Clinical research

Re-assessing the contribution of serum total cholesterol, blood pressure and cigarette smoking to the aetiology of coronary heart disease: impact of regression dilution bias

Jonathan R. Embersona*, Peter H. Whincupb, Richard W. Morrisa, Mary Walkera

aDepartment of Primary Care and Population Sciences, Royal Free and University College Medical School, London, UK
bDepartment of Community Health Sciences, St George’s Hospital Medical School, London, UK

Received 8 May 2003; received in revised form 1 July 2003; accepted 9 July 2003

Aims To estimate the combined contribution of serum total cholesterol, blood pressure and cigarette smoking to coronary heart disease (CHD) risk after adjustment for regression dilution bias.

Methods and results Six thousand, five hundred and thirteen middle-aged British men without CHD were followed for major CHD events over 10 years. The population attributable risk fraction (PARF) was predicted for a range of risk factor thresholds before and after adjustment for regression dilution of serum total cholesterol and blood pressure. Defining 'low-risk' individuals as being in the bottom tenth of the population distributions of serum total cholesterol (<5.2 mmol/l) and diastolic blood pressure (<70 mmHg) and a non-cigarette smoker, the PARF was 75%, increasing to 86% after adjustment for regression dilution. Regardless of the threshold criteria chosen, the PARF was substantially greater than 65% before adjustment for regression dilution and greater than 75% after adjustment. Exclusion of ex-smokers and passive smokers from the low-risk group increased estimates further. Adjustment for other coronary risk factors had little effect on the results.

Conclusions At least 80% of major CHD events in middle-aged men can be attributed to the three strongest risk factors. Population-wide control of these factors is crucial for effective CHD prevention.

© 2003 Published by Elsevier Ltd on behalf of The European Society of Cardiology.

KEYWORDS
Coronary heart disease; Risk factor assessment; Population attributable risk fraction; Regression dilution bias; Logistic regression

Introduction

Coronary heart disease (CHD) is a major cause of death in Western societies and, increasingly, worldwide.1,2 In the 1950s and 1960s early epidemiological studies documented the importance of environmental and behav-

* Correspondence to: Dr J. R. Emberson, Royal Free and University College Medical School, Primary Care and Population Sciences, Rowland Hill Street, London NW3 2PF, UK. Tel: +44 20 7830 2239; Fax: +44 20 7794 1224
E-mail address: j.emberson@pcps.ucl.ac.uk (J.R. Emberson).
It is increasingly recognized that imprecision of measurement and variation in measurements over time can lead to underestimation of the importance of a risk factor (through regression dilution bias). This phenomenon results from the fact that between-person differences in exposure levels measured at a baseline examination are, on average, greater than true underlying differences. As a result, when baseline measurements are used to quantify how disease risk varies across a range of risk exposure levels, the observed disease gradient underestimates the true gradient across that range. In the past, several reports have confirmed that substantial regression dilution occurs in the relationships between individual factors (particularly blood cholesterol and blood pressure) and CHD. However, to date, the combined effects of measurement imprecision in more than one risk factor on the proportion of CHD cases that can be attributed to high blood cholesterol, high blood pressure and cigarette smoking together (the population attributable risk fraction) have not been examined. Though some recent attempts have been made to reassess the contributions of established risk factors and in particular to refute the suggestion that only half of CHD cases are attributable to established risk factors, these have not taken regression dilution bias into account.

In this paper we use data from a prospective study of cardiovascular disease in middle-aged men (the British Regional Heart Study) to assess the validity of the 'only 50% claim'. In particular, we estimate the population attributable risk fraction for blood cholesterol, blood pressure and cigarette smoking, taking account of measurement imprecision in blood pressure and blood cholesterol. In a subsidiary analysis the effects of using different thresholds for cigarette smoking exposure taking account of previous active smoking and passive smoking are examined.

