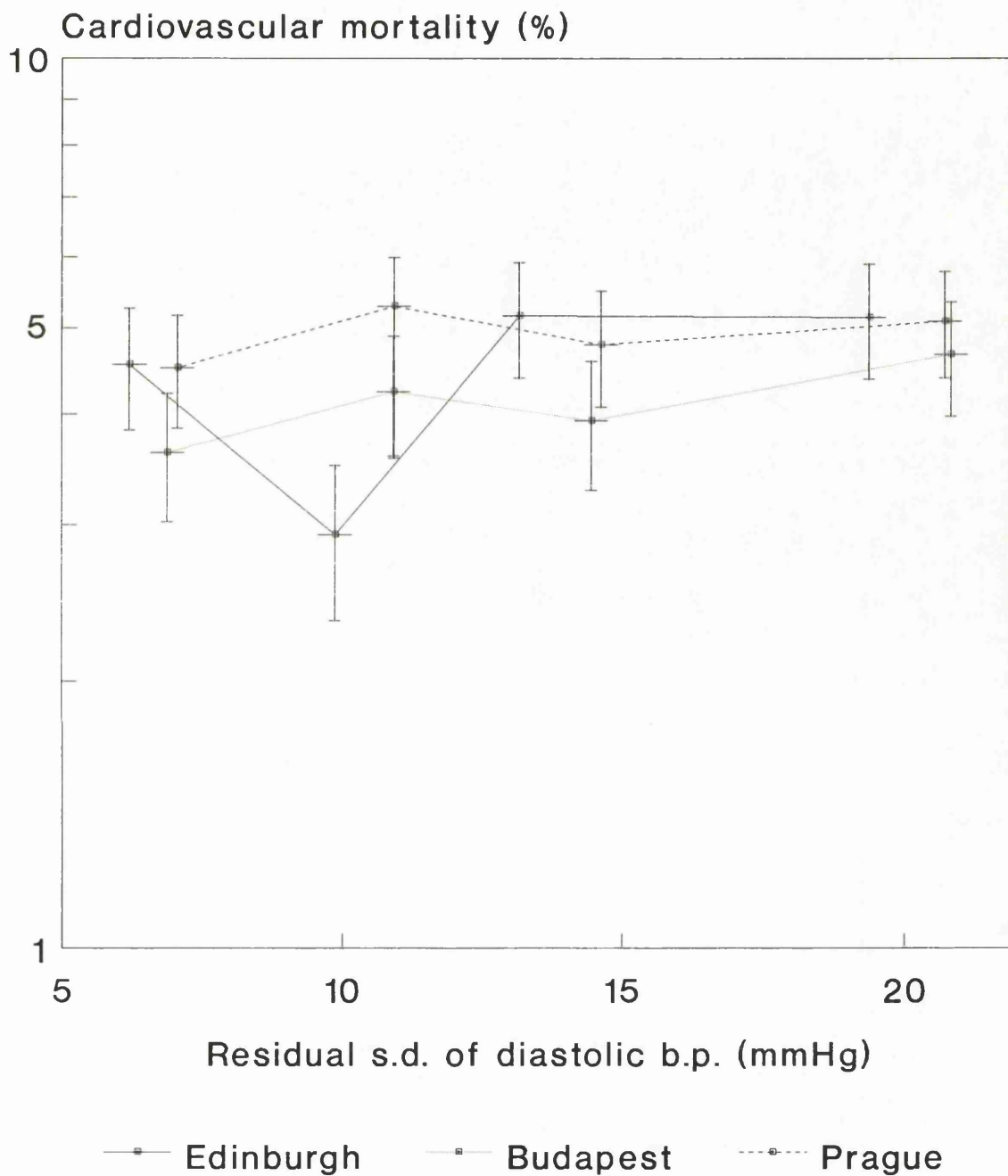


Figure 4.6: Crude cardiovascular mortality rates by quartile of residual standard deviation of diastolic pressure (after fitting mean and trend to a sequence of four annual observations), within town (bars are \pm one standard error)

Function of diastolic pressures at visits V ₋₃ to V ₀	Risk coefficient, $\hat{\beta}$	Standard error of $\hat{\beta}$	-2 log likelihood at maximum
mean	0.0522	0.0054	8521.77
maximum	0.0408	0.0046	8536.99
minimum	0.0434	0.0052	8544.82

Table 4.4: Risk coefficients for functions of the four observed diastolic blood pressures when included separately in a model also with age, cholesterol, smoking status, town and treatment group.

model suggests that the mean is a better explanatory variable than either the minimum or the maximum. Of course, both the maximum and minimum reflect information contained in four pressure measurements. Therefore, just as the risk coefficient associated with the mean of four measurements is larger than the coefficient associated with a single measurement due to its improved precision in estimating underlying level, so are the coefficients associated with the maximum and minimum. Similarly, that their coefficients are smaller than that associated with the observed mean is a reflection of their comparatively lower precision in estimating the underlying level.

Before presenting the full models developed, two further issues are worth considering. First, as the study cohort is taken from three towns, there is the opportunity to consider whether the association between diastolic pressure level and risk varies between towns. Fitting a model with separate terms for diastolic pressure within each

Town	Risk coefficient, $\hat{\beta}$, for mean of 4 dbp's	Standard error of $\hat{\beta}$	-2 log likelihood at maximum
Edinburgh	0.0504	0.0090	8516.62
Budapest	0.0373	0.0098	
Prague	0.0670	0.0089	

Table 4.5: Risk coefficients for the mean of four observed diastolic blood pressures, by town, when included simultaneously in a model also with age, cholesterol, smoking status, town and treatment group.

of the three towns gives the results presented in table 4.5. The estimated effect for Prague is about twice that for Budapest, with Edinburgh intermediate. The difference in -2 times the log likelihood between this model and the model with a common risk coefficient for the observed mean diastolic pressure is 5.15 which is of marginal statistical significance ($P=0.08$). This result contrasts with the overview of MacMahon et al (1990) which found no evidence of heterogeneity between effects observed in several epidemiological studies carried out in different geographical locations. Thus caution should be used before interpreting this finding as evidence for between-town variability in the risk relationship. In addition, it is possible for the association of risk with underlying mean pressure to be the same in each town, but for the relationship with observed pressure to be different between towns because of regression dilution: the extent of induced differences is dependent on the relative magnitudes of within- and between-person variance components which the models developed in chapter 3 suggest do vary markedly between towns. This issue is returned to later when the effects of regression dilution are considered more fully. Despite this, if the variability in effects

between towns is real, then applying these coefficients to differences in the mean of four observed pressures in each town of one standard deviation in the respective town's population (given in table 3.5) produces very varied and practically important differences in the associated risks: 54%, 34% and 62% increase in risk per standard deviation in Edinburgh, Budapest and Prague, respectively. Note also that for predicting risk in an individual from observed diastolic pressures, as distinct from understanding risk relationships, these differences imply that the appropriate town's model should be used.

The second issue is one of potential confounding that might arise from the use of trial data. Earlier, the influence of clofibrate in producing a step decline in blood pressure level from visit V₄ to visit V₃ was noted. Therefore it is relevant to assess whether the lower blood pressures found in the clofibrate group produce a different risk relationship from that found for similar levels in the placebo groups. This can be assessed by including an interaction term of diastolic pressure and treatment with clofibrate. For the interaction of treatment with the mean of the four observed pressures, the test does not achieve conventional statistical significance ($P=0.38$). Thus, it seems that the lower blood pressures achieved in the clofibrate group lead to the same risk relationship with cardiovascular mortality as similar pressures would in individuals not taking clofibrate, though it is important to recognize the limited power of this study to detect an interaction.

Thus this section has established that given the sequence of observed diastolic blood pressure measurements, after the mean level has been included in a survival model for predicting risk of dying from cardiovascular causes, no other characteristic of the

sequence significantly improves prediction. If the mean diastolic pressure is based upon fewer than four observations, then the estimated risk coefficients for diastolic pressure can be replaced by values of 0.0335 (giving an increase in risk of 3.40% per mmHg) if based on one diastolic pressure measurement, 0.0436 (4.46% per mmHg) if based on two, or 0.0485 (4.97% per mmHg) if based on three. This shows very clearly the importance of knowing how many pressure measurements the mean is based upon when using a survival model for future predictions. Note though that this is a result of reducing the within-person component of variance as the number of measurements on which the mean is based is increased. This illustrates well the issue of regression dilution: the risk coefficients are smaller the less precisely the observed pressures reflect the underlying level. Section 4.3 takes up this issue and relates risk to an individual's underlying average diastolic pressure rather than to an observed diastolic pressure.

4.2.3: Estimates of risk association: other factors

Coefficients relating to other variables included in the model with the mean of four diastolic blood pressures are presented in table 4.6. Requiring particularly careful interpretation are the associations between risk of cardiovascular mortality and both cholesterol and treatment group. The cholesterol distribution of subjects entering the trial is very peculiar including all subjects with a single measurement in the upper third of their town's distribution (groups I and II) and half of those in the lower third (group III). Cholesterol, like blood pressure, varies considerably within an individual, so treatment group as well as cholesterol level at visit V_0 are both markers of an individual's underlying average level. This is reflected by the non-zero risk coefficient

Variable	Risk coefficient, $\hat{\beta}$	Standard error of $\hat{\beta}$	Increase in risk*
Mean dbp from visits V_{-3} to V_0 (mmHg)	0.0525	0.0054	5.39% per mmHg
Cholesterol observed at V_0 (mg/dl)	0.0041	0.0012	0.413% per mg/dl
Age at V_0 (years)	0.1006	0.0082	10.6% per year
Smoking status:			
never smoked	-	-	0%
ex or current cigar/ tobacco only	0.2672	0.2579	30.6%
ex cigarette smoker	0.0443	0.1418	4.5%
1-19 cigs/day	0.5926	0.1389	80.9%
20 cigs/day	0.7795	0.1506	118.0%
>20 cigs/day	0.8625	0.1511	136.9%
Town:Edinburgh	-	-	0%
Budapest	-0.1385	0.1269	-12.9%
Prague	-0.1424	0.1245	-13.3%
Treatment group:			
I: high chol./ clofibrate	0.2459	0.1030	+27.9%
II: high chol./ placebo	-	-	0%
III: low chol./ placebo	-0.3233	0.1459	-27.6%

* for continuous variables increase in risk per unit on the scale given; for binary variables (ie smoking categories, town and treatment, increase in risk relative to that in the base group specified).

Table 4.6: Risk coefficient and the associated estimates of relative risk for all variables included in the model with mean diastolic blood pressure over the four visits, V_{-3} to V_0 .

for group III subjects (low cholesterol at entry) compared with group II subjects (high cholesterol at entry) despite the term for cholesterol level already included in the model and the fact that both groups received the placebo. Thus it is not possible to generalize the risk coefficient found for cholesterol in this study to other populations nor is it appropriate to compare it with the findings of other observational studies.

Recall also that the drug clofibrate is used to lower cholesterol and that the cholesterol levels at visit V_0 for subjects in group I reflect about four years of treatment with it. Thus the risk coefficients for cholesterol and treatment group should be considered as a pair and their interpretation limited to describing outcome for what is a very specific cohort: selected on the basis of their cholesterol level at one visit, treated with either clofibrate or placebo for four or more years and then having a cholesterol measurement at visit V_0 . For instance, consider two hypothetical subjects from Edinburgh, one from each of treatment groups I and II and having observed cholesterol levels at visit V_0 equal to the average in their group: that is (from table 2.3) 225mg/dl for the subject in group I taking clofibrate and 252mg/dl for the subject in group II taking the placebo. Then the treatment group/cholesterol component of their risk scores are 1.1684 (that is, $225 \times 0.0041 + 0.2459$) and 1.0332 (that is, $252 \times 0.0041 + 0$), respectively. Thus the subject in the clofibrate-treated group has a risk increased by a factor of $\exp(1.1684 - 1.0332) = 1.14$ compared with the subject in group II who received the placebo. Given that the difference in their cholesterol levels at visit V_0 might be attributed to clofibrate treatment (because the two groups are derived from the same population using random allocation), then this increased risk might also be attributed to clofibrate. This is a little higher than the crude relative risk in this study's cohort (group I: 217/3808 (5.70%) died; group II: 198/3788 (5.22%) died; relative risk=1.09)

and that reported for all randomised patients (group I: 404/5331 (7.58%) died; group II: 374/5296 (7.06%) died; relative risk=1.07; Committee of Principal Investigators, 1978 and 1984).

Some comment is also useful on other covariates included in the model. The increase in risk with age is, of course, not surprising and its magnitude represents a doubling of risk about every seven years. This is very comparable to the findings of other studies: risk coefficients for coronary heart disease from logistic regression models of 0.1174 (standard error 0.0298) for middle-aged Framingham men and 0.1091 (standard error 0.0038) for MRFIT white screenees are reported by Kannel et al (1986), and for more elderly European subjects in the EWPHE trial, a coefficient of 0.1017 for cardiovascular mortality from a proportional hazards model (Amery et al, 1986). For smoking status, the well-known increase in risk for current smokers compared with subjects who have never smoked is apparent. Again the risk relationship compares well with the findings for MRFIT screenees and Framingham men: estimated coefficients of 0.87 and 0.92, respectively, for men smoking about 20 cigarettes per day. The ex-cigarette smokers have negligible increase in risk compared with those who have never smoked though many of the ex-smokers in the trial had been non-smokers for a substantial number of years. For cigar/tobacco smokers, the observed risk is elevated though the small number of subjects in this category means that it is not statistically significant in comparison with those who have never smoked and, as found in the British Regional Heart Study (Cook et al, 1986), is well below the increase in risk associated with cigarette smoking. Finally, the two East-European towns, Budapest and Prague show very similar but lower observed risks than Edinburgh (though not significantly so).

4.3 Underlying blood pressure level and risk: correcting for regression dilution

4.3.1 Background to regression dilution

The problem of regression dilution has been understood for many years though not by that name: it is more often termed the problem of “errors in variables”. Standard text books (for example: Snedcor and Cochran, 1980, p171; Draper and Smith, 1980, p122) consider the problem in the context of simple linear regression when the predictor variable is measured with error. In this case a correction factor to the estimated regression coefficient is obtainable which, under certain conditions, removes the bias induced by the variable measured with error. It is useful to restate the problem as it forms the basis for deriving estimates of the magnitude of risk associated with underlying blood pressure in the context of survival models.

Consider the simple linear regression situation with dependent variable Y and predictor variable, ξ , where

$$Y = \alpha + \beta^* \xi + \epsilon$$

with $E(\epsilon)=0$ and $V(\xi)=\sigma_{\xi}^2$ in the population sampled. In practice, ξ cannot be measured directly and instead X is obtained where

$$X = \xi + \delta$$

with $E(\delta)=0$ and $V(\delta)=\sigma_{\delta}^2$. Then the regression model that should be fitted is

$$Y = \alpha + \beta^* x + \epsilon^*$$

where $\epsilon^* = \epsilon - \beta^* \delta$ and so the error term also depends on β^* . However, if the random variables, δ and ϵ , are uncorrelated and normally distributed, and are also uncorrelated with ξ , then the usual least squares estimator, $\hat{\beta}$, based upon observed values of Y and X is biased but is directly related to β^* :

$$E(\hat{\beta}) = \beta^* \left[1 + \frac{\sigma_\delta^2}{\sigma_\xi^2} \right]^{-1} \quad (4.2)$$

Thus, under these assumptions, the estimated regression coefficient, $\hat{\beta}$, based on the observed values needs to be increased by a factor of $(1 + \sigma_\delta^2 / \sigma_\xi^2)$ in order to provide an unbiased estimate of β^* .

In moving away from simple linear regression, Carroll (1989) notes that such simple correction factors are generally not obtainable and the problem becomes dependent on the true regression coefficients. In the next section, the dependence of the correction factor on the true coefficient is considered further in the context of the proportional hazards model for survival data.

4.3.2 Regression dilution in the proportional hazards model: background

Prentice and co-workers have studied the problem of covariate measurement error in the context of proportional hazards models (Prentice, 1982; Pepe et al, 1989). Their approach has been to assume the proportional hazards model for the underlying covariate, ξ , and then to determine the form of the induced hazard model for the covariate, X, which represents ξ measured with error. They show that the induced hazard function cannot be written in the form of (4.1), that is a base hazard function

dependent on time multiplied by a risk function dependent on the covariate but not on time. Thus, the usual estimation methods cannot be applied. However, they do obtain some general results for the form of the induced hazard ratio $h(t|x)/h_0(t)$, where $h_0(t)=h(t|x=0)$ at any time t , from which some conclusions can be drawn. In particular, these include that the risk coefficient estimated using the covariate x , $\hat{\beta}$, is related to the coefficient associated with the underlying covariate, β^* , by the same relationship, approximately, as derived for simple linear regression (given by equation (4.2)) when either the event rate is low or the distribution of ξ conditional on x is “concentrated”. However, the sensitivity of these results to departures from “low” rates or less “concentrated” distributions was not examined. This is investigated in the next sub-sections using an alternative approach: the expected value of the coefficient $\hat{\beta}$ is found by solving the usual equations for maximum likelihood based, incorrectly, on the observed covariate, x and relating this to the risk coefficient, β^* , associated with the underlying covariate, ξ . In section 4.3.3, the situation when there is no censorship is considered and then the results are extended, in section 4.3.4, to assess also the impact of censorship.

4.3.3 Regression dilution in the proportional hazards model with no censorship

Standard texts on survival analysis (for example: Cox and Oakes, 1984, chapter 7; Kalbfleisch and Prentice, 1980, chapter 4) show that the maximum likelihood estimate, $\hat{\beta}$, for the risk coefficient associated with the covariate x in the proportional hazards model given by (4.1) is the solution to the equation:

$$\sum_{i \in D} \left[x_i - \frac{\sum_{k \in D_i} x_k \exp(\hat{\beta} x_k)}{\sum_{k \in D_i} \exp(\hat{\beta} x_k)} \right] = 0 \quad (4.3)$$

where the set D (indexed by i) is the set of patients with known times of death (that is, uncensored observations), and the set D_i (indexed by k) is the set of patients still being followed up at the time immediately prior to the time of death of patient i . Interest here is in finding the expected value of $\hat{\beta}$ when the covariate, x , corresponds to a predictor variable, ξ , measured with error. To study this, a similar situation to that described in section 4.3.1 for simple linear regression is considered where, for convenience, the scale of measurement is transformed:

$$\gamma = \frac{\xi - \mu}{\sigma_\xi}$$

so that $\gamma \sim N(0, 1)$, and hence $x \sim N(\gamma, \sigma_\delta^2 / \sigma_\xi^2)$. The variance ratio, $\sigma_\delta^2 / \sigma_\xi^2$, will be denoted by λ .

The following analysis is asymptotic: the situation considered is where the sample size is very large so that the estimate, $\hat{\beta}$, arising from (4.3) equals its expected value which will be denoted by β , and the summations in (4.3) can be replaced by appropriate expectations. With no censoring, so that all subjects are followed to death, this gives

$$E_x(x) - E_t \left[\frac{E_{x|T \geq t} [x \exp(\beta x)]}{E_{x|T \geq t} [\exp(\beta x)]} \right] = 0. \quad (4.4)$$

To evaluate this, note first that $E(x)=0$ since $E(\gamma)=0$ so that the left hand expectation disappears. To evaluate the expectations within the bracket requires the probability density function for the covariate, x , conditional upon a survival time

greater than t . Using the notation $f(x)$ to denote the density function for any random variable, R , then

$$f(x|T \geq t) = \int_{-\infty}^{\infty} f(x, \gamma|T \geq t) d\gamma.$$

Now, if $S(t)$ is the usual survivor function (giving the probability that a subject's survival time is greater than t), then

$$\begin{aligned} f(x, \gamma|T \geq t) &= \frac{S(t|x, \gamma) f(x, \gamma)}{S(t)} \\ &= \frac{S(t|\gamma) f(x, \gamma)}{S(t)} \end{aligned}$$

if the assumption is made that survival is determined only by the underlying covariate, γ , and hence that knowledge of x adds no information about survival when γ is known.

Therefore

$$f(x, \gamma|T \geq t) = \frac{S(t|\gamma) f(x|\gamma) f(\gamma)}{\int_{-\infty}^{\infty} S(t|\gamma) f(\gamma) d\gamma},$$

and so

$$f(x|T \geq t) = \frac{\int_{-\infty}^{\infty} S(t|\gamma) f(x|\gamma) f(\gamma) d\gamma}{\int_{-\infty}^{\infty} S(t|\gamma) f(\gamma) d\gamma}. \quad (4.5)$$

Using this density function, then

$$E_{x|T \geq t} \exp(\beta x) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(\beta x) S(t|\gamma) f(x|\gamma) f(\gamma) d\gamma dx}{\int_{-\infty}^{\infty} S(t|\gamma) f(\gamma) d\gamma}.$$

Carrying out the integration with respect to x in the numerator gives

$$E_{x|T \geq t} \exp(\beta x) = \frac{\int_{-\infty}^{\infty} \exp(\beta \gamma + \frac{1}{2} \beta^2 \lambda) S(t|\gamma) f(\gamma) d\gamma}{\int_{-\infty}^{\infty} S(t|\gamma) f(\gamma) d\gamma}. \quad (4.6)$$

Similarly,

$$E_{x|T \geq t} x \exp(\beta x) = \frac{\int_{-\infty}^{\infty} (\gamma + \beta \lambda) \exp(\beta \gamma + \frac{1}{2} \beta^2 \lambda) S(t|\gamma) f(\gamma) d\gamma}{\int_{-\infty}^{\infty} S(t|\gamma) f(\gamma) d\gamma}. \quad (4.7)$$

Substituting (4.6) and (4.7) into (4.4) gives

$$E_t \left[\frac{\int_{-\infty}^{\infty} (\gamma + \beta \lambda) \exp(\beta \gamma + \frac{1}{2} \beta^2 \lambda) S(t|\gamma) f(\gamma) d\gamma}{\int_{-\infty}^{\infty} \exp(\beta \gamma + \frac{1}{2} \beta^2 \lambda) S(t|\gamma) f(\gamma) d\gamma} \right] = 0$$

and so

$$E_t \left[\frac{\int_{-\infty}^{\infty} \gamma \exp(\beta \gamma) S(t|\gamma) f(\gamma) d\gamma}{\int_{-\infty}^{\infty} \exp(\beta \gamma) S(t|\gamma) f(\gamma) d\gamma} + \beta \lambda \right] = 0. \quad (4.8)$$

To evaluate this requires the probability density function for t . This is the marginal density function:

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} f(t, \gamma) d\gamma \\ &= \int_{-\infty}^{\infty} f(t|\gamma) f(\gamma) d\gamma. \end{aligned}$$

But $f(t|\gamma) = -\frac{d}{dt} S(t|\gamma)$

where for the log-linear proportional hazards model

$$S(t|\gamma) = \exp\left[-\exp(\beta^*\gamma) \int_0^t h_0(u) du\right]. \quad (4.9)$$

Therefore

$$f(t|\gamma) = h_0(t) \exp(\beta^*\gamma) S(t|\gamma)$$

and hence

$$f(t) = \int_{-\infty}^{\infty} h_0(t) \exp(\beta^*\gamma) S(t|\gamma) f(\gamma) d\gamma. \quad (4.10)$$

Using (4.10), the expectation in (4.8) can be evaluated giving

$$\int_0^{\infty} \left[\frac{\int_{-\infty}^{\infty} \gamma \exp(\beta\gamma) S(t|\gamma) f(\gamma) d\gamma \int_{-\infty}^{\infty} h_0(t) \exp(\beta^*\gamma) S(t|\gamma) f(\gamma) d\gamma}{\int_{-\infty}^{\infty} \exp(\beta\gamma) S(t|\gamma) f(\gamma) d\gamma} \right] dt + \beta\lambda = 0. \quad (4.11)$$

An important result can be obtained by making the change of variables:

$$v = H_0(t) = \int_0^t h_0(u) du$$

so that $dv = h_0(t)dt$, with $v=0$ when $t=0$ and, if the reasonable assumption is made that $S(t) \rightarrow 0$ as $t \rightarrow \infty$ (that is, everyone dies eventually), $v \rightarrow \infty$ as $t \rightarrow \infty$. Then with expansion of $S(t|\gamma)$ using (4.9), equation (4.11) becomes

$$\int_0^{\infty} \left[\frac{\int_{-\infty}^{\infty} \gamma \exp(\beta\gamma) \exp[-v \exp(\beta^*\gamma)] f(\gamma) d\gamma \int_{-\infty}^{\infty} \exp(\beta^*\gamma) \exp[-v \exp(\beta^*\gamma)] f(\gamma) d\gamma}{\int_{-\infty}^{\infty} \exp(\beta\gamma) \exp[-v \exp(\beta^*\gamma)] f(\gamma) d\gamma} \right] dv + \beta\lambda = 0. \quad (4.12)$$

Thus the relationship between β and β^* does not depend upon the form of the underlying hazard function. This result requires no assumptions about the distributional forms for γ and δ , beyond their mutual independence. In general, this equation requires a numerical rather than an analytical solution to determine the relationship between β , the expected risk coefficient obtained using the observed covariate values and β^* , the risk coefficient related to the underlying covariate, γ .

Taking $f(\gamma)$ as the normal density function with mean zero and variance one, then given β^* , equation (4.12) can be solved numerically for β for different values of λ : here values between 0 and 2.5 in steps of 0.1 are considered. Figure 4.7 shows the relationship between the correction factor β^*/β and λ for a variety of values of β^* from 0 to 2 (corresponding to a wide range of risk increases from 0 to 639% for a standard deviation change in the underlying covariate value in the population). Also shown is the correction factor found for simple linear regression: $\beta^*/\beta=1+\lambda$. Very clear is the dependence of the correction factor on the underlying risk coefficient, β^* . For small values of β^* the correction needed is close to $1+\lambda$ for all values of λ . This agrees with the conclusion of Prentice (1982) that this is an appropriate factor to use if the event rate is low. However, as β^* increases the correction required also increases markedly. Even for values of λ and β^* which might be more relevant in practice (say β^* up to 1 and λ up to 1), the correction factor needed to estimate the underlying risk coefficient can be substantially more than $1+\lambda$. For example, with $\beta^*=1$ and $\lambda=1$, $1+\lambda=2$ whereas the required factor is 2.644.

Another feature shown in figure 4.7 is the approximate linear relationship between the correction factor, β^*/β , and the variance ratio, λ . This is particularly good for smaller

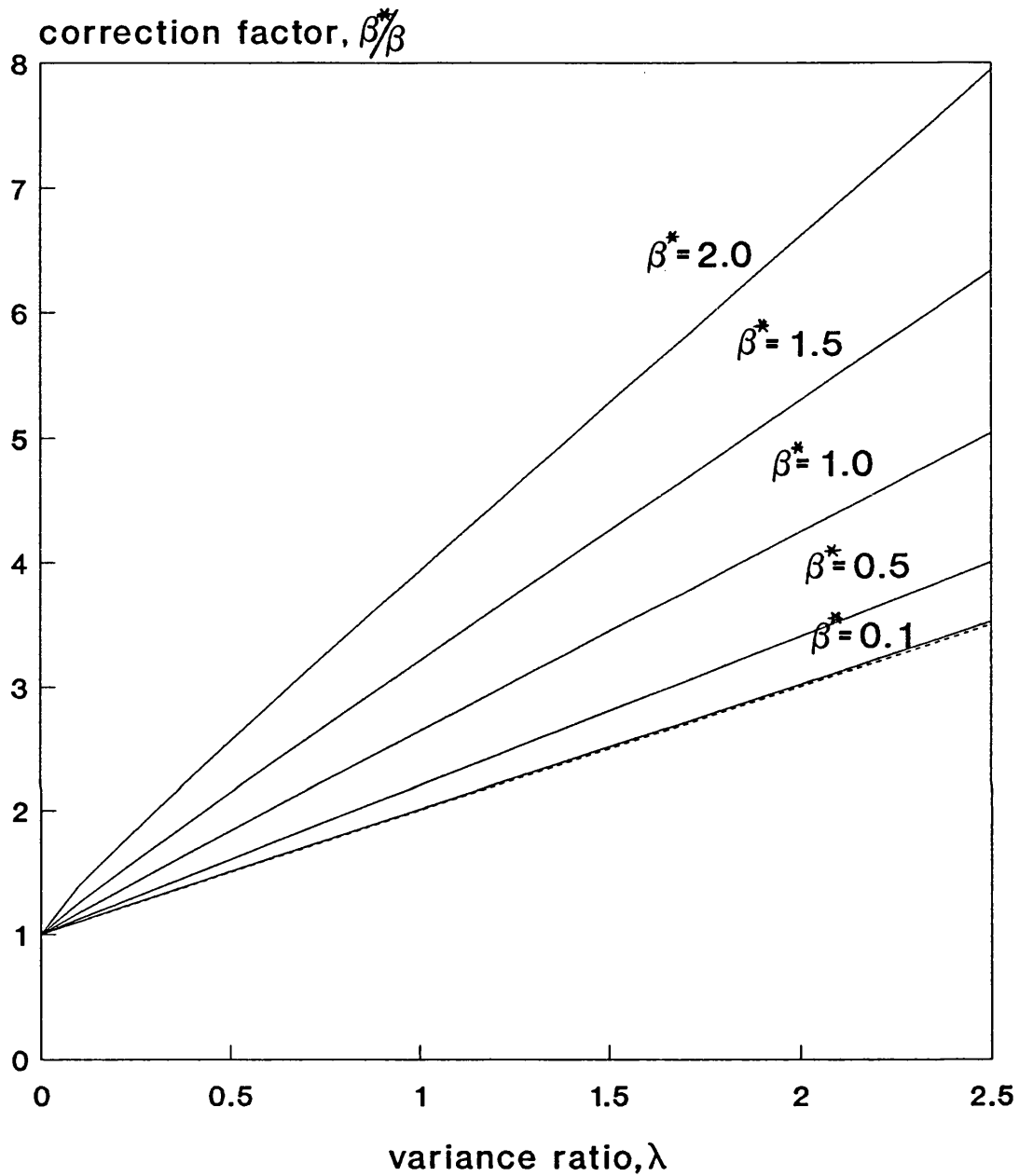


Figure 4.7: Regression dilution: the relationship between β^*/β and the variance ratio, λ , for the proportional hazards model when the covariate is measured with a normally distributed error and there is no censorship (dashed line is where $\beta^*/\beta=1+\lambda$)

values of β^* (say, up to 1) having also the convenient intercept of $\beta^*/\beta=1$ for $\lambda=0$. Thus if the gradient of this relationship can be described in terms of a simple dependence on β^* , if only approximately, then a usable relationship between β^* and β might be obtained.

Of course, the lack of knowledge about the underlying risk coefficient, β^* , makes the direct application of figure 4.7 difficult. Therefore it is also useful to consider a more direct form of the relationship between the correction factor, β^*/β and β : then given β , this would allow some inference concerning the value of β^* . Figure 4.8 shows this for several different values of the variance ratio, λ . Note that in the case of simple linear regression, $\beta^*/\beta=(1+\lambda)$ and so the correction factor is independent of β and would appear as a horizontal line on the figure with intercept $(1+\lambda)$. This figure shows very clearly the departures of the correction factor from $(1+\lambda)$ as β increases; this is particularly marked the larger λ is. Thus, unless β is close to zero, correction for regression dilution using the factor $(1+\lambda)$ would give an underestimate of β^* possibly by a considerable amount.

Recall that β is the expected value of $\hat{\beta}$, the risk coefficient obtained by solving equation (4.3) based on the observed covariate values. In practice, $\hat{\beta}$ will not equal β but if the sample size is sufficiently large then the difference between the two will be small and so figure 4.8, together with knowledge of λ , can be used to give a guide as to the likely magnitude of the underlying risk coefficient, β^* . In small samples, it would be advantageous to reduce λ by taking several measurements on each individual and to find $\hat{\beta}$ using the mean of these instead of a single value as then the correction factor necessary is not so sensitive to the departure of $\hat{\beta}$ from β .

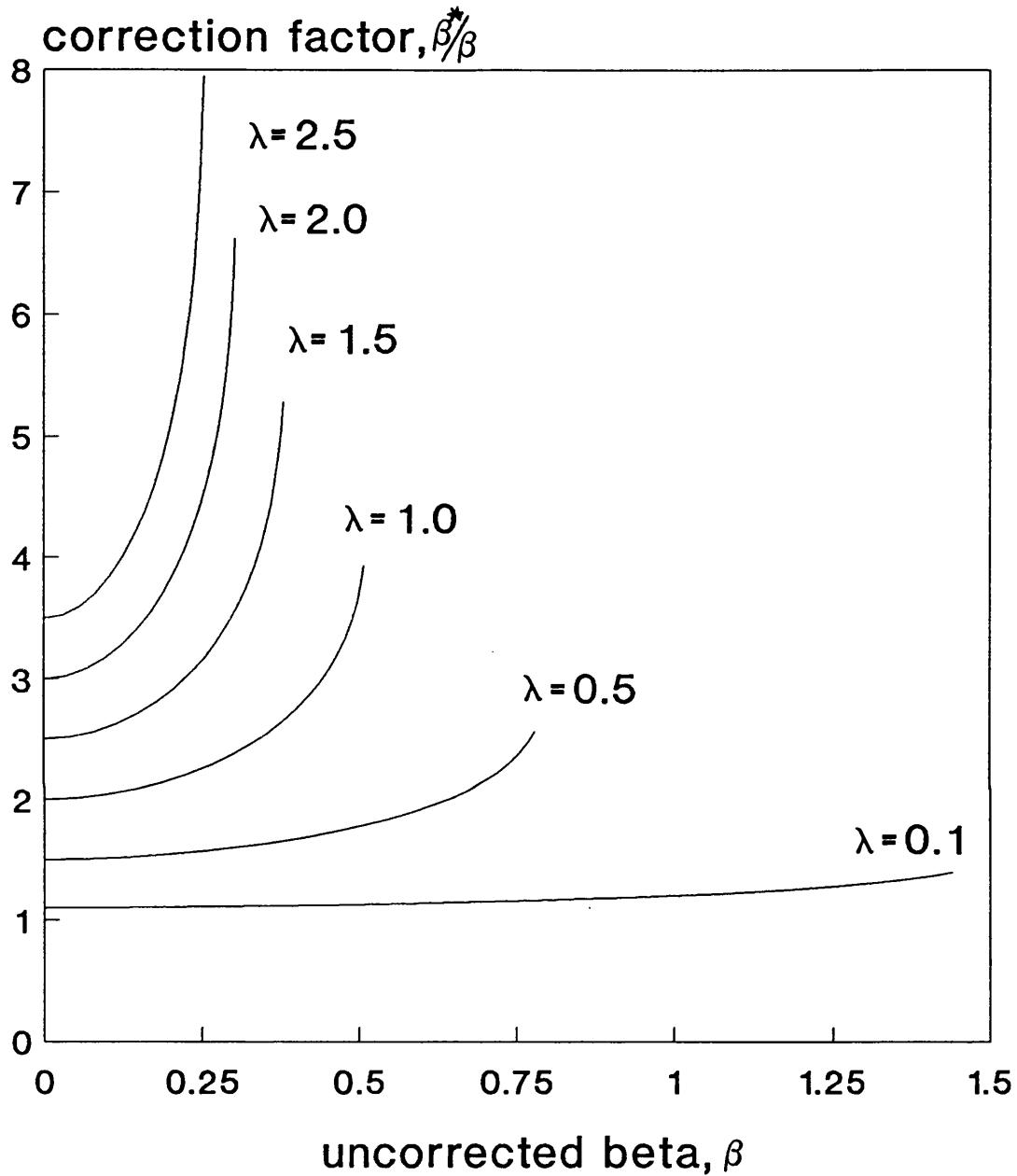


Figure 4.8: Regression dilution: relationship between the correction factor β^*/β and the expected risk coefficient relating to the observed covariate, β , in the proportional hazards model with no censorship present and normally distributed measurement errors

4.3.4 Regression dilution in the proportional hazards model with censorship present

In practice, censoring mechanisms are often difficult to describe well, so only one simple form will be considered explicitly here. This is where all subjects are followed to death or some maximum time, t_0 say, whichever is the shorter time. This type of mechanism might arise in practice, for example, in a clinical trial where a fixed length of follow-up is prescribed for each patient randomized. The effect of this censorship in equation (4.4) is to truncate the distribution of survival times, t , used in determining the right hand term and to produce a non-zero expectation for the left-hand term:

$$E_{x|T \leq t_0}(x).$$

Proceeding in a similar manner as when no censorship was present, consider first the probability density function of survival times that are observable. This is simply the density function (4.10) truncated at the time t_0 giving

$$f(t) = \frac{\int_{-\infty}^{\infty} h_0(t) \exp(\beta^* \gamma) S(t|\gamma) f(\gamma) d\gamma}{\int_{-\infty}^{\infty} [1 - S(t_0|\gamma)] f(\gamma) d\gamma} \quad 0 < t < t_0,$$

and being zero if $t > t_0$. Thus, as in obtaining equation (4.12),

$$\begin{aligned} & E_{t|T \leq t_0} \left[\frac{E_{x|T \geq t} \exp(\beta x)}{E_{x|T \geq t} \exp(\beta x)} \right] \\ &= \int_0^{t_0} \left[\frac{\int_{-\infty}^{\infty} \gamma \exp(\beta \gamma) S(t|\gamma) f(\gamma) d\gamma \int_{-\infty}^{\infty} h_0(t) \exp(\beta^* \gamma) S(t|\gamma) f(\gamma) d\gamma}{\int_{-\infty}^{\infty} \exp(\beta \gamma) S(t|\gamma) f(\gamma) d\gamma \int_{-\infty}^{\infty} [1 - S(t_0|\gamma)] f(\gamma) d\gamma} \right] dt + \beta \lambda \end{aligned}$$

$$= \int_0^{H_0(t_0)} \left[\frac{\int_{-\infty}^{\infty} \gamma \exp(\beta\gamma) \exp[-v \exp(\beta^*\gamma)] f(\gamma) d\gamma \int_{-\infty}^{\infty} \exp(\beta^*\gamma) \exp[-v \exp(\beta^*\gamma)] f(\gamma) d\gamma}{\int_{-\infty}^{\infty} \exp(\beta\gamma) \exp[-v \exp(\beta^*\gamma)] f(\gamma) d\gamma \int_{-\infty}^{\infty} [1 - \exp[-H_0(t_0) \exp(\beta^*\gamma)]] f(\gamma) d\gamma} \right] dv + \beta\lambda. \quad (4.13)$$

To determine the expectation of x conditional upon $T \leq t_0$ requires the probability density function of x given a survival time less than t_0 . Akin to obtaining $f(x|T \geq t)$ given by (4.5),

$$f(x|T \leq t_0) = \frac{\int_{-\infty}^{\infty} [1 - S(t_0|\gamma)] f(x|\gamma) f(\gamma) d\gamma}{\int_{-\infty}^{\infty} [1 - S(t_0|\gamma)] f(\gamma) d\gamma}.$$

Thus,

$$\begin{aligned} E_{x|T \leq t_0}(x) &= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x [1 - S(t_0|\gamma)] f(x|\gamma) f(\gamma) dx d\gamma}{\int_{-\infty}^{\infty} [1 - S(t_0|\gamma)] f(\gamma) d\gamma} \\ &= \frac{\int_{-\infty}^{\infty} \gamma [1 - S(t_0|\gamma)] f(\gamma) d\gamma}{\int_{-\infty}^{\infty} [1 - S(t_0|\gamma)] f(\gamma) d\gamma} \\ &= \frac{-\int_{-\infty}^{\infty} \gamma S(t_0|\gamma) f(\gamma) d\gamma}{1 - \int_{-\infty}^{\infty} S(t_0|\gamma) f(\gamma) d\gamma} \quad \text{since } E(\gamma) = 0. \end{aligned} \quad (4.14)$$

Using (4.13) and (4.14), equation (4.4) gives, in the presence of censorship at time t_0 ,

$$\begin{aligned}
& \int_{-\infty}^{\infty} \gamma S(t_0|\gamma) f(\gamma) d\gamma + \beta\lambda \left[1 - \int_{-\infty}^{\infty} S(t_0|\gamma) f(\gamma) d\gamma \right] \\
& + \int_0^{H_0(t_0)} \left[\frac{\int_{-\infty}^{\infty} \gamma \exp(\beta\gamma) \exp[-v \exp(\beta^*\gamma)] f(\gamma) d\gamma \int_{-\infty}^{\infty} \exp(\beta^*\gamma) \exp[-v \exp(\beta^*\gamma)] f(\gamma) d\gamma}{\int_{-\infty}^{\infty} \exp(\beta\gamma) \exp[-v \exp(\beta^*\gamma)] f(\gamma) d\gamma} \right] dv \\
& = 0 \tag{4.15}
\end{aligned}$$

where $S(t_0|\gamma) = \exp[-H_0(t_0)\exp(\beta^*\gamma)]$.

This is dependent on the underlying hazard function only through $H_0(t_0)$, the cumulative base hazard function evaluated at t_0 , or equivalently through $S_0(t_0)$, the base survivor function evaluated at time t_0 . Although there is a one-to-one relationship between this and the proportion in the population that would be censored, this depends on the distribution of the underlying covariate, γ , in the population as well as β^* and, in general, can only be obtained by solving this relationship numerically.

Keeping $\gamma \sim N(0,1)$, figures 4.9 and 4.10 show the relationship between the correction factor β^*/β and the variance ratio, λ , obtained by solving (4.15) for varying levels of censorship when $\beta^*=0.5$ and 1.0 , respectively. Also shown on both figures is the dashed line $\beta^*/\beta=1+\lambda$. Clear from these figures is that increasing levels of censorship reduces the correction factor needed towards $1+\lambda$ though the magnitude of this reduction is dependent upon the value of β^* .

For other forms of censoring mechanism, both survival times and the values of the covariate x amongst patients observed to die are likely to be more varied than those arising from the censoring mechanism considered here. Hence, the impact of censorship

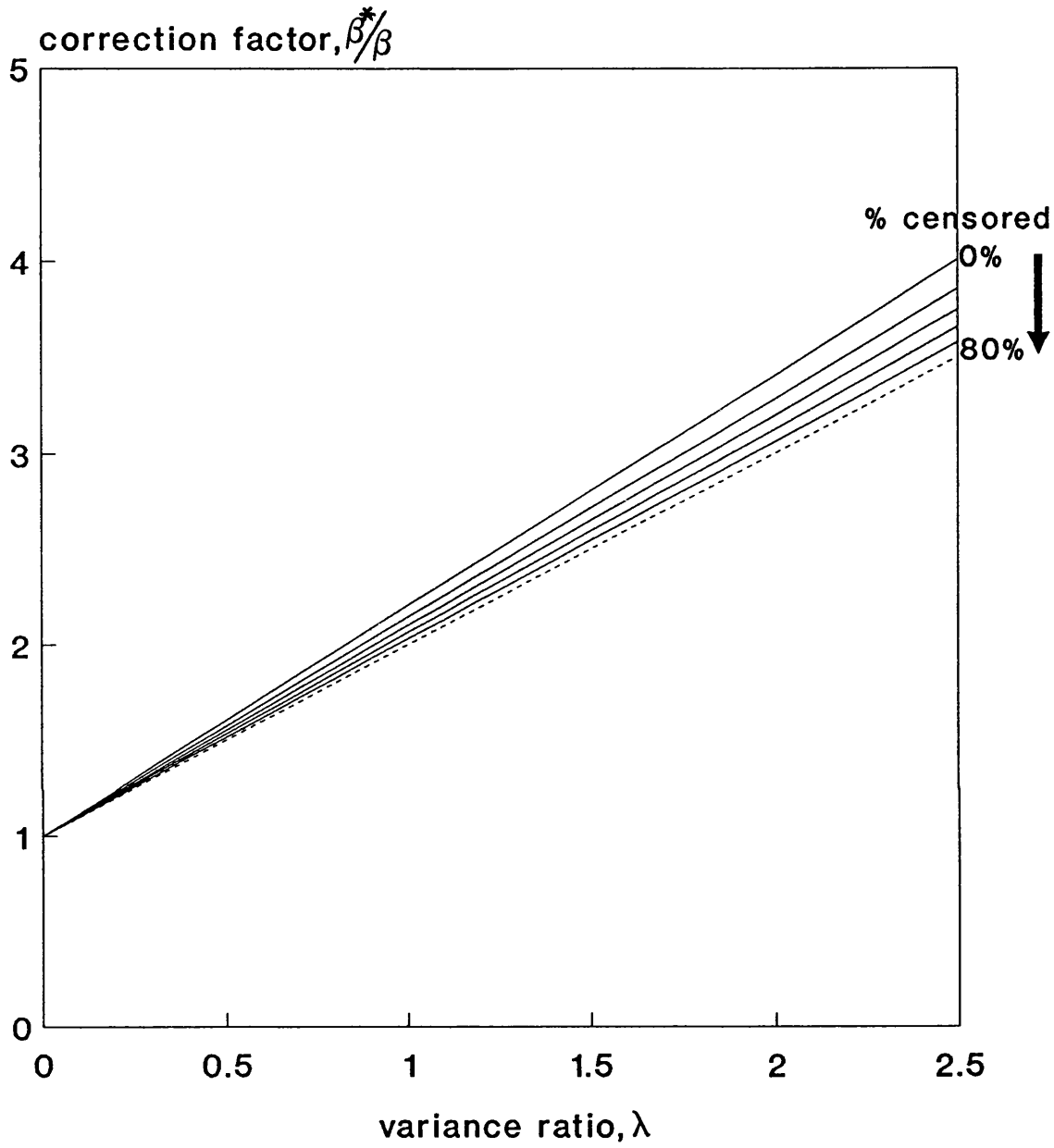


Figure 4.9: Regression dilution in the proportional hazards model with censorship present: relationship between β^*/β and λ when $\beta^*=0.5$ for different levels of censorship (dashed line is where $\beta^*/\beta=1+\lambda$)

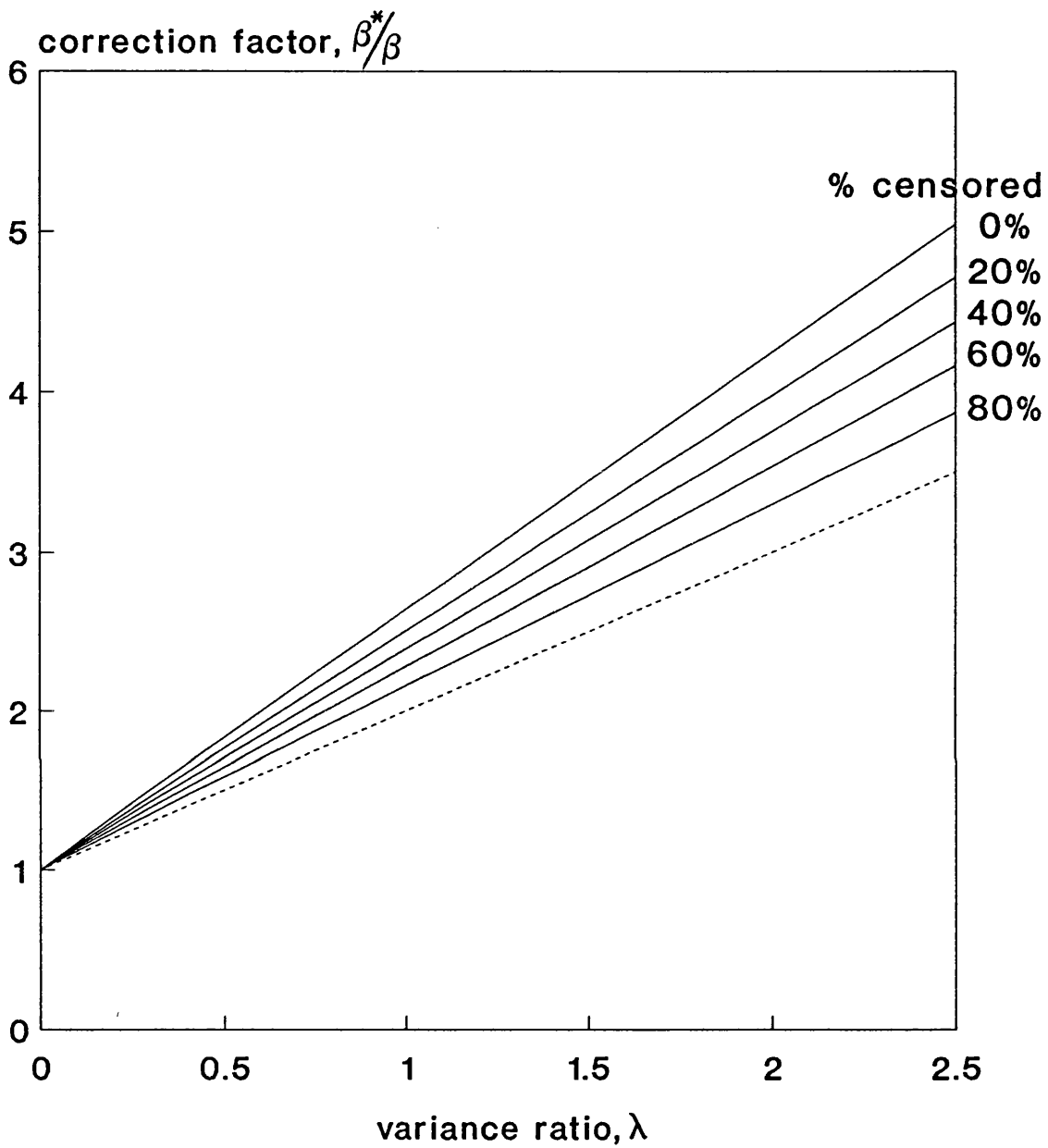


Figure 4.10: Regression dilution in the proportional hazards model with censorship present: relationship between β^*/β and λ when $\beta^*=1.0$ for different levels of censorship (dashed line is where $\beta^*/\beta=1+\lambda$)

is likely to be less marked with the relationship between the correction factor, β^*/β , and the variance ratio, λ , closer to that found with no censorship than the linear relationship with $\beta^*/\beta=1+\lambda$.