Methods

The British Regional Heart Study (BRHS) is a prospective study of cardiovascular disease in men, based in one General Practice in each of 24 British towns. In 1978–1980, 7735 participants aged 40–59 years were enrolled. After a detailed baseline examination they have been followed up for all-cause mortality and cardiovascular morbidity, with fewer than 1% of participants being lost to follow-up. Ethical approval was obtained from all relevant local Research Ethics Committees. In two towns (one with high and one with low CHD mortality), men were re-examined after 16 and 20 years of follow-up. These measurements were used to provide estimates of regression dilution ratios for total cholesterol and blood pressure for the present report.

Baseline assessment

Participants were asked about recall of any doctor diagnosis of coronary heart disease (myocardial infarction and definite/possible angina) and whether they had ever had a history of severe chest pain lasting half and hour or more that caused them to consult with a doctor. Participants also completed a WHO (Rose) angina questionnaire providing evidence of definite or possible angina and an electrocardiogram (ECG) test providing evidence of definite or possible myocardial ischaemia or infarction was performed. Seated blood pressure was measured twice in succession in the right arm using the London School of Hygiene and Tropical Medicine sphygmomanometer. Blood pressure was adjusted for observer variation within each town. Non-fasting serum total cholesterol was measured by a modified Liebermann–Burchard method on a Technicon SMA 12/60 analyser. Cigarette smoking status (current/ex/never) was ascertained using an administered questionnaire. For men in 18 towns, serum samples were frozen (~20 °C) and stored. In 2001, cotinine was measured using a gas-liquid chromatography method (detection limit 0.1 ng/ml).

Four year repeat risk factor assessment

Data were collected from two towns after 16 and 20 years of follow-up. A questionnaire provided information about doctor diagnosed CHD and current cigarette smoking habit. On both occasions, trained nurses measured blood pressure twice in the right arm using the Dinamap 1846SX oscillometric recorder and a fasting serum sample was taken for analysis of total cholesterol on a Hitachi 747 automated analyser (using the method of Siedel).

Assessment of incident major CHD

Information on incident mortality was collected through the established 'tagging' procedures provided by the National Health Service registers which provide automatic notification of death and its cause. Fatal coronary events were defined as deaths from ischaemic heart disease including sudden death of presumed cardiac origin (ICD-9 410–414). Non-fatal heart attacks were ascertained from general practitioner reports supplemented by systematic 2-year patient record reviews throughout the study period. A non-fatal heart attack was diagnosed according to World Health Organization criteria, (any report of myocardial infarction accompanied by at least two of: history of severe chest pain, electrocardiographic evidence of myocardial infarction, and enzyme markers of cardiac changes associated with myocardial infarction). Major CHD events were defined as definite non-fatal myocardial infarction and death from coronary heart disease.

Statistical methods

Associations between usual risk factor levels and major CHD risk

The association between baseline risk factors and 10-year major CHD risk was assessed using logistic regression; goodness of fit was assessed using the Hosmer–Lemeshow test. All analyses were adjusted for age, serum total cholesterol, diastolic blood pressure and current cigarette smoking status (there were no significant interactions between these variables) and were restricted to men with no diagnosis or symptoms of CHD. It was assumed throughout that blood pressure and serum total cholesterol were measured with error and subject to variation over time; the effects of this variation were assessed over a 4-year period, as this enables estimation of usual risk factor levels over the first decade of follow-up.

Since observed regression coefficients may underestimate or overestimate true regression coefficients when two or more variables are measured with error, we used the multivariate techniques developed by Rosner et al. to correct for regression dilution. These methods estimate the vector of true regression coefficients $\beta$ from the observed or 'apparent' regression coefficients $\hat{\beta}$ through the equation $\hat{\beta}=\Lambda^{*}\beta$, where $\Lambda$ is a matrix of regression coefficients relating 4-year
follow-up levels of cholesterol and blood pressure to baseline levels (see Rosner et al.26 for further details).