4.3.5 Blood pressure and risk: adjusting for regression dilution

This section describes how an adjustment to the risk coefficients obtained in section 4.2.2 for relating risk of cardiovascular mortality to observed blood pressure can be made to give an estimate of the risk coefficient relating to underlying diastolic blood pressure. Key to this derivation is the result from the previous chapter that the high level of censorship in this study's cohort (just over 95% in each of the three towns) implies that applying a factor of $1+\lambda$ will give an estimate that is almost unbiased particularly as the risk coefficient β^* is comparatively small (by considering the coefficient related to the mean of four measurements it will be of the order of 0.5 for a standard deviation change in underlying blood pressure level). As mentioned earlier, Prentice (1982) also suggests the use of this correction factor if the distribution of ξ given x is "concentrated". This might be interpreted as meaning that the variance ratio, λ is small. This can be better achieved by applying the correction to the risk coefficient gained for the mean of four observations as then the value of λ is reduced by a factor of four compared with when using the risk coefficient is based on a single observation.

The difficulty with applying the adjustment to the risk coefficient associated with observed blood pressures given at the end of section 4.2.2 is in defining the ratio, λ , when it is known that the data is derived from three distinct towns and that the

variance components differ between towns. Thus the approach adopted is to fit the full model within each town and then apply the adjustment factor to each town's risk coefficient based upon the variance components derived for the respective town in chapter 3. The model formulated for these components related an observed blood pressure, x , to the underlying pressure, ξ , in the following manner:

$$x \sim N\left[\xi, \sigma_W^2[1+c(\xi-\mu)]^2\right]$$

$$\xi \sim N(\mu, \sigma^2).$$

This model allows the within-person variance to change with the level, ξ . This is an additional complexity not considered in the previous section. It imposes a correlation structure between ξ (and hence γ) and δ which will have an effect on the correction factor needed. However this correlation is very weak and so, drawing on a parallel in simple linear regression (Draper and Smith, 1980, p123), its impact on the correction factor would be small. Thus, in obtaining the variance ratio, λ , the average within-person variance will be used for the term, σ_δ^2 :

$$\begin{aligned} \sigma_\delta^2 &= E\left[\sigma_W^2[1+c(\xi-\mu)]^2\right] \\ &= \sigma_W^2[1+c^2\sigma^2]. \end{aligned}$$

As the model where the observed pressure is in fact the mean of four measurements is being used, this average variance needs to be divided by four to give the appropriate value of σ_δ^2 and so

$$\lambda = \frac{\sigma_{\delta}^2}{\sigma_{\xi}^2} = \frac{\sigma_W^2 [1 + c^2 \sigma^2]}{4\sigma^2}.$$

The parameter estimates for c , σ_W^2 and σ^2 (obtained in chapter 3) are presented again in table 4.7 for each town together with the consequent correction factors $1 + \lambda$. Also given in table 4.7 are the risk coefficients associated with the observed mean of four diastolic pressures to which the correction factors are applied to give the risk coefficients shown in the same table for the underlying diastolic pressure. Standard errors are also presented; the adjusted errors being obtained from the model estimates with the same correction, $1 + \lambda$, applied. This ignores the error in estimating λ but, given the large sample size, this will be negligible.

The risk coefficients show quite marked variability between one another particularly between that for Prague and those for Edinburgh and Budapest. A pooled estimate, β_p^* , for the risk coefficient, β^* , can be obtained by finding the weighted average of the individual estimates where the weight for the i th estimate, w_i , equals the inverse of the variance of the associated adjusted estimate (that is, the square of its adjusted standard error):

$$\beta_p^* = \frac{\sum_{i=1}^3 w_i \beta_i^*}{\sum_{i=1}^3 w_i} \quad \text{with standard error} \quad \left[\sum_{i=1}^3 w_i \right]^{-\frac{1}{2}}$$

where β_i^* denotes the adjusted risk coefficient for the i th town. This gives an estimate of 0.0672 with standard error 0.0069. Thus the pooled estimate indicates an increase in risk of 6.95% for a unit increase in underlying diastolic pressure (95% confidence interval: 5.51% to 8.41%). An approximate test of homogeneity can be framed, as in

	Edinburgh	Budapest	Prague
<u>Observed mean of four diastolic pressures</u>			
Risk coefficient	0.0509	0.0381	0.0673
Standard error	0.0091	0.0100	0.0090
<u>Variance components</u>			
σ^2	60.44	46.58	36.66
σ_W^2	45.24	55.00	53.59
$c (x10^3)$	10.43	12.67	13.88
Correction factor*	1.1884	1.2974	1.3680
<u>Underlying mean diastolic pressure</u>			
Risk coefficient	0.0605	0.0494	0.0921
Standard error	0.0108	0.0129	0.0123

* Correction factor = $1 + \frac{\sigma_W^2(1+c^2\sigma^2)}{4\sigma^2}$

Table 4.7: Derivation of coefficients for associating underlying diastolic blood pressure with risk of dying from cardiovascular causes by town.

overviews, to assess the evidence for heterogeneity in the risk coefficients between towns (DerSimonian and Laird, 1986). The test statistic is

$$\sum w_i(\beta_i^* - \beta_p^*)^2.$$

Under the null hypothesis of homogeneity of coefficients between towns, this is

distributed as chi-square on two degrees of freedom. Evaluating the statistic's value gives 6.37 ($0.025 < P < 0.05$). Thus there is some evidence of heterogeneity in underlying risk coefficients between towns.

MacMahon et al (1990) have also studied the problem of regression dilution to obtain an estimate of the relationship between underlying diastolic pressure and risk of cardiovascular events. Their correction factor was obtained graphically using repeated blood pressure measurements made in the Framingham Study. Essentially, they grouped subjects into five diastolic pressure bands on the basis of pressure measurements at one, "baseline", visit. By relating the mean diastolic pressure within each group found at a visit four years after baseline, to the group means at baseline, they found that the difference between the means of the lowest and highest groups was reduced by about 60% between baseline and the later visit. This is the effect of regression to the mean and consequently they increase the risk coefficients based on observed diastolic measurements by 60% to correct for regression dilution.

This is lower than the correction factors for adjusting risk coefficients relating to a single diastolic measurement obtained in this study: increases of 75%, 119% and 147% for Edinburgh, Budapest and Prague, respectively. Notably, their correction factor is closest to that of Edinburgh where a similar policy of rounding to the even digit was practised as in the Framingham Study. However, it does illustrate the possibility, as they recognize, that a larger correction factor may be necessary than they used. It also shows the possibility for error if a factor obtained from one study's repeat measurements is applied to the estimated risk coefficient from another study carried out in a population where the within- and between-person variance components are

different to those in the first study.

Because their study does not have correction factors from each study, they are unable to investigate the possibility of heterogeneity between studies in the risk relationship with underlying diastolic pressure level. Thus there is no evidence to confirm or refute the possible heterogeneity found between the three towns in this study. Thus direct comparison can only be made between their estimate of the magnitude of the relationship and the pooled estimate obtained here: the reduction in risk associated with a reduction in underlying pressure of 7.5mmHg (as they used) from their overview was 46% for stroke and 29% for coronary heart disease compared with 40% for cardiovascular mortality in this study (36%, 31% and 50% for Edinburgh, Budapest and Prague, respectively). Thus their estimate for coronary heart disease and stroke combined (about 30% or so) would be lower than the pooled estimate obtained here though this might just reflect their use of a lower correction factor rather than real differences in risk relationships.

4.4 Summary of main results with discussion

The results of this chapter can be divided into two types. The first are those with practical implications for assessing the components of risk of cardiovascular mortality associated with different characteristics of blood pressure variability over time. The second are those concerning methodological aspects of regression dilution in proportional hazards models, that is the effect on risk assessment of a covariate that is measured with error.

Consider first risk prediction. A sequence of four measurements taken at approximately annual intervals was used to assess the value of repeated measurements. The primary result found was that mean pressure level is the dominant determinant of risk and that other characteristics such as trend and residual variability (about the trend) carry little, if any, predictive value after allowing for mean level. The result for residual variability is not surprising given that homogeneity of within-person variance across individuals was established, after allowing for level, for diastolic pressure in chapter 3, and so differences between individuals in this parameter can be attributed to sampling variation. Trends during the three-year period have been established to be small so that this study lacks power to demonstrate any predictive ability that might exist, unless risk changes dramatically with changes in trend. Therefore it was not surprising that trends were not found to be important. This supports the finding of Hofman et al (1983) who studied the predictive power of longer-term trends using data from the Framingham Study.

When considering diastolic pressure level as a predictor of risk, it is important to distinguish between models developed for predicting risk in a new individual on the basis of their observed blood pressure and models aimed at understanding the aetiology of the disease. For the first, risk coefficients relating to the mean of a given number of observed blood pressure measurements are required. For the second, it is more useful to consider the risk coefficient related to underlying pressure as this enables direct comparison with other populations and other risk factors. Table 4.8 summarizes the changes in risk derived in this study for unit changes in mean of between one and four diastolic pressure measurements and also for the underlying level (equivalent to the mean of an infinite number of measurements). Clear from this table is the importance

No. of measurements on which mean diastolic pressure based	Increase in risk for unit change in mean	95% confidence interval for increase in risk
1	3.40%	2.57%, 4.25%
2	4.46%	3.49%, 5.44%
3	4.97%	3.92%, 6.03%
4	5.36%	4.26%, 6.48%
∞ (\equiv underlying level*)	6.95%	5.51%, 8.41%

* based on pooled estimate of effects adjusted for regression dilution within each of the three towns studied

Table 4.8: Estimated risk gradients associated with the means of different numbers of diastolic blood pressure measurements

of applying the appropriate risk gradient, based on the number of measurements taken, to give predictions. However, it is equally important to understand why the coefficients change: increasing the number of measurements increases the precision with which each individual's level is assessed so that the distribution of observed means in the population becomes less variable. This is the "regression to mean" phenomenon and is the reason why "regression dilution" is such an appropriate term for the effect. An important consequence of this is that the risk coefficients that are obtained in any study for observed mean pressures depend on the relative sizes of the within- and between-person variances for diastolic pressure and so are not readily applicable to other populations unless the variance components are the same in each population. This latter requirement is unlikely to be fulfilled if only because population distributions of pressure are partly determined by age profiles which can vary considerably between populations. This is why the determination of the risk gradient with underlying level is so important as it does not depend on these population traits.

This requires the adjustment for the effect of regression dilution of the risk coefficients relating to observed pressures.

The problem of regression dilution in the log-linear proportional hazards model was examined by relating the risk coefficient, β^* , associated with the underlying covariate, ξ , to the expected value, β , of the risk coefficient, $\hat{\beta}$, which is the solution to the maximum likelihood equations with the covariate x (which is equal to ξ measured with error) replacing ξ . With a proportional hazards model assumed for ξ , the proportional hazards model does not apply for x . Even so, β^* was found to be related to β in a way which does not depend upon the base hazard function. However it does depend upon the value of β^* and on the proportion of observations that are censored.

When the measurement error follows a normal distribution with zero mean, figure 4.7 shows the relationship between β^*/β and λ , the ratio of the within- to between-subject variance components, for the case when there are no censored observations. An important result here is that $\beta^*/\beta \rightarrow 1 + \lambda$ as $|\beta^*| \rightarrow 0$, which is similar to the relationship found for the errors in variable problem in simple linear regression where $\beta^*/\beta = 1 + \lambda$ for all values of β^* . Thus applying the correction $(1 + \lambda)$ to the risk coefficient obtained using the observed covariate values, x , will give a lower bound for the estimate of β^* . This will be a good approximation to an unbiased estimate in many practical applications where the risk gradient is relatively small.

For larger values of β^* , the dependence of the ratio β^*/β on λ is approximately linear except for very small values of λ . Further work would be useful to determine whether a simple analytical form for this relationship can be found. Without this, figure 4.8 can

be used to give a guide for relating a risk coefficient, $\hat{\beta}$, estimated using observed pressures to the underlying risk gradient, β^* . However, this figure demonstrates the sensitivity of the value of β^* found to the precision to which the risk coefficient for observed pressure is determined. Hence its value will particularly lie in those situations where large sample sizes are available.

One specific type of censoring was considered: when all subjects are observed to some maximal time unless death occurs before this. In this setting, the extent of the adjustment needed depends on the proportion of the population who are censored. No simple result follows directly from this because this proportion also depends upon the underlying risk gradient, β^* . However, one important result found is, again, that the ratio $\beta^*/\beta \rightarrow 1+\lambda$ if there is a high level of censorship. This would apply for many population studies of chronic diseases where only a small proportion of subjects followed-up are observed to death. This agrees with a conclusion of Prentice (1982) that when the probability of survival during the follow-up period is close to unity, a proportional hazards model for the covariate x is a good approximation, and that then the ratio of the risk gradients associated with ξ and x is approximately $1+\lambda$. This limit also imposes a lower bound on the value of β^* which can be used in conjunction with an upper bound obtained from either figure 4.7 or 4.8. Again, further work is suggested to see whether an analytical approximation that describes the relationship between β^*/β and λ with the level of censorship through its dependence on β^* can be found.

Some further comment is also useful on the issue of comparing risk gradients found in different populations. As has been seen, the magnitude of risk coefficients relating to

observed blood pressures depends both on the underlying risk gradient, β^* , and also on the value of the variance ratio, λ . Thus with covariates measured with error, differences in the associated risk coefficients between studies might be attributable either to differences in the underlying risk gradients or to differences in the confounding parameter, λ , or to both. Thus, in conducting an overview of studies, the ideal would be to obtain estimates of β^* which reflect adjustment for regression dilution. This requires information about λ (and hence also, implicitly, about the number of measurements on which an observed mean pressure is based). Then an analysis of these can be carried out to assess the evidence for heterogeneity and, if appropriate, pooled estimates obtained. If the required information is available, this would be relatively straightforward in a simple linear regression context. However, the dependence, in survival models, of the adjustment on the unknown magnitude of β^* makes this difficult unless the studies are all large or the level of censorship in each is high.

If these conditions are not satisfied, then some consideration of heterogeneity might be possible if it can be established that the values of λ are similar across studies. Then evidence for heterogeneity in the risk coefficients related to observed blood pressures can be assessed and a pooled estimate obtained for which, because it will be based on a larger sample size, the appropriate factor for adjusting for regression dilution might be determined. Of course, this demonstrates the importance of obtaining information about the within- and between-subject variance components in any study or, alternatively, removing the need for marked corrections to risk coefficients by determining precisely each individual's underlying pressure level through repeated measurements of blood pressure on several occasions.

5. Screening blood pressure

5.1 Introduction

On what basis should a doctor decide that intervention is desirable to try to lower an individual's blood pressure? Whilst this is ultimately a question requiring a clinical decision, there is an important role for statistical methodology in formulating aids for the decision-making process. In particular, this contribution needs to assess what evidence there is that an individual's underlying mean blood pressure is unduly elevated, and to what extent, given a sequence of blood pressure measurements. More generally, the problem concerns the use, for screening, of repeated measures taken on a continuous variable that varies about some unknown level and where risk of disease varies, also continuously, with that level. Thus the classical diagnostic tests associating a positive or negative result, from a single test, with a binary disease state are not of a clear direct use. Whilst this chapter focuses on screening blood pressure because of its association with risk of cardiovascular disease, the concepts put forward are generalizable to other situations with continuous risk factors measured "with error".

There is certainly an awareness amongst some doctors of the problem that the variability of blood pressure causes in reaching a decision about the need for intervention. This has been formalized by the publication of guidelines for treating mild hypertension by the World Health Organization (Participants of the W.H.O./I.S.H. meeting, 1986). Formulated by an expert committee on the basis of a review of the scientific evidence, these guidelines (shown, diagrammatically in figure 5.1) illustrate the framework for a screening programme. On the basis of the accumulated information from a set of repeat blood pressure measurements (generally their mean), a

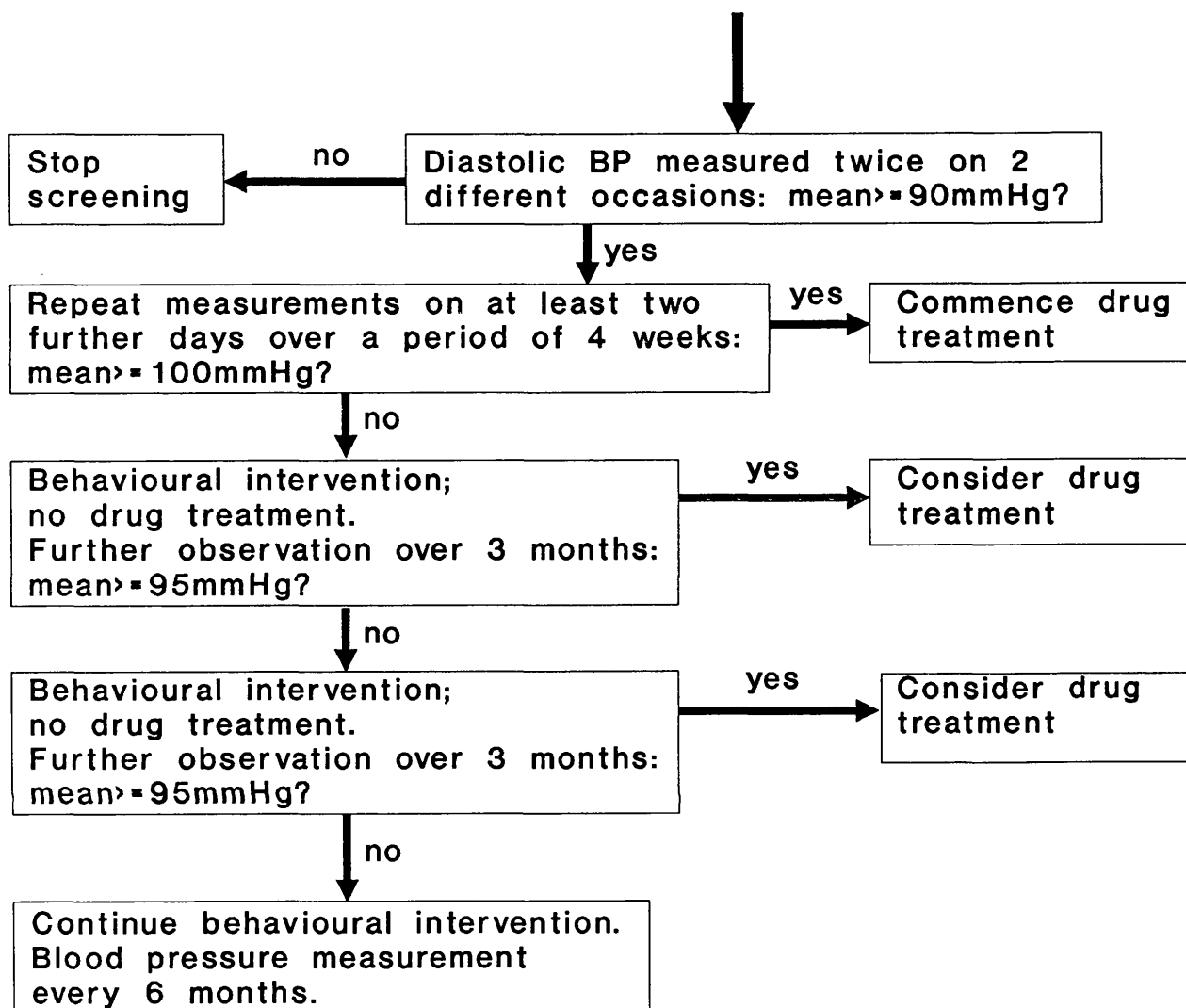


Figure 5.1: Approach to screening for elevated diastolic blood pressure (phase V) suggested by the World Health Organization guidelines (from: Participants of W.H.O./I.S.H. meeting, 1986)

decision is made either to consider pharmacological intervention or to continue screening (whilst encouraging behavioural change such as stopping smoking, and altering diet to lower blood pressure, to lower serum cholesterol and to control diabetes), or to stop screening. Note, though, that these guidelines only allow this last option after the first two measurements have been taken and not later. In contrast, the British Hypertension Society working party on the treatment of mild hypertension (1989) only specify a mean diastolic blood pressure in excess of 100mmHg over three to four months as an indication for drug treatment but do not specify how many measurements should be taken. This fails to reflect fully the fact that, for example, for any individual a mean of two measurements is more likely to exceed this level than a mean of four because of the reduced variance of the sample mean obtained by increasing the number of measurements on which it is based.

The screening programme studied in this chapter is a generalization of that proposed in the World Health Organization guidelines. Essentially, after the i th measurement, a decision is made as follows:

- (i) if $\bar{x}_i > U_i$, then consider intervention
- (ii) if $\bar{x}_i < L_i$, then stop screening
- (iii) else if $L_i < \bar{x}_i < U_i$, then continue screening to visit $(i+1)$,

where \bar{x}_i is the mean of measurements taken up to and including the i th visit. Here the \underline{L} and \underline{U} are critical values to be determined, for which the methods for doing so form the basis of this chapter. Two approaches are put forward. The first emphasizes the assessment of evidence for blood pressure raised above some defined level such as that

level at which intervention is justifiable (section 5.3): this develops previous work in this field. Section 5.4 extends this to consider the evidence for a raised risk in the individual patient. In particular this section develops quantification of the extent of risk elevation as the focus of screening. It then considers the role of other risk factors in screening for risk of cardiovascular disease which requires assessment of absolute risk and not just relative risk.

The second approach takes a public health perspective with an emphasis on identifying that section of the population, of a defined size, for intervention with the greatest expected event rate subject to constraints such as those imposed by intervention and screening costs (section 5.5). First, though, in the next section we briefly review the previous statistical work on screening blood pressures.

5.2 Review of statistical developments for screening blood pressure

Rosner (1977) explored the parallel between classical diagnostic tests and hypothesis testing to generate a screening rule for blood pressure based not on the existence of disease, *per se*, but on the availability of a treatment for use in individuals with raised underlying levels. This then forms the dichotomy used in these tests and so reduces the problem to one of relating the information from the repeat measurements to the underlying level, ξ . He formulated a rule to test the alternative hypothesis that $\xi > \xi_T$ against the null hypothesis that $\xi \leq \xi_T$ where ξ_T is some threshold for the underlying mean blood pressure level, above which treatment for high blood pressure is indicated. The problem of choosing ξ_T is considered in the discussion. By deriving a t-statistic, Rosner suggests labelling an individual as “hypertensive” if, after a pre-defined number

of blood pressure measurements, the null hypothesis can be rejected at some chosen level; otherwise he is labelled “normotensive”. The choice of the number of measurements is determined by the rate of type II error: that is, being labelled normotensive when truly hypertensive. Of necessity, two visits are required to obtain an estimate of within-person variability and Rosner suggests four visits so as to reduce the type II error rate sufficiently; indeed Liu et al (1978) suggest, in a letter commenting on Rosner’s approach, that a larger number than this is necessary to give adequate power. If all individuals have such a number of measurements then this is a considerable practical constraint both in terms of screening workload and maintaining the cooperation of the screenees. Thus, the dependence of this frequentist approach on information from screening the individual concerned, alone, and on an error rate for the individual’s underlying mean level being equal to ξ_T might explain why this approach has not been pursued further.

There is an overlap in methodology between screening the individual for potentially high blood pressure and screening the individual for entry into a clinical trial of hypertensive treatments. For the latter, the aim is to identify individuals with underlying mean blood pressures within a certain range. Goldman (1976) considered the latter problem and formulated a rule whereby an individual was entered into a trial if at each visit i , the observed blood pressure (not the mean of the accumulated observations) was between limits L_i and U_i (which may vary between visits). Using this definition, the probability of an individual, with given underlying mean blood pressure and within-person variability, entering the trial could be derived. Also, if population distributions for the underlying means and variabilities can be defined, then the distribution of underlying means for those entered into the trial could be derived

(using Bayes theorem) as well as the screening costs per trial entrant of different screening policies (i.e. different choices of L_i and U_i). Two difficulties arise from this work: first, the population distributions used were empirical distributions of observed blood pressures without attempting to relate this to the underlying parameters and so incorporated a component of variance reflecting sampling variation; second, the only information used at any visit is the observed blood pressure at that visit rather than some summary of that and previous pressures measured.

These limitations have been considered by Rosner and Polk (1983) in their development of the “predictive value screening rule”. Using normal models to describe the within-person variability about the individual’s underlying mean and the distribution of underlying levels in the population, they obtained estimates for the parameters of the underlying distributions for various age/sex/race groups, using data from both a community blood pressure study and the Hypertension Detection and Follow-up Program. Their predictive value for an individual is then the probability that $\xi > \xi_T$ conditional upon the mean of the blood pressures so far observed in the screening process for that individual. Essentially the same approach was used by Wilson and Hebel (1988) in their development that uses information from both diastolic and systolic blood pressures to make inferences about an individual’s underlying diastolic pressure level.

In the next section, the results of chapter 3 are used to derive a screening model for individual assessment similar to that put forward by Rosner and Polk. The effect of allowing for increasing within-person variability with level is considered for this model and the consequences for prediction of differing parameter values between towns

assessed. Following this, the model is extended so that the association between risk of death from cardiovascular disease and elevated blood pressure contributes to the screening process (section 5.4).

5.3 Screening the individual for treatment

5.3.1 The general model

The relationship between a sequence of observed blood pressures, \underline{X} , and an individual's underlying mean level, ξ , in a population can be described as

$$\begin{aligned}\underline{x} &\sim f(\underline{x}|\xi, \underline{\theta}) \\ \xi, \underline{\theta} &\sim g(\xi, \underline{\theta})\end{aligned}$$

where $\underline{\theta}$ is a vector of other parameters that are not associated with risk which may vary between individuals (for example, within-person variability). This assumes that the level ξ is the important factor to control because of its association with risk but the concepts are extendable if other factors are also relevant. In screening the individual, inference can then be based upon the individual's underlying level, ξ , given one or more observed blood pressures, \underline{x} . This is achieved employing Bayes Theorem to give

$$h(\xi|\underline{x}) = k \int f(\underline{x}|\xi, \underline{\theta}) g(\xi, \underline{\theta}) d\underline{\theta}$$

where k is a normalization constant. The difficulty, already mentioned, is how to turn this conditional distribution of belief about ξ into a rule to guide decision-making as whether to intervene to lower blood pressure or not.

Following Rosner and Polk (1983), assume that there is a level of underlying blood pressure, ξ_T , above which intervention is indicated. The probability that $\xi > \xi_T$ given the observed blood pressures is then

$$\begin{aligned} p(\xi > \xi_T | \underline{x}) &= \int_{\xi_T}^{\infty} h(\xi | \underline{x}) d\xi \\ &= k \int_{\xi_T}^{\infty} \int_{\underline{\theta}} f(\underline{x} | \xi, \underline{\theta}) g(\xi, \underline{\theta}) d\underline{\theta} d\xi . \end{aligned}$$

The decision to intervene is then taken if $p(\xi > \xi_T | \underline{x}) > p_U$ for some value p_U .

In practice, it is highly desirable for any screening rule to be applied sequentially so that intervention begins as soon as there is good evidence of a sufficiently elevated underlying level. Conversely, it is difficult to justify repeated screening if this blood pressure level is adequately low, both on cost grounds and in the interest of cooperation from the individual being screened. To achieve this, the following screening rule can be defined after i screening visits have been completed:

- (i) if $p(\xi > \xi_c | \underline{x}_i) > p_U$ then consider intervention
- (ii) if $p(\xi > \xi_c | \underline{x}_i) < p_L$ then stop screening
- (iii) else if $p_L < p(\xi > \xi_c | \underline{x}_i) < p_U$ then continue screening to visit $(i+1)$.

where \underline{x}_i denotes the vector of measurements up to visit i . There is no necessity for p_L and p_U to be constant from visit to visit, though such a constraint does imply that the same level of evidence is required at any visit before a decision is reached. In contrast, though, it might be reasonable for p_L and p_U to vary from individual to individual if

different individuals have different perceptions of the balance of benefits to disbenefits of treatment or if there are other aspects of their clinical state that might influence the decision to intervene. In particular, the age of the patient might be an important determinant of p_L and p_U : given the potential for side-effects of pharmacological treatment, a greater degree of certainty might be required for raised blood pressure before a younger subject is recommended for treatment in comparison with an older subject. In contrast, if a population perspective is taken for screening, then it may be desirable to use higher values of p_U and lower values of p_L at earlier visits to reduce the extent of misclassification that might be expected at the earlier visits when less information is available. This issue is returned to in section 5.5.

5.3.2 The normal model for screening the individual for treatment

To put the general model into practice, distributional forms for f and g must be assumed. Rosner and Polk (1983) used a simple model with normal distributions for both:

$$\begin{aligned} x &\sim N(\xi, \sigma_x^2) \\ \xi &\sim N(\mu, \sigma^2) \end{aligned} \tag{5.1}$$

where μ , σ^2 and σ_x^2 were assumed constant within given age/sex/race groups. In this case, a standard application of Bayes Theorem after i visits gives:

$$h(\xi|\underline{x}_i) = N \left[\left[\frac{i\bar{x}_i}{\sigma_x^2} + \frac{\mu}{\sigma^2} \right] \left[\frac{i}{\sigma_x^2} + \frac{1}{\sigma^2} \right]^{-1}, \left[\frac{i}{\sigma_x^2} + \frac{1}{\sigma^2} \right]^{-1} \right] \tag{5.2}$$

where \bar{x}_i is the mean of the blood pressures observed from visits 1 to i . Thus

$$p(\xi > \xi_T | \bar{x}_i) = 1 - \Phi \left[\xi_T \left[\frac{i}{\sigma_x^2} + \frac{1}{\sigma^2} \right]^{1/2} - \left[\frac{i\bar{x}_i}{\sigma_x^2} + \frac{\mu}{\sigma^2} \right] \left[\frac{i}{\sigma_x^2} + \frac{1}{\sigma^2} \right]^{-1/2} \right] \quad (5.3)$$

where Φ is the standard normal cumulative distribution function.

However, in chapter 3, the following distributional form for diastolic pressure was found to fit the data well:

$$\begin{aligned} x &\sim N \left[\xi, \sigma_W^2 [1 + c(\xi - \mu)]^2 \right] \\ \xi &\sim N(\mu, \sigma^2) \end{aligned} \quad (5.4)$$

This introduces into σ_x^2 a dependence on ξ and produces a problem which requires a numerical solution:

$$p(\xi > \xi_T | \bar{x}_i) = \int_{\xi_T}^{\infty} \sqrt{(i/2\pi)\sigma^{-1}\sigma_W^{-1}[1+c(\xi-\mu)]^{-1}} \exp \left[-\frac{1}{2} \left[\frac{i(\bar{x}_i - \xi)^2}{\sigma_W^2 [1+c(\xi-\mu)]^2} + \frac{(\xi - \mu)^2}{\sigma^2} \right] \right] d\xi \quad (5.5)$$

This is the model that we pursue further comparing it with the model ignoring the dependence of σ_x on ξ given by (5.1). First, however, it is worth noting two other complexities that might also be introduced as parameters in θ . First, the time interval between screening visits needs to be sufficiently long so that serial correlation between measurements within an individual is not a problem: for blood pressure, Wilson et al (1981) show that for measurements seven or more days apart, the first-order

autocorrelation is negligible. Similarly, screening visits should not be too widely spaced as then long-term trends in underlying level are also relevant. From the results of chapter 3, the additional component of variance from ignoring trends over three years can be considered small compared with the within-person residual variance. Thus for blood pressure screening, a programme over a period of up to two or three years and with visits at least a week apart would seem reasonable. Second, within-person variance is assumed to be a constant between individuals (except for its dependence on level). Goldman (1976) allowed this to vary using an empirical distribution obtained from the work of Armitage and Rose (1966). However this overlooked sampling variability in the estimate of the variance which would be particularly marked given the small number of observations. Chapter 3 suggested that there was no noticeable heterogeneity (apart from that explained by its dependence on underlying level) and so, for diastolic pressure, constant within-person variance would seem reasonable. However, this may not be so in other screening applications and it may be important to recognize its contribution.

In constructing a sequential screening programme, an appropriate choice for the distribution $g(\xi)$ needs to be made. This should describe the distribution of ξ in subjects being screened. If there is no selection of subjects, this can be taken as the distribution of underlying diastolic pressures in the general population. However, if it can be sensitised to the individual being considered, then this will aid precision and so reduce the screening required. Simple ways to achieve this are to employ distributional forms appropriate to the individual's sex and age.

To illustrate the development of a screening programme, consider a programme

directed at middle-aged men in Edinburgh. If the men in this study's cohort are representative of the wider population, then appropriate estimates of μ and σ^2 (within age-groups), σ_W^2 and c are given in table 3.10. These estimates of within-person variability will overestimate the true variability if the programme for any individual is over a short time period but this will only have a minor effect given the small difference between these values and those obtained after adjusting for trends. If ξ_T is taken as 95mmHg and intervention is to be considered if $p(\xi > \xi_T | \bar{x}_i) > 0.9 = p_U$, say, at any visit i , then critical values, U_i , can be found by solving equation (5.5) such that greater values imply intervention. Similarly if $p(\xi > \xi_T | \bar{x}_i) < 0.1 = p_L$, say, then critical values, L_i , can be found so that smaller values of \bar{x}_i imply no further screening visits. Individuals with $L_i < \bar{x}_i < U_i$ at a visit i would then be asked to attend for a further measurement at visit $(i+1)$. Table 5.1 shows the screening programme obtained for up to six visits. Also shown in this table are the critical values obtained when $g(\xi)$ is not chosen to reflect an individual's age-group. Each critical value is given to an accuracy of one decimal place despite the fact that rounding to the nearest even digit was practised. However, the component of variance attributable to this rounding practice is negligible (shown in chapter 3 to be less than 1mmHg^2 when the total within-person variance averages about 45mmHg^2), so that these figures would be applicable when greater measurement accuracy is obtained possibly even by using automatic measuring machines.

To illustrate the use of this programme, consider three individuals from Edinburgh in the age group 50-54. Their blood pressures at the four visits considered were:

Age-group		Screening visit					
		1	2	3	4	5	6
≤ 39	U	121.9	110.8	106.6	104.4	102.9	101.9
≤ 39	L	96.6	94.3	93.7	93.5	93.4	93.3
40-44	U	119.5	109.4	105.6	103.6	102.3	101.4
40-44	L	94.4	93.1	92.9	92.8	92.9	92.9
45-49	U	118.4	108.8	105.2	103.3	102.0	101.2
45-49	L	93.3	92.6	92.5	92.5	92.6	92.7
50-54	U	116.2	107.6	104.4	102.6	101.5	100.8
50-54	L	91.6	91.7	91.9	92.1	92.2	92.4
55-59	U	116.6	107.8	104.5	102.7	101.6	100.8
55-59	L	91.6	91.6	91.9	92.1	92.2	92.4
60+	U	116.1	107.5	104.3	102.6	101.5	100.8
60+	L	91.7	91.7	91.9	92.1	92.2	92.4

All ages	U	118.3	108.7	105.2	103.2	102.0	101.2
	L	93.1	92.5	92.4	92.5	92.6	92.6

Table 5.1: Screening the individual for treatment in Edinburgh: critical values for L_i and U_i for visit i by age-group when $\xi_T=95\text{mmHg}$, $p_U=0.9$ and $p_L=0.1$

individual 1: 110 104 112 100

individual 2: 90 82 88 80

individual 3: 106 90 98 104.

For the first individual, blood pressure at visit 1 lies between the critical values and so screening continues. After two visits, the mean is 107 and is between the critical

values for visit 2. After three visits, the mean is 108.67 and is now greater than U_3 and so screening would stop here and intervention to lower blood pressure would be considered because there is reasonable certainty that $\xi > 95$ mmHg for that individual. For the second individual, the first blood pressure is below L_1 (=91.6) and so screening would stop concluding that ξ is indeed below 95 mmHg, with reasonable certainty, for that individual. However for the third individual, the mean blood pressures after one, two, three and four measurements are 106, 98, 98 and 99.5 mmHg, respectively. These are all between the appropriate critical values and so screening would continue as the evidence is not sufficiently conclusive about that individual's need for intervention. Note, though, how this ignores how high or low ξ may be when making any decision about stopping or continuing screening, or about considering intervention; this issue is returned to in section 5.4.

Comparing the screening programmes in table 5.1 between age-groups, it is clear that after three or four visits, the difference between corresponding critical values is, for practical purposes, small relative to the accuracy with which the mean is determined (for example, 0.5mmHg for a mean of four measurements). Similar practical considerations apply at earlier visits except, perhaps, for the youngest age-group where the critical values are more discrepant. Thus, for simplicity, it might be reasonable to use the common programme for all ages (also given in table 5.1). From a statistical point of view, this similarity of critical values reflects the constancy of the within-person variance, $\sigma_{\bar{x}}^2$, across the age-groups and the fact that the dependence of the distribution $h(\xi|\bar{x}_i)$ on the age-related parameters μ and σ^2 is considerably reduced as the number of visits increases. This is readily seen by considering equation (5.2) without the dependence of $\sigma_{\bar{x}}^2$ on ξ when rewritten as

$$\begin{aligned}
h(\xi|\bar{x}_i) &= N \left[\left[\bar{x}_i + \mu \frac{\sigma_x^2}{i\sigma^2} \right] \left[1 + \frac{\sigma_x^2}{i\sigma^2} \right]^{-1}, \frac{\sigma_x^2}{i} \left[1 + \frac{\sigma_x^2}{i\sigma^2} \right]^{-1} \right] \\
&\rightarrow N \left[\bar{x}_i, \frac{\sigma_x^2}{i} \right]
\end{aligned}$$

quickly as the number of visits, i , increases since the ratio of within- to between-person variance has order of magnitude, unity.

There is no particular reason for taking p_L as 0.1 and p_U as 0.9. Indeed, any pair of values might be chosen. In table 5.2, the critical values for various choices of $p(\xi > 95|\bar{x}_i)$ are given for all age-groups combined for Edinburgh, Budapest and Prague. Even here, where the ratio σ_W^2/σ^2 varies two-fold from 0.75 in Edinburgh to 1.18 in Budapest and to 1.46 in Prague, there are only small differences between the corresponding critical values between towns particularly after the second or third visit. Given that some of the difference between towns in the within-person variability can be attributed to the digit preferences operating, it might be reasonable to employ a single screening programme in all three towns particularly if blood pressure reading is standardized and, as with the World Health Organization guidelines, two or three measurements are required of any individual before any decision is made.

Worth noting is the similarity of the values corresponding to a probability of 0.1 in Edinburgh and Budapest, and to 0.05 in Prague. Although the lines will increase toward ξ_T (=95mmHg in this example) for larger numbers of visits than shown, these would provide useful rules in deciding that no further screening is justified simply by using a fixed cut-off level regardless of the visit number. For example an observed mean diastolic blood pressure below 92mmHg at any visit might be appropriate in

$p(\xi > \xi_T \bar{x})$	Screening visit					
	1	2	3	4	5	6
<u>Edinburgh</u>						
0.99	128.1	115.3	110.3	107.6	105.9	104.7
0.95	121.7	111.0	106.9	104.8	103.4	102.4
0.9	118.3	108.7	105.2	103.2	102.0	101.2
0.8	114.1	106.0	103.0	101.4	100.4	99.7
0.7	111.0	104.0	101.4	100.1	99.2	98.7
0.5	105.9	100.6	98.8	97.9	97.3	96.9
0.3	100.7	97.3	96.2	95.7	95.4	95.2
0.2	97.6	95.3	94.6	94.3	94.2	94.1
0.1	93.1	92.5	92.4	92.5	92.6	92.6
0.05	89.4	90.1	90.6	90.9	91.2	91.4
0.01	82.3	85.7	87.2	88.1	88.7	89.2

<u>Budapest</u>						
0.99	135.4	119.8	113.6	110.3	108.1	106.6
0.95	128.1	114.9	109.8	107.0	105.2	104.0
0.9	124.1	112.2	107.7	105.2	103.7	102.6
0.8	119.3	109.0	105.2	103.1	101.8	100.9
0.7	115.7	106.7	103.3	101.6	100.5	99.7
0.5	109.7	102.8	100.3	99.0	98.2	97.7
0.3	103.6	98.9	97.3	96.5	96.0	95.7
0.2	99.9	96.5	95.4	94.9	94.6	94.5
0.1	94.5	93.1	92.8	92.7	92.8	92.8
0.05	89.9	90.3	90.7	91.0	91.2	91.4
0.01	80.9	85.0	86.6	87.6	88.3	88.8

<u>Prague</u>						
0.99	138.0	121.5	114.8	111.2	108.8	107.2
0.95	130.7	116.5	110.9	107.9	105.9	104.6
0.9	126.7	113.8	108.8	106.1	104.4	103.2
0.8	121.8	110.6	106.3	103.9	102.5	101.5
0.7	118.2	108.2	104.4	102.4	101.1	100.3
0.5	112.2	104.2	101.3	99.8	98.9	98.2
0.3	105.9	100.2	98.2	97.2	96.6	96.2
0.2	102.1	97.8	96.3	95.6	95.2	95.0
0.1	96.6	94.3	93.7	93.4	93.3	93.3
0.05	91.8	91.4	91.5	91.6	91.7	91.8
0.01	82.4	85.9	87.3	88.2	88.7	89.2

Table 5.2: Screening the individual for treatment: critical values for L_i and U_i for visit i for all ages by town when $\xi_T=95\text{mmHg}$

$p(\xi > \xi_T \bar{x})$	Screening visit					
	1	2	3	4	5	6
0.99	125.1	112.7	108.2	105.9	104.4	103.4
0.95	119.1	108.9	105.3	103.4	102.2	101.4
0.9	115.8	106.9	103.7	102.1	101.0	100.3
0.8	111.9	104.4	101.8	100.4	99.6	99.0
0.7	109.1	102.6	100.4	99.3	98.6	98.1
0.5	104.4	99.7	98.1	97.4	96.9	96.6
0.3	99.8	96.8	95.9	95.4	95.2	95.0
0.2	96.9	95.0	94.5	94.3	94.2	94.1
0.1	93.0	92.6	92.6	92.7	92.8	92.8
0.05	89.8	90.5	91.0	91.3	91.6	91.8
0.01	83.7	86.7	88.0	88.8	89.4	89.8

Table 5.3: Screening the individual for treatment when within-person variability is assumed not to depend on ξ : critical values for L_i and U_i for visit i for all ages in Edinburgh when $\xi_T = 95 \text{mmHg}$

giving reasonable evidence against the need for treatment. In practice, given the accuracy of blood pressure measurement, this may be sufficient.

Rosner and Polk (1983) assumed that $\sigma_{\bar{x}}^2$ is not dependent on the underlying level, ξ . It is useful to consider what differences such an assumption produces compared with the results so far presented where the dependence is acknowledged. Table 5.3 is comparable to the Edinburgh section of table 5.2 but with the dependence of the within-person variance on the level, ξ , removed (that is, the parameter c in the model for blood pressure variability in (5.4) is taken as zero). The width of the interval $U_i - L_i$ at any visit i is less wide when $c=0$. This reflects less strict criteria both for considering intervention and for stopping screening caused by the underestimation of

the within-person variance by ignoring its dependence on ξ at higher underlying pressures. Similar orders of difference between the programmes with c equal or not equal to zero were found for other values of ξ_T . Thus decisions will be reached sooner than is appropriate if the dependence is not taken account of. Inevitably, this will result in some additional misclassification of individuals to intervention when it is not warranted.

In putting this form of screening programme into practice, a graphical display as shown in figure 5.2 for Edinburgh men (without $g(\xi)$ determined by age-group) is useful. This allows an individual's observed mean blood pressure from measurements up to any visit to be easily related to the probability of having an underlying diastolic blood pressure that is sufficiently high to warrant intervention. Thus the decision can be tailored more closely to the needs of the individual concerned: a greater certainty about the need for or value of intervention may be required if, for example, the subject is younger or has a low cholesterol level.

The figure shows clearly the funnel shape obtained for the programme with the curves having as an asymptote $\bar{x}_i = \xi_T$ as i increases. However, note that although the critical values tend, in general, monotonically toward ξ_T , there are values of the probability $p(\xi > \xi_T | \bar{x}_i)$ for which this does not apply. For example, for $p(\xi > \xi_T | \bar{x}_i) = 0.2$ the critical value for \bar{x}_1 is greater than 95, but for the other values of i shown, \bar{x}_i moves progressively below 95 though, for larger i , it does turn back up toward 95. This behaviour is a consequence of the relative strengths of the between-person variance σ^2 and the within-person variance σ_x^2 , and the consequent movement of the conditional distribution for ξ , $h(\xi | \bar{x}_i)$, away from the population distribution,

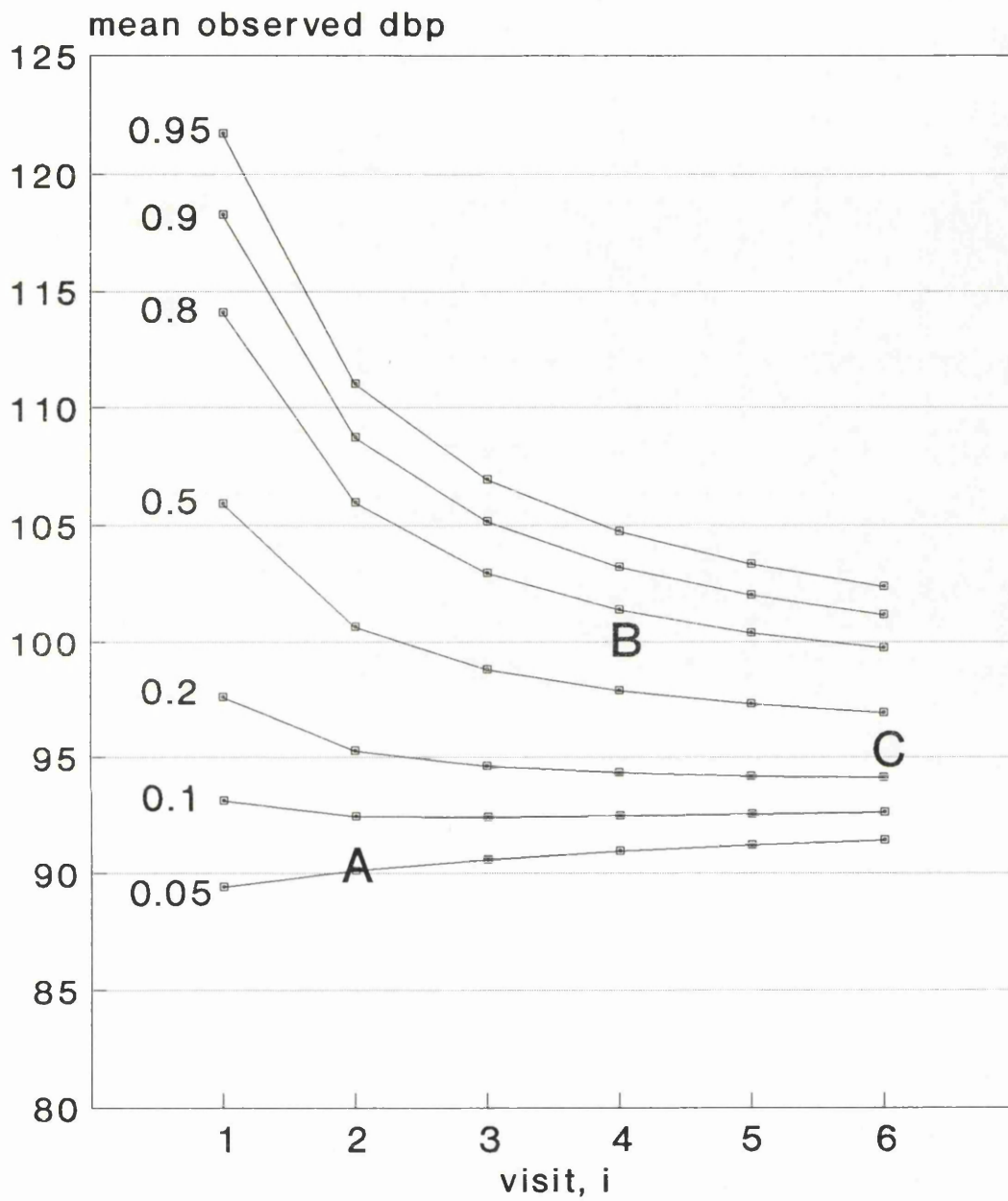


Figure 5.2: Screening the individual for treatment: critical values by visit giving defined probabilities $p(\xi > \xi_T = 95 | \bar{x})$ for all ages in Edinburgh. Points A, B and C show the W.H.O. guideline: screening stops if \bar{x}_2 is below A, or drug treatment starts if \bar{x}_4 is above B or is considered if \bar{x}_6 is above C.