**Estimating the usual risk factor distribution**

The estimates of measurement imprecision obtained from the 4-year repeated data were used to estimate true variances and correlations from observed ones (specific details available from authors on request). These corrected estimates were then used along with the estimated mean risk factor levels in the population to estimate the true ‘usual’ risk factor distribution of total cholesterol, diastolic blood pressure and cigarette smoking in the population, assuming normality in total cholesterol and diastolic blood pressure and assuming the same cholesterol-blood pressure covariance structure for smokers and non-smokers.

**Predicted population attributable risk fraction (PPARF) before and after adjustment for regression dilution**

The population attributable risk fraction is the proportion of disease events in the population that would be avoided if a particular risk factor (or combination of risk factors) was not present. To determine population attributable risk fractions it is necessary to distinguish a ‘low risk’ group (assumed to be free of the relevant risk exposures) from a remaining ‘high risk’ group in which the exposures are present. For the continuous, threshold-free relations between total cholesterol, blood pressure and CHD, a range of cut-off points for the low-risk group have been used, including the usual 10th and 20th centiles in the population and an estimate based on the level above which medical intervention is recommended to reduce CHD risk (total cholesterol 5 mmol/l and diastolic blood pressure 85 mmHg).27 For cigarette smoking, ‘current non-smokers’ have been used as the low-risk group in the main analysis; the effects of restricting the low-risk group to ‘never smokers’ and subjects unexposed either to active or passive smoking have been considered in subsidiary analyses. In each analysis, the absolute risk of major CHD within 10 years for given levels of total cholesterol, blood pressure and smoking status was estimated using the linear predictor obtained from a logistic regression model. For each ‘threshold’ criterion used to define high-risk individuals, the expected probability of major CHD within 10 years for high and low risk individuals was calculated numerically by sampling 100 000 random observations from the observed joint risk factor distribution in the population and calculating the mean predicted risk in each group. The relative risk (RR) of major CHD (for high risk individuals relative to low risk individuals) and the proportion of the population at high-risk (P) were evaluated, and the PPARF calculated by the equation $\frac{P}{1 + P(RR - 1)}$. Adjustment for regression dilution was performed by replacing the unadjusted logistic regression coefficients with those corrected for regression dilution and by sampling risk factor observations from the estimated ‘usual’ population risk factor distribution instead of the observed baseline risk factor distribution. Approximate 95% prediction intervals for the PPARF were calculated to take into account uncertainty in the logistic regression coefficients and uncertainty in the magnitude of measurement imprecision. An empirical Bayes approach was used whereby the regression coefficients (both those before and after correction for regression dilution) were assumed to come from a (normal) probability distribution. The PPARF was then calculated a large number of times (1000) using randomly selected observations from these distributions, and the prediction interval was calculated by taking the 2.5 and 97.5 percentiles of these values. Subsequent analyses that assessed the individual contributions to the population attributable risk fraction of several separate high risk groups (e.g. high total cholesterol only, high blood pressure only), yielded combined attributable risk estimates very similar to those obtained using a single high risk group.

**Results**

**BRHS baseline data and 10-year major CHD follow-up**

Of the 7735 men examined at the baseline screening in 1978–1980, 1159 (15.0%) had a doctor diagnosis or symptoms of CHD (angina or myocardial infarction) and were excluded from this analysis. Of the remaining men, 6513 (99%) had complete data on diastolic blood pressure, serum total cholesterol and cigarette smoking status. Table 1 shows their observed baseline risk factor levels and estimated usual levels over the study period. Baseline mean serum total cholesterol was the same amongst smokers and non-smokers, diastolic blood pressure was slightly higher amongst men who did not smoke (83 mmHg vs 81 mmHg). A weak positive correlation between baseline total cholesterol and baseline diastolic

**Table 1** Observed and usual population risk factor levels for the 6513 men with complete risk factor data and no history of CHD (previous diagnosis or current symptoms)