$g(\xi)$, to reflect an individual's own underlying level. This is clear, analytically, in the case with σ_x^2 not dependent on ξ as given by (5.2). Despite the mathematical validity of this behaviour, in practice it would be less appealing to the doctor dealing with the individual patient: for Edinburgh, if p_L is taken as 0.2, then an individual with a blood pressure of 95mmHg at the first screening visit would be followed no further but an individual with a mean pressure of 95mmHg by any subsequent visit would be advised to return for further follow-up.

It is useful to consider how the World Health Organization guidelines relate to this type of screening programme. It is not a strict comparison as the data considered here are phase IV diastolic pressures whereas the guidelines are based on the lower phase V pressure. On figure 5.2 are marked three points (A, B and C), one each at the second, fourth and sixth visits. According to the guidelines, individuals with a mean at their second visit below 95mmHg (point A) would not undergo further screening. This is quite a stringent limit as $p(\xi > \xi_T | \bar{x}_2 = 90) \approx 0.05$. However, the decisions for those that continue screening and for whom pharmacological intervention might be considered ($\bar{x}_4 > 100$ or $\bar{x}_6 > 95$; above points B and C, respectively) have probabilities $p(\xi > \xi_T | \bar{x}_4 = 100) \approx 0.7$ and $p(\xi > \xi_T | \bar{x}_6 = 95) \approx 0.3$. Even allowing for the fact that phase V diastolic pressures will be lower than phase IV ones, the evidence for elevated blood pressure required before intervention is considered is quite weak.

Of course, this presupposes that the expert committee considered that an individual with an underlying level of greater than 95mmHg would benefit from treatment. This seems likely given that they use 95mmHg as the level for considering drug treatment for individuals undergoing repeated follow-up over a long period. However, they also

suggest that “mild hypertension in adults is defined as a diastolic pressure persistently between 90 and 104mmHg” and that intervention is justified in such individuals. Thus, it is worth also considering a critical level, ξ_T , of 90mmHg and looking at the guidelines in this context. Figure 5.3 shows the screening programme for this situation with the three points A, B and C also marked.

In comparing this figure with figure 5.2, the effect of the change in ξ_T is clear and substantial. At the first visit, taking ξ_T as 90mmHg produces a shift of corresponding points for observed blood pressures down by about 10mmHg compared with when ξ_T is 95mmHg. As visits accumulate, the difference between the two figures reduces and tends to a shift in corresponding points of about 5mmHg. Point A shows that the evidence required to stop screening at the second visit is comparatively weak [$p(\xi > \xi_T | \bar{x}_2 = 90) \approx 0.3$]. In contrast, the evidence required for pharmacological intervention (points B and C) is strong. This would certainly be better criteria for a screening programme in terms of the level of evidence required than that shown in figure 5.2. However, if the distribution of underlying phase V diastolic blood pressures in a population is considered as a translation downwards from the phase IV pressures, then the distance of μ from ξ_T in phase V terms is more akin to that used in constructing figure 5.2 for phase IV pressures. Thus the logical basis for the World Health Organization guidelines might be questioned.

Another possible difficulty with screening programmes of this type is the potential for a large number of screening visits to be required to obtain sufficient evidence of an individual’s underlying blood pressure in order to stop screening or to recommend intervention. This is particularly relevant to those individuals with borderline

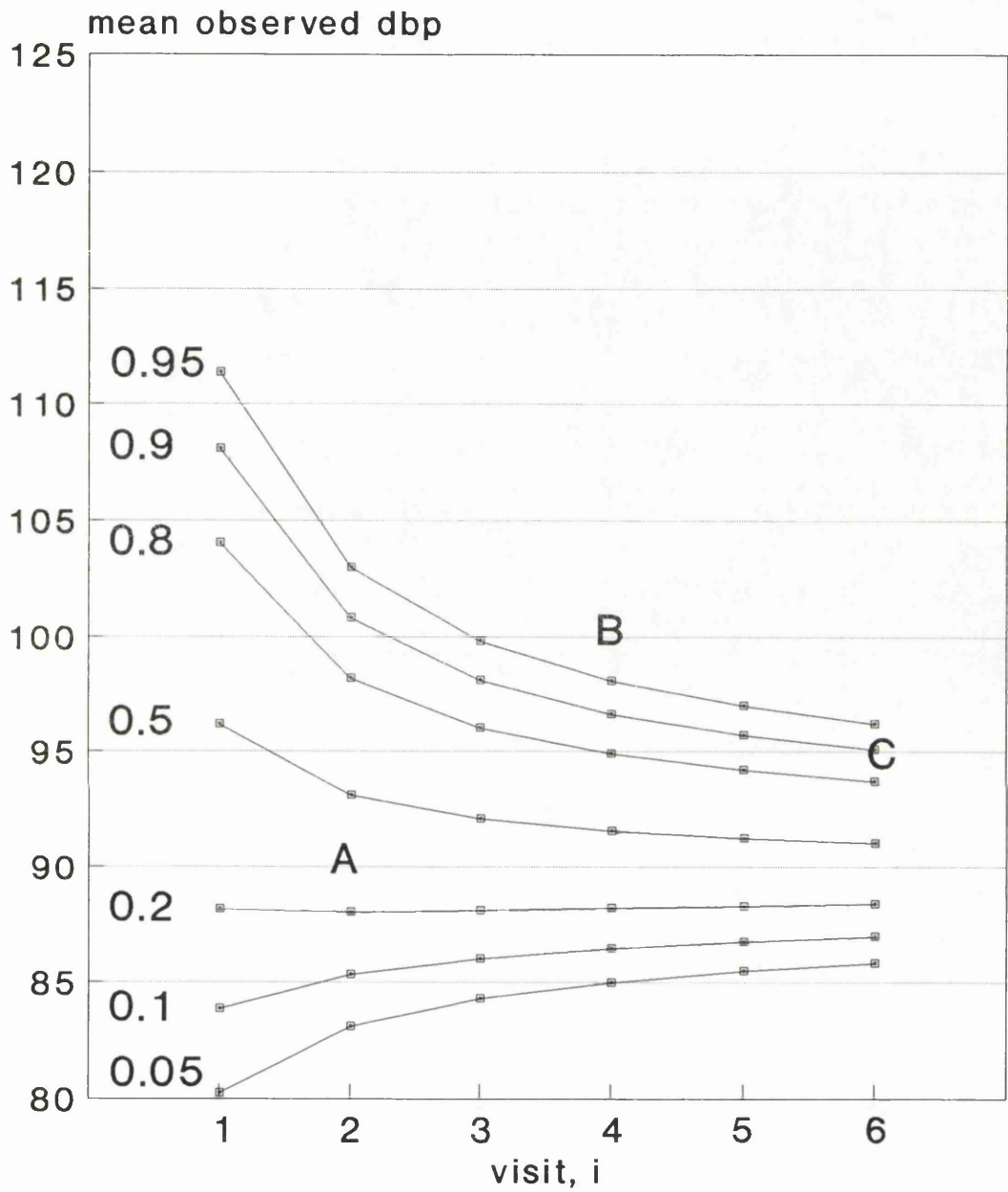


Figure 5.3: Screening the individual for treatment: critical values by visit giving defined probabilities $p(\xi > \xi_T = 90 | \bar{x})$ for all ages in Edinburgh. Points A, B and C show the W.H.O. guideline: screening stops if \bar{x}_2 is below A, or drug treatment starts if \bar{x}_4 is above B or is considered if \bar{x}_6 is above C.

Visit	$\xi < 95$ (n=94683)		$\xi > 95$ (n=5317)		all ξ (n=100 000)	
	stop screening	intervention considered	stop screening	intervention considered	stop screening	intervention considered
1	83628 (88.3%)	4 (0.0%)	1461 (27.5%)	49 (0.9%)	85089 (85.1%)	53 (0.1%)
2	4764 (4.9%)	11 (0.0%)	253 (4.8%)	231 (4.3%)	5017 (5.0%)	242 (0.2%)
3	1789 (1.9%)	21 (0.0%)	102 (1.9%)	265 (5.0%)	1891 (1.9%)	286 (0.3%)
4	903 (1.0%)	19 (0.0%)	57 (1.1%)	275 (5.2%)	960 (1.0%)	294 (0.3%)
>4	3544 (3.7%)		2624 (49.4%)		6168 (6.2%)	

Table 5.4: Simulation of 100 000 Edinburgh men being screened for treatment using $\xi_T=95\text{mmHg}$, $p_U=0.9$ and $p_L=0.1$ according to whether their underlying level ξ is greater or less than 95mmHg. (numbers in brackets are percentages of column totals, n)

underlying levels (that is, close to ξ_T). For example, if p_U is taken as 0.9 and p_L as 0.1, then table 5.4 shows the percentages of the population, in Edinburgh, who would require a measurement at each successive screening visit. These are obtained from a simulation of 100,000 individuals from the distributional form described for Edinburgh in table 3.11 and assume that a blood pressure observed at any screening visit is measured precisely. The most striking feature is the percentage of individuals (6.2% of the population) for whom no decision is reached on completing the fourth visit. For

this group of individuals many further repeat measurements would be required (for just over three-quarters of them, no decision would have been reached on completing the sixth visit). However, the distribution of underlying levels, ξ , in this group is approximately normal with mean 94.3mmHg and standard deviation 3.8mmHg. Thus, the substantial majority will have underlying diastolic pressures in the range 90mmHg to 100mmHg and so a relaxation of the strength of evidence before considering intervention may be acceptable. For instance, individuals for whom $p(\xi > 95 | \bar{x}_4) > 0.5$ might be considered for intervention. Alternatively, the whole group might, as suggested in the World Health Organization guidelines, be encouraged to modify their lifestyle whilst undergoing long-term but much less frequent screening.

If screening is stopped at the fourth visit, then the cost of the screening programme in terms of the average number of screening visits required per person in the population can be evaluated. For this example, this gives 1.32 visits per person. Similarly, with some decision-making criteria at the fourth visit, the cost of intervention in terms of the percentage of the population considered for intervention can be evaluated. For the simulation shown, this could range from 0.9% if none of those with $p(\xi > 95 | \bar{x}_4) < 0.9 = p_U$ are considered, to 7.1% if all of those with $p(\xi > 95 | \bar{x}_4) > 0.1 = p_L$ are considered. Clearly, then, the criteria for considering intervention at the fourth visit has considerable intervention cost implications.

Also shown in table 5.4 is a description of what happens, in this simulation, to those individuals targetted for intervention (that is those with $\xi > 95$ mmHg) and those not (that is those with $\xi < 95$ mmHg). Clear is that very few individuals with $\xi < 95$ mmHg are considered for intervention. In contrast, a substantial percentage of those with

$\xi > 95$ mmHg are not considered and this is particularly attributable to those for whom screening stops at the first visit (for over one quarter of individuals with $\xi > 95$ mmHg). The sensitivity of the screening programme might be defined as the percentage of those with $\xi > 95$ mmHg who are considered for intervention. This can be as low as 15.4% if none of those with $p(\xi > 95 | \bar{x}_4) < 0.9$ are considered or as high as 64.8% if all of those with $p(\xi > 95 | \bar{x}_4) > 0.1$ are considered. Clearly then, sensitivity is very dependent on the action taken for those individuals with $p_L < p(\xi > 95 | \bar{x}_4) < p_U$, but is never high at least when $p_L = 0.1$. In contrast, defining specificity as the percentage of those with $\xi < 95$ mmHg who are not considered for intervention, this will be between 96.2% and 99.9% depending, again, on the action taken at the fourth visit. Note that the sensitivity and specificity are very little changed if the observed pressure at each screening visit is rounded to the nearest even digit as was the practice for the Edinburgh data. Of course, other choices of p_L and p_U can make substantial changes to the sensitivity and specificity. Also, as mentioned earlier, different values of p_L and p_U might be used at different visits: in particular, decreasing p_L at the earlier visits would increase the sensitivity markedly. This issue is not pursued further here but is discussed, briefly, in the context of the population screening method described in section 5.5.

5.4 Screening the individual for expected risk

In the previous section, the decision to consider intervention or to cease screening was based upon an assessment that an individual's underlying blood pressure level is above or below some critical value. Thus, it ignores the extent by which it is above or below this value. Extent can be measured on various scales: notably those of underlying

pressure, of risk of cardiovascular disease and of reduction in risk of cardiovascular disease due to intervention. Clearly, these possibilities represent an increasing progression in the evidence and assumptions needed for models. Conversely, they also represent an increasing association with the long-term aim of a screening programme: to identify high-risk individuals and consequently to attempt to reduce their risk. As the risk scale parallels that of risk reduction if the reduction achieved is proportional to risk, an assumption that might be justified in our context on the basis of evidence from trials of anti-hypertensive treatments (Collins et al, 1990), the use of a risk scale will be pursued further here.

Approaching the problem generally, assume that there is a function $R(\xi)$ relating risk of disease to the underlying level, ξ . Then given observations up to the i th visit, \underline{x}_i , the expected risk for an individual is

$$E(R|\underline{x}_i) = k \int R(\xi) f(\underline{x}_i|\xi) g(\xi) d\xi \quad (5.6)$$

where

$$k = \int f(\underline{x}_i|\xi) g(\xi) d\xi$$

is a normalization constant. By taking

$$\begin{aligned} R(\xi) &= 1 \text{ if } \xi > \xi_T \\ &= 0 \text{ if } \xi < \xi_T \end{aligned}$$

a useful link with the “screening for treatment” approach can be made. In this case

$$E(R|\underline{x}_i) = k \int_{\xi_T}^{\infty} f(\underline{x}_i|\xi) g(\xi) d\xi$$

$$= P(\xi > \xi_T | \underline{x}_i).$$

Thus the “screening for treatment” approach is equivalent to defining a step risk function that switches from 0 to 1 at ξ_T , and then considering intervention if $E(R|\underline{x}_i)$ is sufficiently close to 1 and ceasing screening if it is sufficiently close to 0.

A parallel for any risk function $R(\xi)$ can be drawn:

- (i) if $E(R|\underline{x}_i) > R_U$ then consider intervention
- (ii) if $E(R|\underline{x}_i) < R_L$ then stop screening
- (iii) else if $R_L < E(R|\underline{x}_i) < R_U$ continue screening to visit $(i+1)$.

This has turned the decision making process away from statements concerning the probability that risk is above some cut-off point (which is equivalent to the “screening for treatment” approach for any R monotonic in ξ), to an assessment based on the expected risk. The logic for this might be that R_U is the level above which the potential risk reduction due to intervention definitely outweighs the potential side-effects of intervention. Conversely, below R_L the potential for side-effects is considered to outweigh the possibility for risk-reduction. The window R_L to R_U then reflects uncertainty about the relative balance of risk reduction to side-effects. This contrasts with the “screening for treatment approach” which assumes a defined cut-point, ξ_T , at which the potential for risk reduction and for side-effects is considered to be in balance. However it still ignores the issue of how high, or low, risk might plausibly be.

For the purposes of presenting such a screening programme, two approaches might be adopted. The first is to show expected risk as a ratio to some base risk. The choice of base risk is a practical and not a mathematical decision as different values will only affect all ratios by a constant factor. This base risk could be the threshold risk at which treatment begins to be beneficial (the value ξ_T as used previously) or the expected risk in the population, ignoring blood pressure level. The second is to show risk in absolute terms such as the expected annual (cause-specific) death rate or cause-specific incidence rate. Of course, the second is just a relabelling of the risk scale for the first though not a linear transformation of it. However, where there are multiple risk factors for a disease, absolute risk needs to be quantified taking account of all factors for the individual concerned even though the reduction in risk due to treatment will relate to a single factor.

To illustrate these concepts further, consider again Edinburgh middle-aged men as in this study's cohort. Using the log-linear relationship between risk of cardiovascular mortality and underlying blood pressure described in the previous chapter, then

$$E(R|\bar{x}_i) = k \int \exp(\beta\xi) f(\bar{x}_i|\xi) g(\xi) d\xi$$

where \bar{x}_i is, as before, the mean of blood pressures observed up to and including visit i , and β represents the gradient of risk on a log scale associated with the underlying pressure, ξ . Also the expected risk in the population is

$$E(R) = \int \exp(\beta\xi) g(\xi) d\xi$$

$$= \exp\left(\beta\left(\mu + \frac{1}{2}\beta\sigma^2\right)\right) \text{ when } \xi \sim N(\mu, \sigma^2).$$

Thus by setting the risk ratio $E(R|\bar{x}_i)/E(R)$ equal to given values and solving for \bar{x}_i , screening programmes can be defined. Note that the use of this ratio avoids the need to consider other (independent) risk factors as their effect cancels out. Figure 5.4 shows the curves obtained for various values of the risk ratio from 1 to 2.5, taking β as 0.0605 as derived for Edinburgh in the previous chapter. Again, a decision rule for screening can be determined by choosing two curves; considering intervention if an individual's mean observed blood pressure lies above the upper curve, stopping screening if it lies below the lower, otherwise continuing further observation. For example, if $R_U=2.25$ and $R_L=1.5$ and taking the three individuals described in section 5.3.2, then the first individual whose pressure at the first visit was 110mmHg would be recommended for intervention at that visit because of a sufficiently high expected risk, whilst the second would cease screening after the first visit because his measurement of 90mmHg indicates a sufficiently low expected risk. However, the third individual with mean pressures of 106, 98, 98 and 99.5mmHg has intermediate expected risks at the first, second, third and fourth visits, respectively and so screening would continue.

There is a fundamental difference between these curves and those of the "screening for treatment" approach: these do not have a single value (ξ_T) as their asymptote; instead the asymptotes are, for a risk ratio RR,

$$\beta^{-1}\left[\ln(RR)+\left(\beta\mu+\frac{1}{2}\beta^2\sigma^2\right)\right] \text{ as } i\rightarrow\infty.$$

This is a consequence of using a risk function, R , that is continuous in ξ rather than a step-function. It also explains why the curves are flatter than those for "screening for treatment" and are monotonic for all ratios of practical interest. However, this

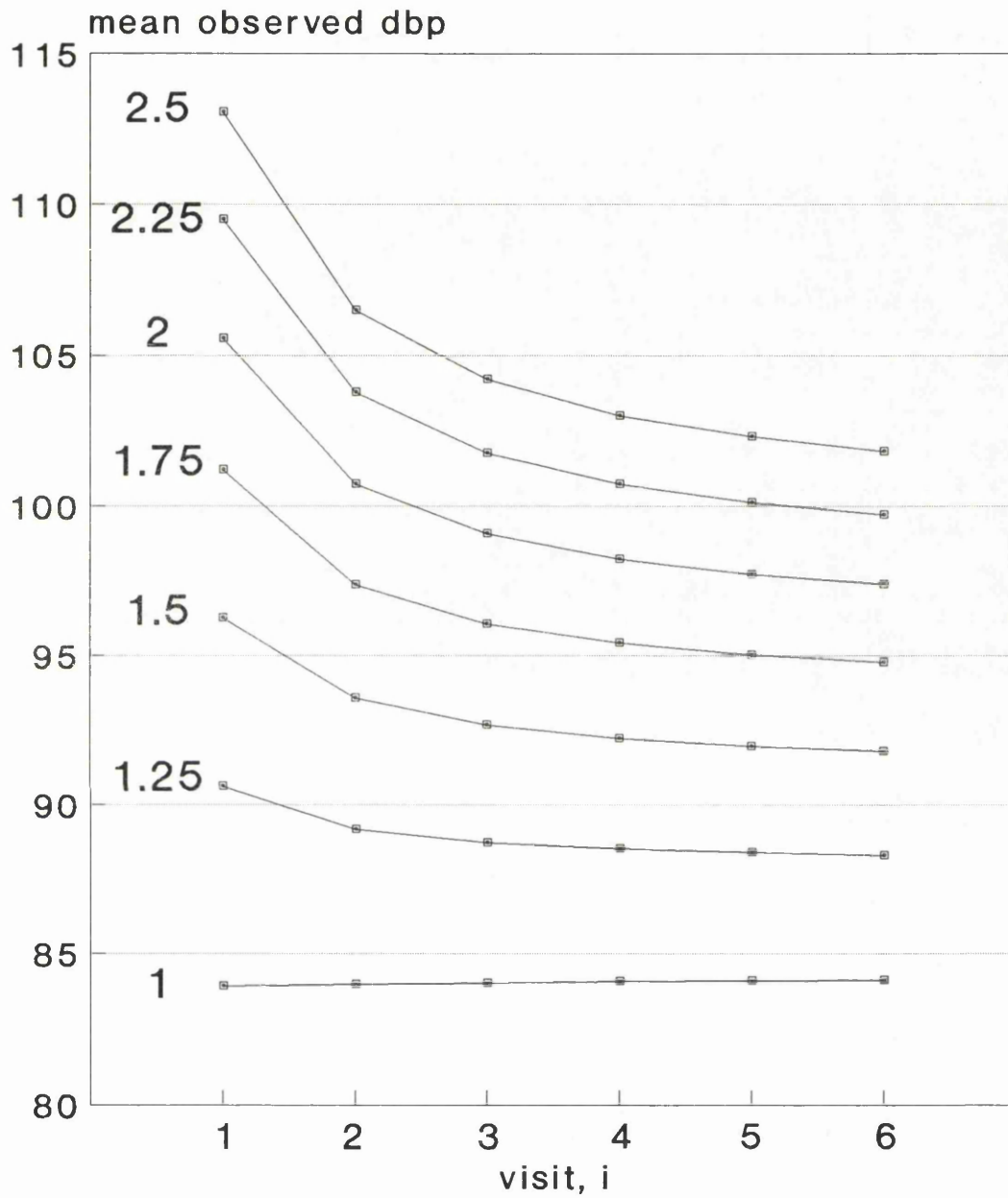


Figure 5.4: Screening the individual for risk: critical values by visit giving defined ratios of individual risk to the average in the population, $E(R|\bar{x})/E(R)$, for all ages in Edinburgh.

approach does not solve the practical problem of the substantial numbers for whom no decision to consider intervention or cease screening has been reached after several visits.

Although the results of the previous chapter were by no means conclusive about the evidence for different risk relationships between cardiovascular mortality and underlying pressure, ξ , it is useful to consider how sensitive such a screening programme is to variation in the risk coefficient, β . To illustrate this, the estimated coefficients determined in the previous chapter for each of the three towns ($\beta=0.0605$, 0.0494 and 0.0921 for Edinburgh, Budapest and Prague, respectively) together with the respective within-town estimates for the blood pressure distribution parameters can be used. Values of the mean \bar{x}_i giving risk ratios of between 1 and 2.5 (as used in formulating figure 5.4) are given in table 5.5. Immediately clear is the effect of increasing the risk gradient associated with a change in underlying diastolic pressure. For example, comparing the values for Budapest (smaller β) and Prague (larger β), the smaller the risk gradient is, the higher are the mean observed pressures giving a defined risk ratio: after four visits, for an expected risk ratio of 2, the mean pressure in Budapest must be 104.4mmHg compared with 96.7mmHg in Prague. The distinction between Edinburgh and Prague is not so clear because of the larger difference in the relative sizes of the within- and between-person variances in these two towns. Thus the development of a common screening programme of this sort for use in different locations is difficult unless there is strong evidence for a common risk coefficient, β , as well as similar values of the variances, σ^2 and σ_x^2 .

Despite this, the “screening for risk” approach may be useful to the doctor faced with

$E(R \bar{x}_i)/E(R)$	Screening visit					
	1	2	3	4	5	6
<u>Edinburgh</u>						
2.5	113.1	106.5	104.2	103.0	102.3	101.8
2.25	109.5	103.8	101.8	100.7	100.1	99.7
2	105.6	100.7	99.1	98.2	97.7	97.4
1.75	101.2	97.4	96.1	95.4	95.0	94.8
1.5	96.3	93.6	92.7	92.2	92.0	91.8
1.25	90.6	89.2	88.7	88.5	88.4	88.3
1	83.9	84.0	84.0	84.1	84.1	84.1

<u>Budapest</u>						
2.5	130.2	118.2	113.7	111.3	109.8	108.8
2.25	124.5	114.0	110.1	108.0	106.7	105.9
2	118.3	109.4	106.1	104.4	103.4	102.7
1.75	111.5	104.3	101.7	100.4	99.6	99.1
1.5	103.8	98.7	96.9	95.9	95.4	95.0
1.25	95.0	92.2	91.3	90.8	90.6	90.4
1	84.8	84.7	84.8	84.8	84.8	84.8

<u>Prague</u>						
2.5	112.2	104.6	101.8	100.4	99.5	98.9
2.25	109.0	102.3	99.9	98.6	97.9	97.4
2	105.5	99.8	97.7	96.7	96.0	95.6
1.75	101.6	97.0	95.3	94.5	94.0	93.7
1.5	97.2	93.8	92.6	92.0	91.7	91.5
1.25	92.0	90.1	89.5	89.2	89.0	88.9
1	85.8	85.7	85.8	85.8	85.8	85.8

Table 5.5: Screening the individual for expected risk: critical values for L_i and U_i for visit i for all ages by town

the individual patient. The decision to intervene and attempt to lower blood pressure concerns the balance of potential side effects, inconvenience etc. with the expected reduction in absolute risk for that patient. For instance a patient in Edinburgh with a mean blood pressure, after four visits, of 95mmHg has a relative risk of just under 1.75 compared with the average population risk. If treatment might be expected to reduce risk by 10 or 20% then relative risk on treatment would be between about 1.4 and 1.6. In absolute terms this change might not be sufficient to outweigh the possible side effects of treatment if the individual is younger, non-smoking and with a low cholesterol level. However for the older individual with a high cholesterol it may be a significant personal risk reduction. Thus relating the observed mean blood pressure to risk without making any formal decision rule for screening might be an adequate aid for the doctor and patient jointly to reach a decision about starting intervention or stopping screening. Of course, this does not negate the possibility for showing on the risk chart, the risk curve below which it is generally accepted that the risk reduction from treatment does not outweigh the side effects for most individuals or does not justify the cost of intervention.

This concept of providing the doctor with risk information without formally declaring a decision rule can be taken a step further. Using the survival model developed in the previous chapter and estimating the base survivor function (that is, the survival curve for a hypothetical individual with covariate values of zero) nonparametrically (Kalbfleish and Prentice, 1980, chapter 4), values of ξ can be transformed to probabilities of dying from cardiovascular causes within, say, 5 years whilst also taking account of an individual's levels of other risk factors including age. In figure 5.5, the conditional distribution $p(\xi|\bar{x}_4)$ is shown for the third individual described in section

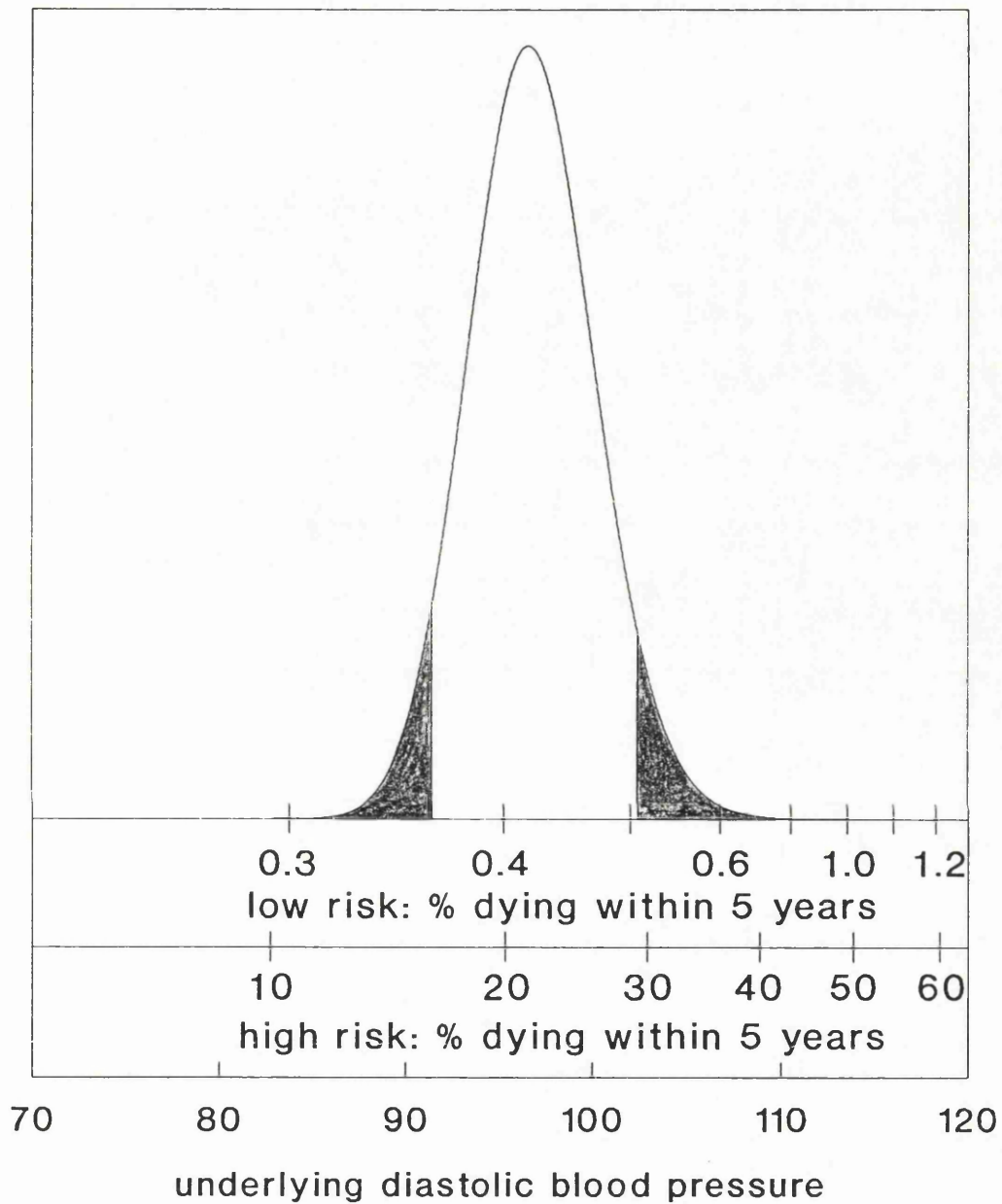


Figure 5.5: Conditional distribution for underlying diastolic pressure, ξ , for an Edinburgh man after four pressure measurements of 106, 90, 98 and 104mmHg, showing also the associated absolute risk scales if the man has other risk factors as low risk (age 40, never smoked, low cholesterol 200mg/dl) or high risk (age 60, >20 cigs/day and high cholesterol 350md/dl)

5.3.2 (with blood pressures at the four visits of 106, 90, 98 and 104mmHg, respectively). Above the scale for ξ are two risk scales showing the estimated risk of dying of cardiovascular causes within 5 years: the first where the individual is 40 years of age, a non-smoker with a low observed cholesterol (200mg/dl at one visit) and the second where he is 60, a heavy smoker (>20 cigarettes per day) with a high observed cholesterol (350mg/dl). Very clear from these plots is the substantial difference in absolute risk depending on the levels of factors other than blood pressure: the expected absolute risks are 0.43% and 23% chances of dying within five years for such low and high risk men, respectively. The shaded tail areas shown each cover a probability of 5% so that the interval in between covers probability 90%. This interval can be used to determine whether there is a significant possibility that risk is sufficiently raised to warrant intervention. If not, then there would be no justification for further screening. Conversely, if the interval contains risk levels which are either sufficiently raised, or sufficiently low, then further screening would be justified to reduce the standard deviation of the distribution and so help to confirm or refute either of these possibilities.

5.5 Screening a population

In the previous sections the screening rules developed have had two characteristics. First they have been open-ended in that some individuals will be recalled repeatedly without a decision being reached. This is clearly undesirable because of the implications for screening costs and the need to maintain the cooperation of the subjects concerned. Second, because the same expected risk or probability of pressure being elevated above the critical level, ξ_T , is used at all visits, a constraint on the

critical values, L_i and U_i , is imposed. This produces a problem of misclassification notably of subjects for whom intervention is recommended or screening is stopped after one or two screening visits. This can be expressed in terms of patient numbers being considered for intervention when their underlying level do not justify it, or vice versa. Alternatively, a more useful way of expressing it is to quantify the proportion of those events that might be preventable by intervention but are not prevented because of misclassification: this then also involves the extent of misclassified risk. The approach considered in this section addresses both of these issues: constraints are defined by limits on the proportion of the population that can be considered for intervention and by the effort (or cost) of the screening programme itself. Subject to such constraints it is then desirable to choose the L_i and U_i so that the proportion of the population selected for intervention is that proportion that is most likely to benefit. Thus the problem is one of constrained optimization. In the next section this screening model is defined more fully in terms of objective and constraint functions that reflect the risk function, $R(\xi)$, and the variability of diastolic pressures both within and between individuals. Then we illustrate its use by considering the screening of an Edinburgh male population similar to that found in the clofibrate trial.

5.5.1 Constrained optimization model for population screening

As before, we consider a screening programme defined by critical values L_i and U_i for visit i at which the following decisions, determined by the mean of the accumulated observations \bar{x}_i , are made:

- (i) if $\bar{x}_i > U_i$ then consider intervention
- (ii) if $\bar{x}_i < L_i$ then stop screening

(iii) else if $L_i < \bar{x}_i < U_i$ then continue screening to visit (i+1).

Note that the programme can be limited to a maximum number of visits by choosing $L_i = U_i$ at the final allowable visit.

We shall restrict our attention to risk-related objective functions though the method follows through in the same manner if functions quantifying the reduction in risk due to intervention are used. Labelling by I the set of observed measurements in individuals considered for intervention, so that

$$I = \{\underline{x} \mid \bar{x}_i > U_i \text{ and } L_j < \bar{x}_j < U_j \text{ for } i \geq 1 \text{ and for all } j < i\},$$

then the problem is to maximize the expected risk for individuals considered for intervention = $E(R \mid \underline{x} \in I)$ subject to constraints on the L_i and U_i .

The constraints considered here are those of “cost” and can be defined in terms of the expected number of visits required per individual entering the screening programme (the screening “costs”) and the proportion of those screened that are considered for intervention (the intervention “costs”). Taking the screening cost constraint first, this gives

$$1 + \sum_{i=1}^{\infty} p(L_i < \bar{x}_i < U_i \mid L_j < \bar{x}_j < U_j \text{ for all } j < i) \leq M, \text{ say,}$$

where M is a constant to be given. In this expression the one reflects all individuals having at least one measurement and each term in the summation reflects the

proportion going forward from visit i to visit $(i+1)$. Next, the intervention cost constraint is

$$\sum_{i=1}^{\infty} p(\bar{x}_i > U_i \text{ and } L_j < \bar{x}_j < U_j \text{ for all } j < i) \leq P, \text{ say,}$$

where P is a second constant to be given. Here each term in the summation reflects the proportion of the population who are considered for intervention as a result of observations up to visit i . Writing the objective function in a similar manner, the problem is to maximize

$$E \left[R(\xi) | \bar{x}_i \in \bigcup_{i=1}^{\infty} \{ \bar{x}_i | \bar{x}_i > U_i \text{ and } L_j < \bar{x}_j < U_j \text{ for all } j < i \} \right]$$

subject to these constraints. In practice, it may also be useful to limit the number of screening visits that can be completed by any individual, say to a maximum of N . This necessitates the additional constraint that $L_N = U_N$, so that a decision is reached for all individuals.

To evaluate the expectation in the objective function, the probability density function for ξ conditional on $\underline{x} \in I$ is needed. This function is

$$\frac{\int_{\underline{x} \in I} p(\underline{x}, \xi) d\underline{x}}{\int_{\xi} \int_{\underline{x} \in I} p(\underline{x}, \xi) d\underline{x} d\xi}$$

where p is the joint density function for \underline{x} and ξ . Given that observations at successive

visits may be considered independent (if the intervening time interval is sufficiently long) then

$$p(\underline{x}, \xi) = g(\xi) \prod_{j=1}^i f(x_j | \xi).$$

Thus the required density function is

$$\frac{\sum_{i=1}^N \int_{\underline{x}_i \in I_i} g(\xi) \prod_{j=1}^i f(x_j | \xi) dx_j}{\sum_{i=1}^N \int_{\xi} \int_{\underline{x}_i \in I_i} g(\xi) \prod_{j=1}^i f(x_j | \xi) dx_j d\xi} \quad (5.7)$$

where I_i is the set considered for intervention at the i th visit ($i \geq 1$):

$$I_i = \{\underline{x}_i \mid \bar{x}_i > U_i \text{ and } L_j < \bar{x}_j < U_j \text{ for all } j < i\}.$$

Note that the denominator of (5.7) is exactly the proportion considered for intervention and so is the constant, P . Thus in determining the maximum of the objective function, the denominator in (5.7) can be ignored.

By defining the set proceeding from visit i to visit $(i+1)$ for further screening by S_i where

$$S_i = \{\underline{x}_i \mid L_j < \bar{x}_j < U_j \text{ for all } j \leq i\},$$

then the problem can be written as

$$\text{maximize (with respect to } \underline{L} \text{ and } \underline{U}\text{): } \sum_{i=1}^N \int_{\xi} \int_{\mathbf{x}_i \in I_i} R(\xi)g(\xi) \prod_{j=1}^i f(x_j|\xi) dx_j d\xi$$

$$\text{subject to: (i) } L_N = U_N$$

$$\text{(ii) } 1 + \sum_{i=1}^{N-1} \int_{\xi} \int_{\mathbf{x}_i \in S_i} g(\xi) \prod_{j=1}^i f(x_j|\xi) dx_j d\xi = M$$

$$\text{(iii) } \sum_{i=1}^N \int_{\xi} \int_{\mathbf{x}_i \in I_i} g(\xi) \prod_{j=1}^i f(x_j|\xi) dx_j d\xi = P \quad (5.8)$$

where N is the maximum number of screening visits allowed ($N > 1$), M is the average number of screening visits per person and P is the proportion of the population to be considered for intervention.

The difficulty in solving this problem is that integrals of dimension $(N+1)$ need to be evaluated and these will not, in general, be amenable to analytic simplification. In addition, the dimension $(2N-1)$ of the parameter space makes this a substantial computational problem when requiring a numerical solution. Hence it is only likely to be soluble for small values of N . It is important also to note that the integral in the objective function may not be bounded and so the problem is not always soluble. In particular, if R is exponential in ξ , as is often assumed, and if $g(\xi)$ is markedly skew, then convergence of the integral to a finite value will not be obtained. This would be so if also ξ is log normally distributed. Thus in modelling the distribution of underlying covariate values, $g(\xi)$, and the risk function, $R(\xi)$, particular care is needed in assessing their fit to the data. Where uncertainty persists it may be useful to use an empirical distribution function for g rather than a poor mathematical approximation.

5.5.2 Maximizing expected risk: population screening of blood pressures in Edinburgh

Here interest is restricted to screening an Edinburgh population of men of ages similar to that entered into the clofibrate trial. The results of chapters 3 and 4 (tables 3.9 and 4.7) are used to provide the necessary parameter estimates for functions required by the optimization problem:

$$f(x_j|\xi) = N\left[\xi, \sigma_W^2[1+c(\xi-\mu)]^2\right].$$

$$g(\xi) = N(\mu, \sigma^2)$$

$$R(\xi) = \exp(\beta\xi)$$

The example considered takes as fixed the intervention cost constraint such that

$$P = \int_{\xi=95}^{\infty} g(\xi)d\xi.$$

Thus if the underlying diastolic pressure, ξ , could be determined, then the intervention group would include all individuals with levels above 95mmHg so that $P=5.27\%$. Note that this choice is an arbitrary one made for illustration. If the volume of risk in a cohort is defined as its average risk times the cohort size, then the volume of risk in the intervention group as a proportion of the volume of risk in the whole population, R_p , is a useful summary measure of the risk that will be intervened upon. If the underlying blood pressures were known, a useful upper bound on the volume of risk in the intervention group is

$$\int_{\xi=95}^{\infty} R(\xi)g(\xi)d\xi = 0.1253R_p.$$

Hence, maximally, 12.53% of future cardiovascular deaths could be expected in a cohort of 5.27% of this Edinburgh population. Therefore any optimal screening programme developed for a given screening cost, M , might be related to this maximum. Note that a lower bound can also be defined by considering the expected risk in the intervention group formed by taking a random sample of size, P , from the population. This, for the Edinburgh situation, would include 5.27% of future cardiovascular deaths and represents $5.27/12.53=42\%$ of the maximum that any screening programme with an intervention group of this size could achieve. Other screening programmes simple in concept can also be formulated for comparison. In particular, programmes in which every member of the population has a fixed number of blood pressure measurements are useful in showing how much very intensive screening effort can achieve. These are equivalent to the general screening programme with the the critical values L_i set to $-\infty$ and U_i set to $+\infty$ for $i < N$ and with $L_N = U_N$ such that the intervention group is of the chosen size.

Returning to the optimization problem, the example considered allows a maximum of four screening visits and considers three values of the screening cost parameter: $M=1.1, 1.25$ and 1.5 visits per person. A quadratic programming method using analytically derived derivatives was used to solve the problem. However, the objective function changes very slowly near its maximum as the parameters \underline{L} and \underline{U} are altered within the constraints and so the optimal solution found by the optimization routine can be mildly sensitive to the numerical integration procedure used, though the value of the objective function at the maximum is well determined. We return to the accuracy of the critical values below.

Table 5.6 shows a comparison of the optimal programmes with the other screening programmes outlined above. The two measures for comparing the payoff from different programmes are given: first, the volume of risk in the intervention group as a percentage of the total population risk (equivalently, the percentage of cardiovascular deaths that might be expected in the group) and, second, the volume of risk in the intervention group as a percentage of the volume achievable in the intervention group if the underlying pressure, ξ , were known. Clear from these measures is that taking a single blood pressure measurement for all individuals in the population and selecting those individuals with the highest values for intervention produces close to a doubling of the identified risk compared with taking a random sample from the population of the desired size. Thus the main gain is obtained by having some minimal knowledge of individuals' blood pressure. Further increasing the number of observations per person produces much smaller gains: going from one to four measurements for each person screened increases the volume of risk in the intervention group from 81.5% to 92.7% of the maximum possible. Similar orders of change are also found for other sizes of the intervention group, P . If $P=16.4\%$ (corresponding to the percentage with $\xi > 90\text{mmHg}$) then the expected risk rises from 86.8% to 95.1% in moving from one to four measurements for each person, and taking $P=1.2\%$ (corresponding to $\xi > 100\text{mmHg}$) gives a change from 76.1% to 89.9%.

Similar sized gains can be made by considering the optimal sequential rules: returning to table 5.6, averaging 1.5 observations per person with a maximum of four observations gives a marginally lower payoff of 91.9% compared with the rule where everyone has four measurements. Note that increasing the maximum number of visits allowed whilst holding the average number per person constant would produce further,

Programme	Critical values	Risk in intervention group as % of	
		total risk in population	risk in intervention group if ξ known
Random sample	None	5.27	42.1
<hr/>			
<u>All individuals have M measurements</u>			
M=1	$U_1=99.5^*$	10.21	81.5
M=2	$U_2=97.5^*$	11.00	87.8
M=3	$U_3=96.7^*$	11.38	90.9
M=4	$U_4=96.3^*$	11.61	92.7
<hr/>			
<u>Optimal sequential rules (max 4 visits)</u>			
M=1.1	$\underline{U}=(102.4, 99.0, 96.4, 94.4)$ $\underline{L}=(95.5, 95.1, 94.7)$	10.88	86.8
M=1.25	$\underline{U}=(106.5, 101.6, 98.1, 95.1)$ $\underline{L}=(92.5, 93.3, 94.2)$	11.28	90.0
M=1.5	$\underline{U}=(111.3, 105.2, 100.1, 95.8)$ $\underline{L}=(88.1, 90.7, 93.3)$	11.52	91.9
<hr/>			
ξ known	intervention if $\xi > 95$	12.53	100

* For $j < i$, values of $L_j = -\infty$ and $U_j = \infty$ are used where i is the total number of measurements allowed in the programme.

Table 5.6: Comparison of screening rules applied to a population of Edinburgh men (similar to that used in the clofibrate trial) when the intervention group has a fixed size=5.27% of the population= $p(\xi > 95)$

but small, gains in payoff. Thus considerable savings in screening costs can be made by using such sequential rules. Despite this, the gain in payoff in terms of volume of risk identified for intervention is small and, if intervention costs (either financial or related to the side-effects of treatment) are low, then there are likely to be greater gains in payoff for fixed total screening plus intervention costs by accepting a larger intervention group on the basis of a single measurement per person than a smaller intervention group found by repeat screening of some of the population.

Also shown in table 5.6 are the critical values, \underline{L} and \underline{U} , found for the sequential screening rules. For each value of M , these are one set of values (correct to one decimal place) which give the maximal risk identified as a percentage of the volume of risk in the whole population (correct to two decimal places). Other sets of values (with critical values varying from those shown by 0.1 or 0.2mmHg) satisfying these criteria for accuracy may be found. This reflects the flatness of the objective function close to its maximum. However the considerable computational problems that need to be overcome in order to increase accuracy and hence find a solution closer to optimal is not warranted when the accuracy to which blood pressure is measured in practice can lead to more significant departures from the optimal payoff. This is illustrated by the following simulation.

Consider a sample of 100 000 Edinburgh men with underlying pressure and within-person variability as described in table 3.9. The men are entered into a screening programme with critical values for considering intervention or the cessation of screening as given for the sequential rules in table 5.6. Table 5.7 shows the results from two situations: first, when measurements are taken exactly and, second, when

	<u>Measured exactly</u>			<u>Measured to nearest even digit</u>		
	\hat{M}	$\hat{P}(\%)$	$\hat{R}/R_P(\%)$	\hat{M}	$\hat{P}(\%)$	$\hat{R}/R_P(\%)$
M=1.1	1.101	5.34	10.89	1.115	5.22	10.76
M=1.25	1.250	5.30	11.27	1.235	5.15	11.01
M=1.5	1.502	5.34	11.59	1.474	5.37	11.63

Column definitions: \hat{M} =average number of visits per person in simulation
 \hat{P} =percentage of men considered for intervention (the programmes were designed with $P=5.27\%$)
 \hat{R}/R_P =volume of risk in the intervention group as a percentage of volume of risk in the whole population (optimal values found were 10.88%, 11.28% and 11.52% for $M=1.1$, 1.25 and 1.5, respectively)

Table 5.7: Results of a simulation of 100 000 Edinburgh men entering screening programmes with the sequential rules given in table 5.6, when their blood pressures are measured exactly or are rounded to the nearest even digit

they are rounded to the nearest even digit as occurred, approximately, in practice in Edinburgh in the clofibrate trial. For each of these situations, the average number of visits per person (\hat{M}), the size of the intervention group (\hat{P}) and the volume of risk in the intervention group as a percentage of the volume of risk in the whole population, (\hat{R}/R_P)%, were evaluated.

Considering first, the situation with blood pressure measured exactly, the values of \hat{M} are within one standard error of that aimed for (the standard error varies from 0.001

when $M=1.1$ to 0.003 when $M=1.5$). Similarly, the values of \hat{P} are within one standard error of the 5.27% used in defining these rules (the standard error is 0.07%). Note however that the rounding of the critical values to one decimal point can affect the value of \hat{P} by a small amount. For example, when $M=1.5$, if U_4 is changed to 95.9 from 95.8 , then \hat{P} drops to 5.23% from 5.34% without altering \hat{M} . Secondly, when the measurements are rounded to the nearest even digit, simulation of the same cohort passing through the screening programme tends to give larger departures from the values expected for each of the screening cost, intervention cost and payoff parameters. Thus, in practice, the accuracy of blood pressure measurement is likely to be a greater determinant of the actual costs and payoff for a particular screening programme than the departures caused by computational limitations in seeking the optimal payoff for constrained costs. Of course, the ideal would be to undertake the optimization allowing also for the accuracy of measurement used but this imposes an additional degree of computational complexity.