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Observed value</th>
<th>Estimated usual value in the population over the study period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum total cholesterol (mmol/l)</td>
<td>Mean 6.3</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>Standard deviation 1.0</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>20th centile 5.4</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>10th centile 5.0</td>
<td>5.2</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg)</td>
<td>Mean 82.0</td>
<td>82.0</td>
</tr>
<tr>
<td></td>
<td>Standard deviation 13</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>20th centile 71.0</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>10th centile 66.0</td>
<td>70</td>
</tr>
<tr>
<td>Proportion current cigarette smokers (%)</td>
<td>40.4</td>
<td>40.4</td>
</tr>
</tbody>
</table>

*This is the observed value after correction for regression dilution bias.*
blood pressure was observed \((r=0.11)\). Both serum total cholesterol and diastolic blood pressure were normally distributed.

During the following 10 years, 426 men (6.5%) had a definite major CHD event. The relations between usual levels of serum total cholesterol and diastolic blood pressure and CHD (adjusted for age, cigarette smoking and each other) were linear with no apparent thresholds \((\text{Fig. 1})\), and were unaffected by adjustment for body mass index, physical inactivity and diabetes. The model relating the risk of major CHD to age, serum total cholesterol, diastolic blood pressure and current cigarette smoking status fitted the data adequately.

**Four-year estimates of regression dilution**

Three hundred and sixty-one men with no previous history of CHD attended the 16-year screening in the two towns studied (80% response). Of these men, 340 survived the following 4 years, of whom 259 (76%) attended the 20-year re-screening. Using these repeat blood pressure and serum total cholesterol measurements, regression dilution ratio estimates of 0.72 (95% confidence interval \((\text{CI})\) 0.61 to 0.82) for serum total cholesterol, and 0.52 (95% CI 0.42 to 0.62) for diastolic blood pressure were obtained.\(^26\)

**Estimates of predicted population attributable risk fractions (PPARFs)**

**Risk factors considered in isolation**

The PPARFs for serum total cholesterol and diastolic blood pressure using different thresholds are shown in \(\text{Fig. 2}\), before and after adjustment for regression dilution. Each curve is adjusted for age and cigarette smoking and mutually adjusted for total cholesterol and diastolic blood pressure. For both risk factors, the PPARF increases as the threshold criteria used to define the low risk group decreases. The PPARF estimates for low risk criteria based on the 20th and 10th centiles of the usual risk factor distributions and on thresholds for current clinical interventions are shown in \(\text{Table 2}\). The PPARF for serum total cholesterol at the 20th and 10th centiles are 42% and 48%, rising to 52% and 60% after adjustment for regression dilution. Corresponding figures for diastolic blood pressure are 31% and 36%, rising to 45% and 52% after adjustment for regression dilution. Using current clinical intervention thresholds to define low-risk individuals the PPARF for high serum total cholesterol alone \((\geq 5 \text{ mmol/l})\) is 50%, increasing to 62% after adjustment for regression dilution, whilst for high diastolic blood pressure \((\geq 85 \text{ mmHg})\), it is 19% before and 25% after adjustment. For current cigarette smoking, the PPARF is 24%.

**Risk factors considered simultaneously**

\(\text{Fig. 3}\) displays the PPARF for high serum total cholesterol, high blood pressure and current cigarette smoking considered simultaneously, both before and after adjusting for the effects of regression dilution (also shown in \(\text{Table 2}\)). Using the 20th centiles of the usual risk factor levels to define the low-risk group, the PPARF is 70%, increasing to 81% after adjustment for regression dilution. Using the 10th centiles of usual risk factor levels, the PPARF is 75%, rising to 86% after adjustment for regression dilution.
Similar estimates are obtained using threshold criteria based on current clinical intervention thresholds for CHD prevention (see Fig. 3(c)).