Figure 5.6 illustrates diagrammatically the critical values (bold lines) for the optimal screening rules given in table 5.6. This shows clearly the effect of increasing the screening costs allowed by the programme: a wider aperture of the funnel shape for higher costs. These rules have been overlaid onto the expected risk ratio curves formed for the “screening for risk” approach for individuals (dashed lines) identical to those given in figure 5.4. This shows, as should be expected, that U_1, U_2, U_3 and U_4 form a decreasing sequence in terms of the expected ratio and L_1, L_2, L_3 and L_4 an increasing sequence, reflecting the use of this expected risk ratio in the objective function of the optimization problem. Particularly noticeable from the overlaid plots is the way in which the optimal rules make cessation of screening difficult at the first and second

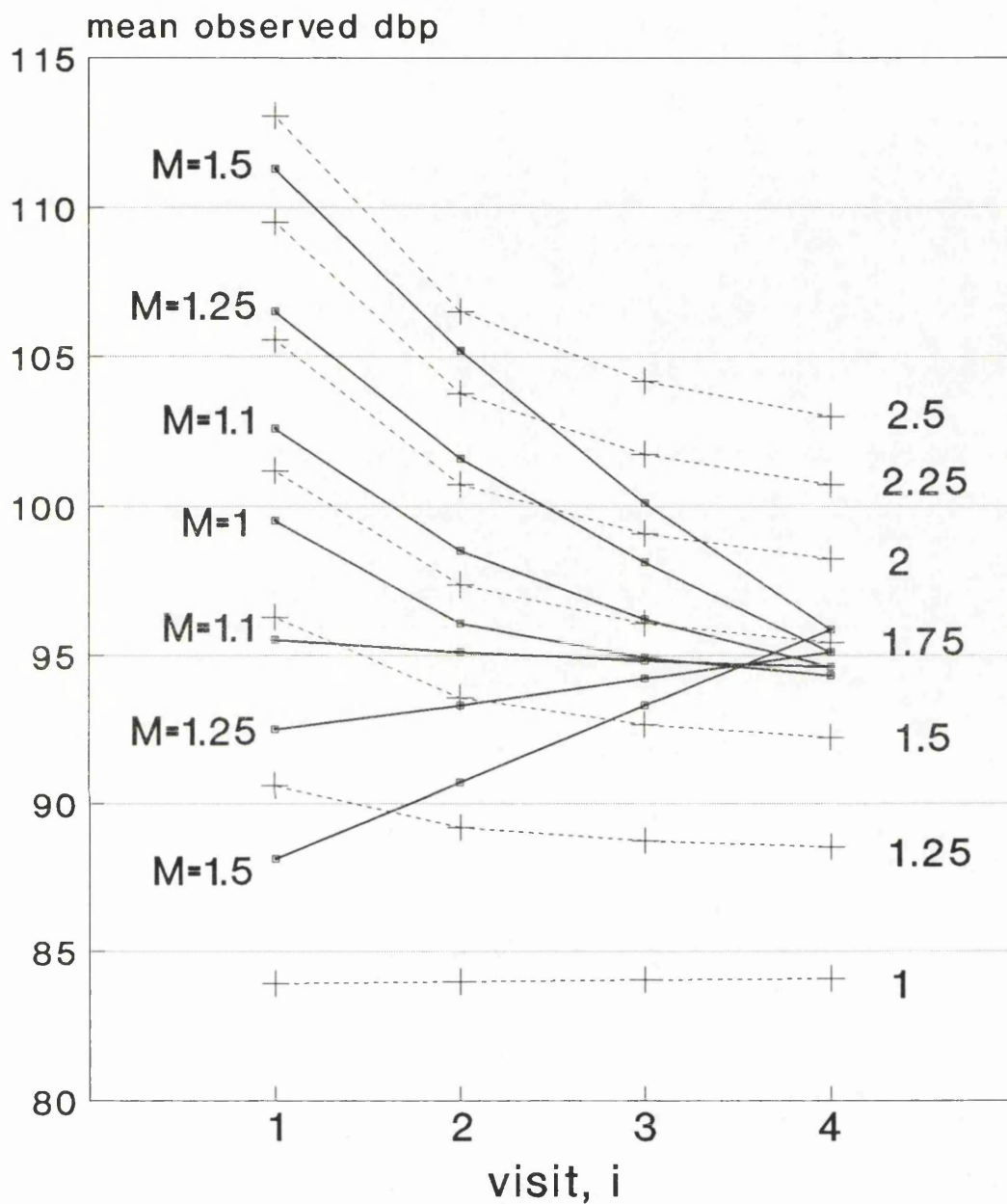


Figure 5.6: Optimal screening rules for an average of M visits per person with 5.27% of the population considered for intervention and a maximum of 4 visits (bold lines). Dashed line show the expected risk as a ratio to the average risk in the population.

visits. As has been discussed earlier in the context of the “screening for treatment” rules, this is where the greatest scope for misclassification arose in the rules presented in sections 5.3 and 5.4; this demonstrates how the optimal rules make their gains in risk classification.

Also shown in figure 5.6 is a screening rule labelled “M=1”. This rule represents the limit as the number of visits per person tends to unity. It is formed by considering the critical value for intervention when each individual has a single diastolic pressure measurement (equal to 99.5mmHg, from table 5.6). Individuals with such a pressure have an expected risk ratio of 1.66 times the average population risk. Thus any optimal rule allowing an average number of visits per person slightly bigger than one must produce a funnel shape that departs slightly from the line shown linking the points with an expected risk ratio of 1.66 at each visit. Thus for any rule allowing a maximum of four measurements, this gives a lower bound for U_4 of 94.3mmHg. An upper bound for U_4 is given by the critical value for considering intervention when each individual in the population has the maximal four screening visits: equal to 96.3mmHg (from table 5.6). Thus for a fixed value of P, U_4 varies very little with the value of M (a useful aid in choosing a starting point for the optimization routine).

We have already seen the relatively small impact that these sequential rules have on the magnitude of identified risk in the intervention group. Thus the gains in population (public health) terms are small by using repeat measurements and so the criteria for choosing the three constraining parameters (N, P and M) might be chosen on the basis of ethical obligations toward the individual subject to any limitation on financial cost. As it would not be desirable to have individuals selected for intervention

for whom the expected risk reduction does not exceed the expected risk of side-effects, then P might be defined in terms of the minimum expected risk for an individual selected into the intervention group. Since the sequence of the U_i 's is decreasing as the visit number, i , increases, this imposes a minimum on the value of U_N . Given that U_N varies very little as M changes (for fixed P), then to a reasonable approximation $E(R/R_P | \bar{x}_N = U_N)$ defines P .

The maximum number of visits, N , can be determined by the range of underlying risk ratios (or equivalently of underlying blood pressures) which are consistent, with reasonable certainty, with an observed mean blood pressure at the final visit equal to U_N . For example, the probability density function $f(\xi | \bar{x}_4 = U_4)$ when $U_4 = 95.1 \text{ mmHg}$ (corresponding to the optimal sequential rule in table 5.6 with $M=1.25$) is shown in figure 5.7. The shaded tails each represent a probability of 0.05 and so we can be 90% certain that an individual with such an observed mean blood pressure has a risk ratio lying between 1.26 and 2.41 (or, equivalently, that their underlying blood pressure lies between 88.0 and 98.8). If this is a satisfactory range for intervention then four visits would be an appropriate value for N . However, if it were felt unethical to treat someone with a risk ratio as low as 1.26, then a higher maximum number of visits would be needed. Of course, if the range is felt to be too narrow then a lower number of visits might be appropriate.

The same concept is useful in choosing the effort put into screening, that is the average number of visits per person, M . For example, figure 5.8 shows the 90% probability interval for ξ given observed mean blood pressures equal to the critical values \underline{L} and \underline{U} given in table 5.6 for the case when $M=1.25$ visits per person. If the intervals

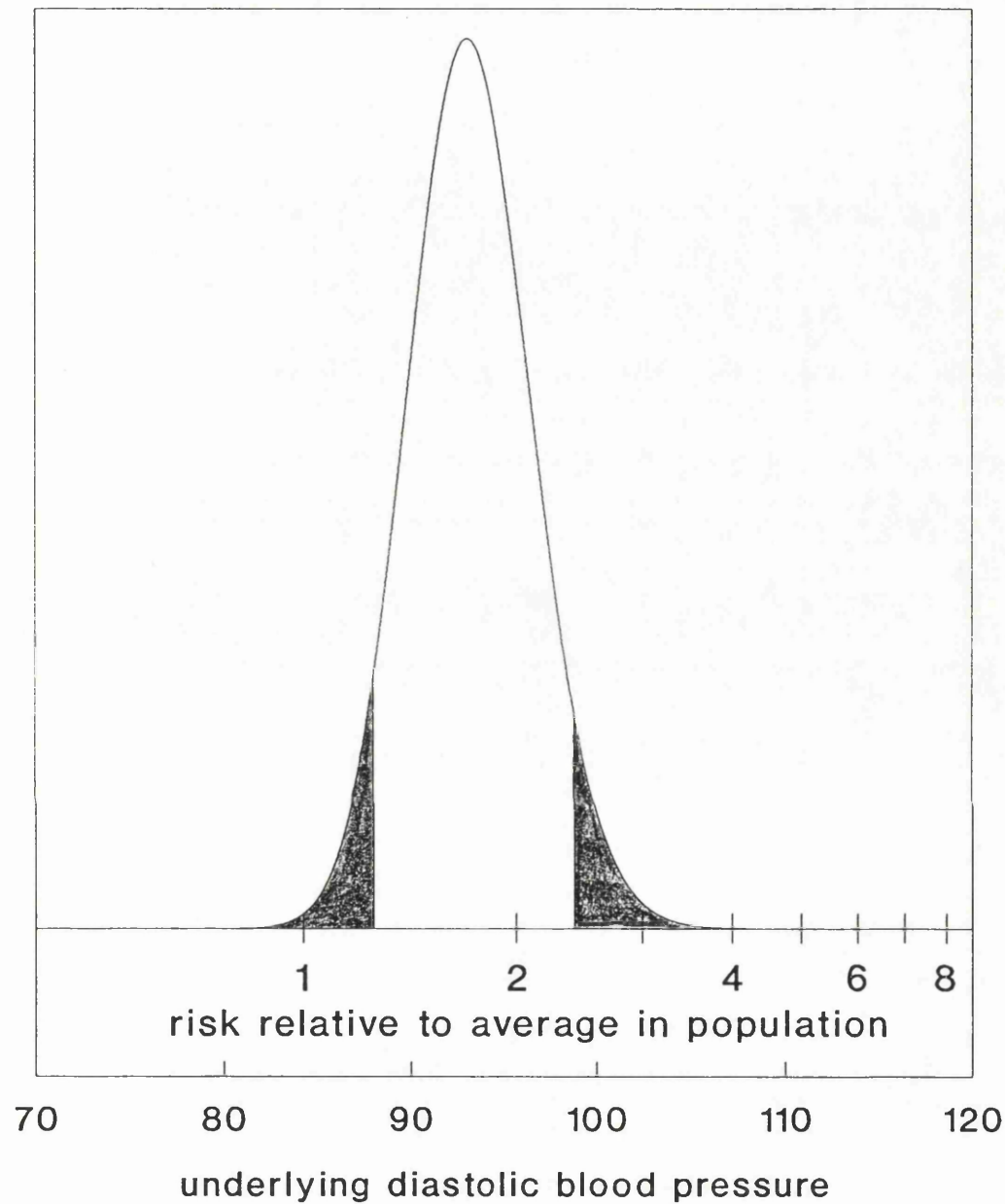


Figure 5.7: Conditional probability density function for ξ and the associated risk relative to the average in the population given an observed mean diastolic pressure of 95.1mmHg after 4 visits (corresponding to the critical value for intervention in the optimal rule with a maximum of 4 visits with $M=1.25$ and $P=5.27\%$).

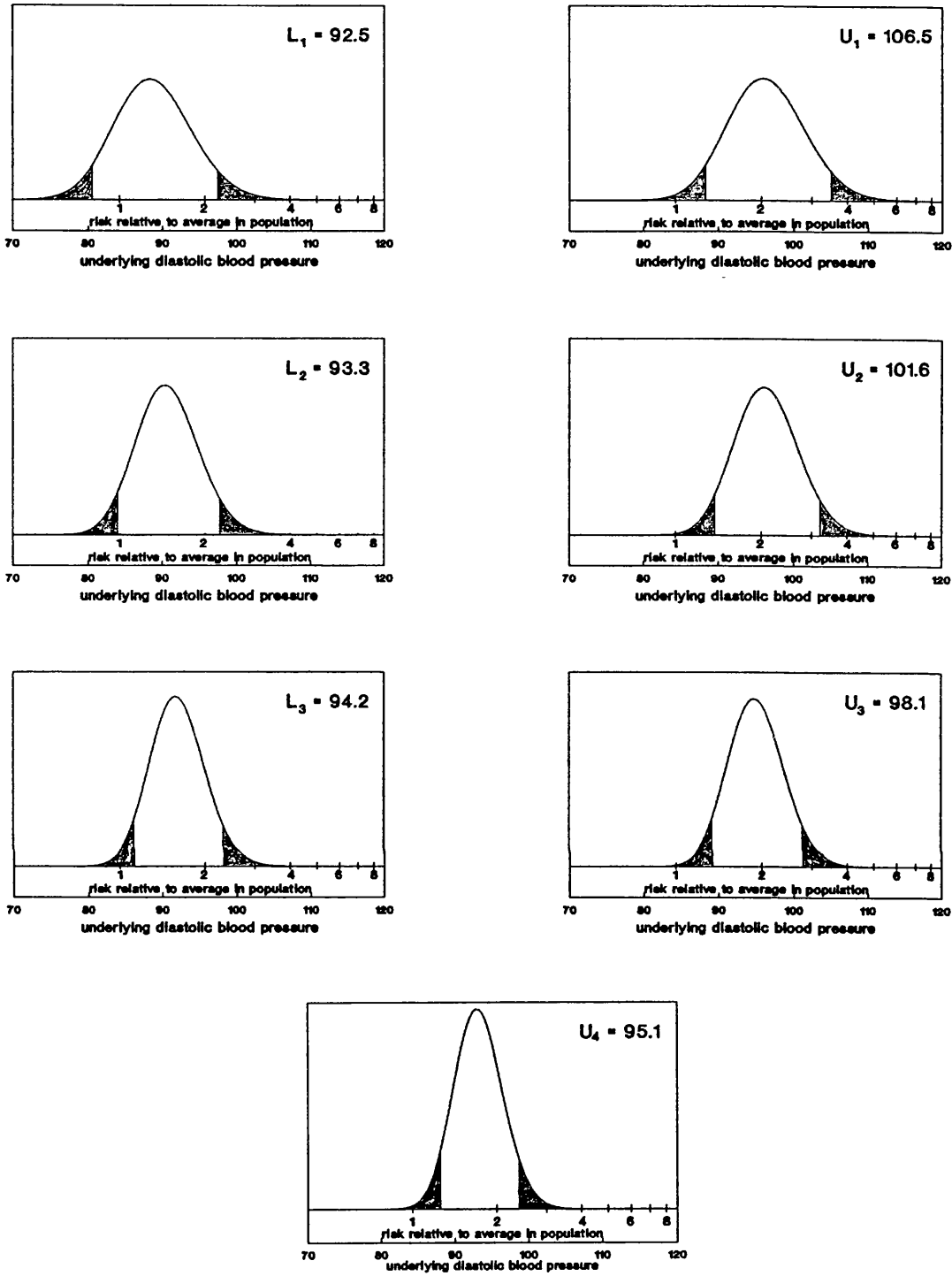


Figure 5.8: Conditional probability density functions for ξ given observed mean diastolic pressures equal to the critical values for considering intervention (U) or stopping screening (L) for the optimal rule with a maximum of 4 visits and with $M=1.25$ and $P=5.27\%$.

conditional on $\bar{x}_i=U_i$, for some visit i , include values that are considered unethical for intervention, or else those intervals conditional on $\bar{x}_i=L_i$, for some visit i , include values that should indicate intervention, then increasing M would be necessary so that the relevant intervals do not show such characteristics. Alternatively, it may be possible to decrease M , so saving screening costs, whilst maintaining adequate evidence of elevated risk before intervention or of low risk before stopping screening. The major difficulty with this approach is that the desire to answer to individual ethics may produce a rule with an excessive financial cost and/or excessive inconvenience for the individual of repeat screening. In this case, a reasonable balance needs to be reached. In practice, this is likely to require, in particular, a decrease in the proportion of the population considered for intervention so that the funnel pattern for the critical values is shifted upward to higher mean observed blood pressures.

5.6 Summary of main results with discussion

In screening for cardiovascular disease, we are dealing with a risk relationship that is multifactorial. However, interventions (particularly pharmacological ones) are generally concerned with producing a change in a single risk factor (though, for instance, dietary change might affect both cholesterol and blood pressure levels). This makes the development of a screening programme for selective intervention (i.e. not for the whole population) difficult because the decision to treat should be based upon the balance between the expected benefits in terms of the reduction in absolute risk and the expected disbenefits such as cost and side-effects. The approaches presented have concentrated on identifying individuals with elevated underlying blood pressure or, similarly, elevated risk of dying from cardiovascular causes attributable to raised blood

pressure. Formal rules can be defined by choosing critical levels for the mean diastolic pressure over measurements taken at several screening visits so that a subject with a mean above the upper limit would be recommended for intervention, a subject with a mean below the lower limit would cease screening and a subject with an intermediate mean would be encouraged to return for further screening.

Two approaches to screening individual subjects that would be particularly suitable for more ad hoc screening situations such as with casual screening by general practitioners have been considered. The first, the “screening for treatment” approach, considers a threshold for underlying pressure, ξ_T , above which it is felt that intervention is desirable and below which it is felt to be undesirable. Intervention is then recommended when it can be established that, given a sequence of repeat measurements, a subject’s underlying pressure, ξ , is greater than ξ_T with some high probability. Conversely, screening ceases when ξ is shown to be below ξ_T with a high probability. The determination of critical values for the mean pressure over a series of visits that satisfy these criteria was described, using Bayes Theorem, in section 5.3. The main drawbacks of this approach are twofold. First, it fails to quantify how high or low the underlying pressure, and hence the risk of cardiovascular mortality, is and, second, it assumes that a level ξ_T can be determined well.

In contrast, the second approach, “screening for risk”, determines, on the basis of the measurements obtained, the expected risk for a subject. This is then compared with two risk levels: an upper one above which intervention is considered advantageous because the potential benefits in terms of risk reduction would outweigh the potential side-effects, and a lower one below which the reverse is felt to apply when screening

would be stopped. This approach allows uncertainty in the knowledge about the level of risk required before treatment is considered but does not consider the degree of uncertainty about how high, or low, an individual's risk might plausibly be.

In effect what is required is a mixture of these two approaches. Without formalizing rules, a graphical display which portrays the required information was suggested at the end of section 5.4 (and illustrated in figure 5.5). This enables a doctor to see visually the expected risk together with a distribution showing the range of risk and underlying diastolic pressure that the subject concerned might plausibly have. The balance of expected risk reduction and potential side-effects can then be made together with an assessment as to whether the decision to intervene, or not, should be influenced by the uncertainty in the risk determination. If the latter is so, then further screening would be warranted to improve the precision of risk estimation. The other distinct advantage of this graphical display is that it can be designed to show absolute risk reflecting also levels of other risk factors which are clearly important in deciding whether intervention is warranted.

For both of these "individual" approaches to screening, the scope for misclassification occurred mainly through the termination of screening at the earlier visits (notably at either of the first two) with the conclusion that risk was adequately low not to warrant intervention. This is addressed in the "population" approach which is concerned with identifying that group in the population of a defined size with the greatest expected event rate for whom intervention would be recommended, subject to a constraint on the amount of screening (expressed in terms of the average number of screening visits per person allowed). This approach also had the desirable feature, not catered for in

the individual approaches, of limiting the maximum number of screening visits. From such a public health perspective, the value of repeated measurement of diastolic blood pressure for cardiovascular disease might be questioned. After a single measurement, further measurements add little in identifying a higher-risk group for intervention. For illustration, in a cohort of middle-aged men similar to that studied from Edinburgh, an example considered for a group size representing 5.27% of the population and chosen on the basis of having the highest diastolic pressures at a single screening visit would be expected to have, in number, 81.5% of the events that would be expected had a group been chosen on the basis of the highest underlying pressures. Increases in the volume of screening then produce only small increases in the expected event rate above this level. This primarily reflects the relatively low rate of change in risk with underlying level and so, in population risk terms, the relatively small penalty paid for misclassification. Nevertheless, the optimal rules presented illustrate how careful selection of individuals for further screening can achieve similar gains in terms of identified risk to that from mass screening where all individuals have repeated measurements.

A particular advantage of considering this approach should be in its use to define screening rules that have desirable properties in the public health context (that is closed screening rules which are close to optimal in terms of population risk identification within given cost constraints) but which also give satisfactory risk assessment for individual patients so that they are not subjected to unnecessary, costly and potentially harmful treatment. This can be achieved by using the somewhat heuristic approach suggested for developing a screening plan which employs the conceptual framework of the population approach (and so is optimal in that context)

but where the constraints are determined by the precision of risk assessment required for individual patients before intervention is considered or screening is stopped. This should give a good balance of public health considerations and individual ethics.

In deciding whether intervention is or is not appropriate, some quantification of the level at which adverse effects or cost outweigh the benefits needs to be made. When there is a single (or dominant) risk factor this is simplified as it can be interpreted as the underlying level, ξ_T , already mentioned above which intervention is indicated because, on balance, it is considered beneficial or cost-effective, and below which it is not. In this chapter, values of 95mmHg and 90mmHg have been used for illustration purposes. However, determining the appropriate value of ξ_T for use in practice is not simple. Although the efficacy of interventions may be established in clinical trials, it is not always clear from the entry criteria at what level efficacy outweighs adverse effects particularly as the criteria are generally defined in terms of observed risk factor levels and not underlying levels. For instance the overview of trials of anti-hypertensive treatments (Collins et al, 1990) showed significant risk reductions in stroke and coronary heart disease in trials with an entry (observed) diastolic pressure in the range of about 90 to 110mmHg. However, without knowledge of the distribution of blood pressures from which the trial entrants were drawn, it is difficult to derive the mean underlying diastolic level for the entrants to the trials and so to use any efficacy result found to make a statement about the level ξ_T . Thus, it would be advantageous for clinical trials to include information about the underlying mean blood pressures of individuals randomized. Perhaps the easiest way to do this would be to randomize patients and then to have a short period (say, a month or so) prior to treatment during which several repeat measurements are made. Then the mean of the

measurements for each individual would be a precise estimate of their underlying level.

In moving from a disease with a single risk factor to one with several, the population approach can still be applied directly to risks and interventions associated with a single factor provided that levels of each factor are (approximately) uncorrelated and the effect of intervention on one factor does not affect the levels of other factors. However in screening the individual, multiplicative risk factors means that total absolute risk needs to be determined so that the balance of risk reduction and cost or adverse effects can be properly assessed. The “screening for treatment” approach is then less applicable as it is more difficult to define the value, ξ_T , as this should depend on the levels of other risk factors in the individual being screened: for example, a high cholesterol level in an older person might suggest a lower critical value for intervention on blood pressure than for a younger person with a low cholesterol level. In contrast, the “screening for risk” approach can be easily extended and the package of interventions chosen that produces the greatest benefit for least risk of adverse effects or lowest cost. To facilitate this, the effect of any intervention needs to be understood for all levels of the risk factor that it affects. In particular, it would be useful to know whether intervention does produce a proportionate reduction in risk regardless of the level of the risk factor or whether the reduction is attenuated at lower levels. Also relevant in this context is whether there is advantage in lowering risk factor levels in younger subjects to reduce long-term risk despite the risk of morbidity in the short-term being small. However, long-term clinical trials that answer this question are rarely undertaken and so the issue can only be addressed by extrapolating results from shorter-term trials and considering the mechanisms that relate the risk factor to morbidity.

An important consideration in applying any of the screening methods in practice is their robustness to the parameter estimates adopted for the risk profiles and risk factor distributions both within- and between-persons. In this respect the population approach is likely to be particularly sensitive: small errors in their estimation could lead to suboptimal solutions and to marked departures from the costs expected. The approach is also sensitive to the functional forms chosen, particularly to the relationship between the shape of the risk function with the underlying level, ξ , and the distribution of ξ in the population. For example, there may be little difference in the departures from normality of the distribution for ξ or for $\log\xi$, but if the risk function is log-linear and a normal distribution is chosen for ξ then the problem is soluble, but if a normal distribution is chosen for $\log\xi$, then it is not soluble because the volume of risk (that is, number at risk times level of risk) is an unbounded function as ξ increases.

For the individual approaches, the importance of the accuracy of estimates for the parameters defining the population distribution for ξ reduces very quickly as the number of repeat measurements accumulates. This is appealing particularly as long-term intervention is unlikely, in practice, to be recommended without a few repeat measurements. Thus the problem of uncertainty about population parameters really concerns those who stop screening after one or two measurements; for these situations it might be appropriate to be more cautious before stopping screening (as is suggested by the population approach). Thus a lower level than a mean of 90mmHg after two visits, as suggested by the W.H.O. guidelines for hypertension screening (Participants of the W.H.O./I.S.H. meeting, 1986) would be useful. Similarly, in casual screening

situations without formal decision rules, doctors should be encouraged to look at the mean of a set of measurements even if taken over several years rather than considering a single low measurement at a current visit as indicative of no requirement for intervention. In practice, with both the development of an emphasis on screening for prevention and with increases in computing facilities in general practice, it would be relatively simple for software to be developed that provides both individual risk profiles updated at each visit whilst also including individual information in a database that is able to update population parameters tailored to local areas. In addition the development of multi-factorial risk scores with screening based on these rather than single factors would reduce the need to measure repeatedly those factors, such as blood pressure and cholesterol, that have substantial components of within-person variability because the risk score will readily give an adequate assessment of the range within which an individual's risk is likely to be.

6. Further perspectives

In this chapter the developments of this study are summarized in the context both of other published work and also of further research that would be useful in developing a statistical framework for screening. The problem considered essentially concerns risk factors that are continuous variables measured “with error” and the consequent decision process that leads to an individual being classified as at “high-risk” of developing a disease and so being considered for intervention to lower that risk. Four particular issues can be identified that contribute to the screening and intervention process. The first concerns the assessment of risk in any individual. The second considers the information that needs to come from clinical trials to aid the decision about when a treatment is likely to be effective. The third is the method by which screening leads to a decision as to whether risk is, or is not, sufficiently elevated to warrant intervention. The fourth concerns the role of monitoring response, particularly in the short-term, following the start of intervention to assess its effectiveness. These issues are considered in turn in the next four sections with a specific emphasis on the context considered in this study, screening diastolic pressure to assess risk of mortality from cardiovascular disease, and also on the role of multiple risk factors in this setting.

6.1 Risk assessment

The risk of cardiovascular disease appears to rise continuously with diastolic blood pressure level (MacMahon et al, 1990) and so lowering pressure in any individual is likely to reduce their risk. However, when interventions used to achieve this can be potentially harmful (as with drugs), costly and/or inconvenient, then these

“disbenefits” need to be balanced against the benefits, that is the expected reduction in risk of developing the disease, in considering the value of intervention. This is necessarily somewhat subjective but does require a reliable assessment of absolute risk for an individual.

Risk scoring schemes to identify high-risk individuals have been developed for cardiovascular disease (for example, Shaper et al, 1986). These can be used as a basis for determining absolute risk. Typically, however, they only require information on risk factors at one screening visit and therefore assume that there is no dependence of risk on past levels of these factors. In chapter 4, this assumption was investigated for the association between risk and changes in observed diastolic pressure in a sequence of four annual measurements. No evidence was found that trends or variability about the observed trend gave additional predictive power when added to a model including mean level. This confirms the findings of Hofman et al (1980) for trends and variability over a longer period. Thus it would be reasonable to conclude that current underlying level of diastolic pressure is the important determinant of risk in middle-aged men at least in the medium term (say, for predictions for up to about 10 years).

Underlying level cannot be measured directly. In chapter 3, a model for variability was established describing annual measurements as being normally distributed about underlying level with standard deviation which increases with level. This model might be expected to apply also to measurements closer together in time, even when separated by as little as a week as then serial correlation is negligible (Wilson and Hebel, 1988). The size of the within-person variability is of a similar order to that between-person, and so the consequent lack of precision in predicting individual risk

needs to be recognized: a risk assessment based on one measurement is substantially less precise than one based on, say, four measurements at separate visits because the latter is a more accurate measure of underlying level. Note that this cannot be achieved by repeated measurements at a single visit since the within-visit variance is much smaller than the between-visit variance (Armitage, Fox et al, 1966). In addition, the component of variance attributable to rounding practices is only a small component of the within-person variance so that improvements in measuring practice, though worthwhile, will only have a small impact on precision of estimation.

Any application of a risk score using information from several visits needs to reflect the problem of regression dilution: the risk gradient associated with a single measurement is smaller than the risk gradient associated with the mean of two (or more) measurements. This is illustrated well by table 4.8: a 10mmHg difference in single diastolic pressure measurements is associated with a 40% increase in risk, whereas a 10mmHg difference in the mean of measurements from four visits is associated with a much larger, 69%, increase in risk. Thus applying a risk score derived on a sample using single pressure measurements to a screening situation where multiple measurements are available will lead to an underestimation of an individual's true risk and hence to an over-emphasis on other risk factors.

An approach which avoids this problem and also indicates the precision of the risk assessment is obtained by using the observed pressures to derive a probability distribution for the underlying level using Bayes Theorem, and then to use a risk score defined in terms of underlying level rather than observed level. Obtaining the risk gradient with underlying level can be achieved by determining the relationship with

measured values (done naively using the likelihood equations for the proportional hazards model even though this model is only valid for the underlying risk factor and not for the observed factor levels), and then applying an adjustment for regression dilution arising from the within-person variability about underlying level. For log-linear proportional hazards models, the size of the adjustment needed to give an unbiased estimator in larger samples was investigated in chapter 4. When within- and between-subject variation are both normally distributed with variances σ_{δ}^2 and σ_{ξ}^2 , respectively, and if there is a high degree of censorship in the cohort being studied, then adjustment by a factor of $(1 + \sigma_{\delta}^2 / \sigma_{\xi}^2)$ is appropriate (as used in simple linear regression). Prentice (1982) also suggested this approximation when the mortality rate is low (that is, close to 0). This will be so in many epidemiological studies of chronic diseases in the general population, as found in this study. When this is not the case, the adjustment needed may be substantially larger: its size increases as the underlying risk gradient increases and as the level of censorship decreases. However, an important result established analytically in chapter 4 shows that the size of the adjustment needed does not depend on the underlying hazard function. This is a particularly useful result as one of the dominant advantages of these models is that the hazard function does not need to be specified in order to derive estimates of relative risk.

On a practical front, wider recognition of the effect of regression dilution on estimates coming from epidemiological studies is needed. In particular, comparison of effects found in different studies can be confounded by differential effects of regression dilution attributable to varied within- and between-person variances across studies. Thus it would be useful for studies to obtain repeat measurements on risk factors that vary with time on part or all of their cohort so that relationships between underlying levels

and risk can be assessed.

In moving to multiple risk factors, the effect of variability in one or more factors on the estimates for risk associations has not been considered. In the context of cardiovascular risk, this is a pertinent question as both cholesterol and systolic pressure contribute to improved risk assessment and both vary within-individuals over time. Armstrong et al (1989) consider this issue for logistic regression models. The adjustment obtained depends on the within-person covariance matrix for the set of risk factors considered. Given the similarity of results obtained from logistic regression and proportional hazards models when the mortality rate is low, then similar results might be expected for the adjustment needed when using the latter model in this setting. Thus the adjustment for diastolic pressure might not be expected to depend much on the variability of cholesterol within-individuals (because their correlation is low) but the adjustments for diastolic and systolic pressures, when fitted in the same model, would have a mutual dependence because of their more marked correlation. In settings where the degree of censorship is not high, further work would be desirable to assess the dependence in proportional hazards models of the magnitude of the factors for adjusting for regression dilution in the presence of correlated covariates.

6.2 When is treatment effective?

As has been mentioned, when considering the individual patient, the effect of treatment in terms of reducing absolute risk needs to be known. However, this is determined by the levels of a set of risk factors whereas clinical trials tend to describe treatment effects in terms of relative risk reductions (though adverse effects are more often

quantified in absolute terms).

A shift of emphasis to show trial results also in terms of absolute risk would be useful. Ideally, for trials of anti-hypertensive treatments, this might show the absolute risk reduction in a two-way classification: by baseline underlying diastolic pressure and by absolute risk of mortality (or morbidity) within a defined period from baseline. This would then enable the balance of risk reduction and side effects to be assessed more readily. Of course, this table might be summarized more readily in terms of a model for relative risk particularly if the effect of treatment is independent, on this scale, either of levels of diastolic pressure or of levels of other risk factors, or if any dependence can be conveniently summarized. Such a model would then be useful in deriving absolute risks if, for instance, they are calculated within computer software developed for general practice screening. On a practical front, however, few trials will have sufficient power to determine such a dependence of treatment effect on baseline risk factor levels and so it would be a useful issue to pursue more fully in overviews.

The other important consideration for clinical trials is the need to relate risk reduction to underlying level and not to observed level so that predictions of the expected risk can be made for any individual. Two methods of achieving this are possible. The first is to measure pressure, in each subject entered into the trial, repeatedly over several weeks (after randomization but before treatment), so that the mean pressure obtained is a precise estimate of underlying level. This has the advantage of allowing the effect of treatment to be assessed in subgroups defined by underlying level but also has the disadvantage of involving considerable screening effort. The alternative approach is to derive the distribution of underlying pressures from the observed pressures using a

model that relates observed and underlying pressures. Such a model should be specific to the population sampled for the trial but this would require repeat measurements on a large cohort in order to estimate its parameters. A good approximation might be obtained by conditioning on results describing within-person variability obtained in other studies, such as obtained here in chapter 3, where the measurement error caused by rounding is similar. Then the parameters for the distribution of underlying levels in the population sampled can be estimated by likelihood methods using data from all subjects who are screened once for entry into the trial. This would require no additional screening effort than that usually undertaken for a trial though it does not give a precise determination of individual underlying levels, only of average levels in the cohort entered into the trial.

6.3 Screening rules

Each of the approaches to screening considered in chapter 5 has focussed on identifying individuals with high components of risk attributable to elevated diastolic pressure. At any of a sequence of screening visits, the decision processes used give critical levels of mean observed diastolic pressure up to that visit (or equivalently, expected risk of cardiovascular mortality attributable to elevated diastolic pressure) that lead to screening being stopped either with or without the recommendation to consider intervention. This dependence on a single risk factor is not desirable when there are multiple risk factors as the decision should be based on total absolute risk. In particular, the “screening for treatment” rule which developed work by Rosner and Polk (1983) presumes that there is a level of underlying diastolic pressure above which intervention is indicated and below which it is not. This is not the case when absolute

risk at that critical level may be large or small depending on the levels of other risk factors. Despite this, the concept of requiring convincing evidence that a subject has elevated pressure before considering intervention can be usefully carried over to decisions based on absolute risk. It would also be relevant if it were established that intervention was not at all effective below a particular level of diastolic pressure.

The “population approach” to screening focussed on identifying a group within the population of a given size with the greatest expected event rate attributable to elevated diastolic pressure, subject to a constraint on screening costs defined as the average number of repeat measurements allowed per person. If the primary “disbenefits” are not side-effects to individual patients but are, for example, financial costs to the health service (as for professional guidance on behavioural interventions), then individual misclassification is less important. In this case, if the correlation between diastolic pressure and other risk factors is small and intervening on diastolic pressure has little impact on other factors, then considering just diastolic pressure in defining a screening rule is appropriate using the methods put forward.

The philosophy of the approach also extends easily to scores incorporating multiple risk factors, though the concept of screening costs becomes more complex. For screening for risk of cardiovascular disease, repeat screening is essentially concerned with blood pressure and cholesterol measurements and so relative costs of obtaining these need to be considered. Fullest flexibility would then be obtained by choosing, for each individual, that combination of blood pressure and/or cholesterol measurements which would most aid risk classification in the population.

This is a very dynamic problem and so simplification would be useful. This can be achieved by ignoring the cost of the first screening visit (where information on all risk factors would be collected) because all individuals attend then, and then considering each successive visit as forming a screening package. This would consist of diastolic and systolic measurements, together with a cholesterol measurement either on a clinic machine giving an instant result or the result obtained from a sample taken at a previous visit but sent to a laboratory for assessment. In this way, costs do not vary from visit to visit (after the first) and so the method of analysis presented in section 5.5 follows through straightforwardly. It also provides maximal information for comparing the benefits of intervention on either blood pressure or cholesterol or both. However, it does presume that a single elevated risk factor is not important when the absolute risk is sufficiently low. Also, it may be possible to have an elevated absolute risk but without any one risk factor being markedly raised or having a level at which the benefits of treatment have not been well established in clinical trials. Such issues may make the application of optimal screening rules in their purest sense difficult but the methods will still be useful in formulating guidelines (as distinct from rules) for screening. In particular such guidelines could be used in the “screening for risk” approach where a distribution of belief is obtained for an individual’s absolute risk which can also be related to scales showing the associated uncertainty in estimating the underlying level of each risk factor.

6.4 Monitoring the effects of intervention

Although this issue has not been specifically addressed in this study, it does have some relevance to the information collected during screening and so deserves brief discussion

in the context of results from this study. Once intervention is commenced there is an ethical obligation to assess its performance: if there is a possibility of treatment side-effects or an intervention is costly, then it is inappropriate to maintain a patient on a treatment which is not giving the desired effect.

In chapter 3, trends in underlying diastolic pressure were found to be sufficiently small in magnitude so that any treatment effect should show as a reduction in underlying diastolic pressure within a reasonable time of its commencement. However, the large within-person variability (a standard deviation of the order of 10mmHg at diastolic pressure levels likely to be treated) means that substantial trends may be observed even if treatment has no effect. Thus, in monitoring a treatment's effect, it is important to establish a reference level prior to the start of treatment with which later measurements can be compared.

Given the desirability of multiple pre-treatment measurements, screening rules might be formulated that recognize this. Effectively, this leads to one-sided rules where the decision after any screening visit concerns only the possibility of stopping with the conclusion that risk is sufficiently low not to warrant intervention. For other subjects screening would continue until some pre-defined maximum number of visits by which time good precision in risk assessment would be obtained to decide the appropriateness of commencing intervention at all. Although in the population context, this will increase screening costs, the impact will be relatively small because the intervention group size generally considered is likely to be small compared with the population size.

6.6 Conclusion

This study has developed a statistical framework for screening rules when a risk factor varies in time about some underlying or “usual” level. Essentially the rules are applied sequentially setting out criteria, as each repeat measurement of the risk factor is completed, as to whether adequate precision has been achieved to make a recommendation to start intervention because risk is high or to stop screening because risk is adequately low. Their development requires statistical models both for risk factor variability and for the relationship between underlying level and risk. The latter can be derived from the relationship between observed level and risk by adjusting for regression dilution. From consideration of these rules, guidelines for use in practice can be developed which would give both good risk classification for individual patients in reaching decisions but which also display desirable characteristics in a public health context in terms of identifying a high risk group for intervention subject to constraints on screening costs. These would represent an advance on the guidelines currently employed, for example for diastolic pressure, which have little statistical basis. The extension of the screening models to multifactorial risk scores is both desirable and conceptually simple. To achieve this some further work is indicated particularly, on the statistical front, in the handling of regression dilution for correlated risk factors in survival models and, in clinical trials, in relating treatment effects to underlying risk factor levels.

References

- Amery, A., Birkenhager, W., Brixko, R., Bulpitt, C., Clement, D., Deruyttere, M., de Schaepdryver, A., Dollery, C., Fagard, R., Forette, F., Forte, J., Hamdy, R., Henry, J.F., Joossens, J.V., Leonetti, G., Lund-Johnson, P., O'Malley, K., Petrie, J.C., Strasser, T., Tuomilehto, J. and Williams, B. (1986) 'Efficacy of antihypertensive drug treatment according to age, sex, blood pressure, and previous cardiovascular disease in patients over the age of 60' The Lancet ii, 589-592
- Armitage, P., Fox, W., Rose, G.A. and Tinker, C.M. (1966) 'The variability of measurements of casual blood pressure. II. Survey experience' Clinical Science 30, 337-344
- Armitage, P. and Rose, G.A. (1966) 'The variability of measurements of casual blood pressure. I. A laboratory study' Clinical Science 30, 325-335
- Armstrong, B.G., Whittemore, A.S. and Howe, G.R. (1989) 'Analysis of case-control data with covariate measurement error: application to diet and colon cancer' Statistics in Medicine 8, 1151-1163
- Boe, J., Humerfelt, S. and Wedervang, F. (1957) 'The blood pressure in a population' Acta Medica Scandinavica 157, Supplement 321
- British Hypertension Society working party (1989) 'Treating mild hypertension: agreement from the large trials' British Medical Journal 298, 694-698
- Caroll, R.J. (1989) 'Covariance analysis in generalized linear measurement error models' Statistics in Medicine 8, 1075-1093
- Chisholm, J.W. (1990) 'The 1990 contract: its history and its content' British Medical Journal 300, 853-856
- Collins, R., Peto, R., MacMahon, S., Hebert, P., Fiebach, N.H., Eberlein, K.A., Godwin, J., Qizilbash, N., Taylor, J.O. and Hennekens, C.H. (1990) 'Blood pressure, stroke, and coronary heart disease. Part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context' The Lancet 335, 827-838
- Committee of Principal Investigators (1978) 'A co-operative trial in the primary prevention of ischaemic heart disease using clofibrate' British Heart Journal 40, 1069-1118
- Committee of Principal Investigators (1980) 'W.H.O. cooperative trial on primary prevention of ischaemic heart disease using clofibrate to lower serum cholesterol: mortality follow-up' The Lancet ii, 379-385
- Committee of Principal Investigators (1984) 'W.H.O. cooperative trial on primary prevention of ischaemic heart disease with clofibrate to lower serum cholesterol: final mortality follow-up' The Lancet ii, 600-604

- Cook, D.G., Shaper, A.G., Pocock, S.J. and Kussick, S.J. (1986) 'Giving up smoking and the risk of heart attacks. A report from the British Regional Heart Study' The Lancet ii, 1376-1379
- Cox, D.R. (1972) 'Regression models and life-tables' (with discussion) Journal of the Royal Statistical Society Series B 34, 187-220
- Cox, D.R. and Oakes, D. (1984) Analysis of Survival Data Chapman and Hall, London
- DerSimonian, R. and Laird, N. (1986) 'Meta-analysis in clinical trials' Controlled Clinical Trials 7, 177-188
- Draper, N.R. and Smith, H. (1981) Applied Regression Analysis Wiley, New York
- Friedman, G.D. , Klatsky, A.L. and Siegelaub, A.B. (1982) 'Alcohol, tobacco and hypertension' Hypertension 4, Supplement III, 143-150
- Gardner, M.J. and Heady, J.A. (1973) 'Some effects of within-person variability in epidemiological studies' Journal of Chronic Diseases 26, 781-795
- Glock, C.Y., Vought, R.L., Clark, E.G. and Schweitzer, M.D. (1956) 'Studies in hypertension: II. Variability of daily blood pressure measurements in the same individuals over a three week period' Journal of Chronic Diseases 4, 469-476
- Goldman, A. (1976) 'Design of blood pressure screening for clinical trials of hypertension' Journal of Chronic Diseases 29, 613-624
- Gordon, T. and Shurtleff, D. (1973) 'Means at each examination and interexamination variation of specified characteristics: Framingham Study, examinations 1 to 10' in Kannel, W.B. and Gordon T. (editors) The Framingham Study: An Epidemiological Investigation of Cardiovascular Disease. Section 29 US Government Printing Office, Washington D.C.
- Gordon, T., Sorlie, P. and Kannel, W.B. (1976) 'Problems in the assessment of blood pressure: the Framingham Study' International Journal of Epidemiology 5, 327-334
- Harlan, W.R., Osborne, R.K. and Graybiel, A. (1962) 'A longitudinal study of blood pressure' Circulation 26, 530-543
- Harrell, F.E. (1986) 'The PHGLM procedure' in SAS Institute Inc. SUGI Supplemental Library User's Guide, Version 5 Edition SAS Institute Inc., Cary, North Carolina
- Hawthorne, V.M., Greaves, D.A. and Beevers, D.G. (1974) 'Blood pressure in a Scottish town' British Medical Journal 3, 600-603
- Heady, J.A. (1973) 'A cooperative trial on the primary prevention of ischaemic heart disease using clofibrate: design, methods, and progress' Bulletin of the World Health Organisation 48, 243-256

- Hebel, J.R., Apostolides, A.Y., Dischinger, P., Entwisle, G. and Su, S. (1980) 'Within-person variability in diastolic blood pressure for a cohort of normotensives' Journal of Chronic Diseases 33, 745-750
- Hofman, A., Feinleib, M., Garrison, R.J. and van Leer, A. (1983) 'Does change in blood pressure predict heart disease?' British Medical Journal 287, 267-269
- Hypertension Detection and Follow-up Program Cooperative Group (1978) 'Variability of blood pressure and the results of screening in the hypertension detection and follow-up program' Journal of Chronic Diseases 31, 651-667
- Janeway, T.C. (1913) 'A clinical survey of hypertensive cardiovascular disease' Archives of Internal Medicine 12, 755-798
- Julius, S., Harburg, E., McGinn, N.F., Keyes, J. and Hoobler, S.W. (1964) 'Relation between casual blood pressure readings in youth and at age 40: a retrospective study' Journal of Chronic Diseases 17, 397-404
- Kalbfleisch, J.D. and Prentice, R.L. (1980) The Statistical Analysis of Failure Time Data Wiley, New York
- Kannel, W.B., Dawber, T.R. and McGee, D.L. (1980) 'Perspectives on systolic hypertension. The Framingham Study' Circulation 61, 1179-1182
- Kannel, W.B., Neaton, J.D., Wentworth, D., Thomas, H.E., Stamler, J., Hulley, S.B. and Kjelsberg, M.O. (1986) 'Overall and coronary heart disease mortality rates in relation to major risk factors in 325348 men screened for MRFIT' American Heart Journal 112, 825-836
- Kannel, W.B., Sorlie, P. and Gordon, T. (1980) 'Labile hypertension: a faulty concept? The Framingham Study' Circulation 61, 1183-1187
- Kesteloot, H. and van Houte, O. (1974) 'An epidemiologic survey of arterial blood pressure in a large male population group' American Journal of Epidemiology 99, 14-29
- Kilpatrick, J.A. (1948) 'The variation of casual, basal and supplemental blood pressures in health and essential hypertension' British Heart Journal 10, 48-56
- Liu, K., Dyer, A. and Stamler, J. (1978) 'Screening for hypertension - some statistical observations' (letter) Journal of Chronic Diseases 31, 643-644
- MacMahon, S., Peto, R., Cutler, J., Collins, R., Sorlie, P., Neaton, J., Abbott, R., Godwin, J., Dyer, A. and Stamler, J. (1990) 'Blood pressure, stroke, and coronary heart disease. Part I, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias' The Lancet 335, 765-774
- Mann, J.I., Lewis, B., Shepherd, J., Winder, A.F., Fenster, S., Rose, L. and Morgan, B. (1988) 'Blood lipid concentrations and other cardiovascular risk factors: distribution, prevalence and detection in Britain' British Medical Journal 296, 1702-1706

- Master, A.M., Dublin, L.I. and Marks, H.H. (1950) 'The normal blood pressure range and its clinical implications' Journal of the American Medical Association 143, 1464-1470
- McKeown, T., Record, R.G. and Whitfield, A.G.W. (1963) 'Variation in casual measurements of arterial pressure in two populations (Birmingham and South Wales) re-examined after intervals of 3-4½ years' Clinical Science 24, 437-450
- Medical Research Council Working Party on Mild to Moderate Hypertension (1985) 'MRC trial of treatment of mild hypertension: principal results' British Medical Journal 291, 97-104
- Miall, W.E. and Lovell, H.G. (1967) 'Relation between change of blood pressure and age' British Medical Journal 2, 660-664
- Participants of the W.H.O./I.S.H meeting (1986) '1986 guidelines for the treatment of mild hypertension: memorandum from a W.H.O./I.S.H. meeting' Bulletin of the World Health Organization 64, 31-35
- Pearson, E.S. and Hartley, H.D. (1954) Biometrika Tables for Statisticians Volume I University Press, Cambridge
- Pepe, M.S., Self, S.G. and Prentice, R.L. (1989) 'Further results on covariate measurement errors in cohort studies with time to response data' Statistics in Medicine 8, 1167-1178
- Prentice, R.L. (1982) 'Covariate measurement errors and parameter estimation in a failure time regression model' Biometrika 69, 331-342
- Rose, G. (1981) 'Strategy for prevention: lessons from cardiovascular disease' British Medical Journal 282, 1847-1851
- Rosner, B. (1977) 'Screening for hypertension - some statistical observations' Journal of Chronic Diseases 30, 7-18
- Rosner, B., Hennekens, C.H., Kass, E.H. and Miall, W.E. (1977) 'Age-specific correlation analysis of longitudinal blood pressure data' American Journal of Epidemiology 106, 306-313
- Rosner, B. and Polk, B.F. (1983) 'Predictive values of routine blood pressure measurements in screening for hypertension' American Journal of Epidemiology 117, 429-442
- Schoenfeld, D. (1982) 'Partial residuals for the proportional hazards regression model' Biometrika 69, 239-241
- Shaper, A.G., Pocock, S.J., Phillips, A.N. and Walker, M. (1986) 'Identifying men at high risk of heart attacks: strategy for use in general practice' British Medical Journal 293, 474-479
- Shaper, A.G., Ashby, D. and Pocock, S.J. (1988) 'Blood pressure and hypertension in middle-aged British men' Journal of Hypertension 6, 367-374

Snedecor, G.W. and Cochran, W.G. (1980) Statistical Methods Iowa State University Press, Iowa

Szklo, M. (1979) 'Epidemiological patterns of blood pressure in children' Epidemiologic Review 1, 143-169

Wilson, P.D. and Hebel, J.R. (1988) 'The probability that long-term mean diastolic blood pressure is extreme: tables for black individuals' Statistics in Medicine 7, 423-433

Wilson, P.D., Hebel, J.R. and Sherwin, R. (1981) 'Screening and diagnosis when within-individual observations are Markov-dependent' Biometrics 37, 553-565

World Health Organization (1959) 'Hypertension and coronary heart disease: classification and criteria for epidemiological studies' World Health Organization Technical Report Series 168

World Health Organization (1962) 'Arterial hypertension and ischaemic heart disease: preventive aspects' World Health Organization Technical Report Series 231