**Further analyses: effect of different thresholds for cigarette smoking exposure**

In the subset of 5883 subjects who had ever smoked cigarettes or had cotinine measurements taken, the effect of using different low-risk thresholds for cigarette smoking on PPARF estimates was examined (Table 3). Restricting the low-risk group to subjects who had never smoked (i.e. adding ex smokers to the high risk group) increased the PPARF estimates slightly. Restricting the low-risk group further to subjects who had never smoked and were not heavily exposed to environmental tobacco smoke (cotinine levels <1.5 ng/ml, excluding subjects exposed to a partner smoking 20 cigarettes or more per day) increased the PPARF estimates still further. The estimates of the PARF of 70–75% before adjustment for regression dilution bias are consistent with estimates from earlier reports using the same risk factors, including those from the National Co-operative Pooling Project (66%), the Whitehall I Study (66%) and the MRFIT Study (75–77%), a slightly higher estimate (82%) was derived from smoking, physical activity, obesity and a composite dietary score emphasizing fat intake.

**Discussion**

We believe this is the first report examining the true contribution of blood cholesterol, blood pressure and cigarette smoking to the population attributable risk of CHD taking regression dilution of both blood cholesterol and blood pressure into account. Defining low-risk individuals as being in the lowest fifth of usual levels of serum total cholesterol and diastolic blood pressure and current non-smokers, the PPARF was 70% before and 81% after adjustment for regression dilution bias. Corresponding values of 75% and 86% were achieved when using the lowest tenths of the usual risk factor distributions. Extending the definition of non-smokers to exclude ex-smokers and those exposed to heavy passive smoking increased the PPARF estimates still further. The estimates of the PARF of 70–75% before adjustment for regression dilution bias are consistent with estimates from earlier reports using the same risk factors, including those from the National Co-operative Pooling Project (66%), the Whitehall I Study (66%) and the MRFIT Study (75–77%), a slightly higher estimate (82%) was derived from smoking, physical activity, obesity and a composite dietary score emphasizing fat intake.

**Validity of analyses**

The validity of the analyses presented depends on the appropriateness of the estimates of regression dilution used and the thresholds used to define individuals as low risk. The estimates of regression dilution used in this
Table 2  Predicted population attributable risk fraction (PPARF) estimates for a range of threshold criteria before and after adjustment for regression dilution bias

<table>
<thead>
<tr>
<th>Definition of 'low-risk' individuals</th>
<th>Threshold used</th>
<th>PPARF estimate (95% prediction interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Before correction</td>
</tr>
<tr>
<td>Usual serum total cholesterol</td>
<td>Intervention(^b)</td>
<td>50%(41%,59%)</td>
</tr>
<tr>
<td>&lt;5.0 mmol/l</td>
<td>Usual 10th centile</td>
<td>48%(39%,56%)</td>
</tr>
<tr>
<td>&lt;5.2 mmol/l</td>
<td>Usual 20th centile</td>
<td>42%(33%,50%)</td>
</tr>
<tr>
<td>&lt;5.5 mmol/l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual diastolic blood pressure</td>
<td>Usual 10th centile</td>
<td>36%(26%,45%)</td>
</tr>
<tr>
<td>&lt;70 mmHg</td>
<td>Usual 20th centile</td>
<td>31%(22%,39%)</td>
</tr>
<tr>
<td>&lt;74 mmHg</td>
<td>Intervention</td>
<td>19%(13%,25%)</td>
</tr>
<tr>
<td>&lt;85 mmHg</td>
<td></td>
<td>24%(16%,32%)</td>
</tr>
<tr>
<td>Non smoker</td>
<td>Usual 20th centile</td>
<td>70%(63%,76%)</td>
</tr>
<tr>
<td>Serum total cholesterol &lt;5.5 mmol/l</td>
<td>Non smoker</td>
<td>24%(16%,32%)</td>
</tr>
<tr>
<td>Serum total cholesterol &lt;5.2 mmol/l</td>
<td>Usual 10th centile</td>
<td>75%(69%,81%)</td>
</tr>
<tr>
<td>Serum total cholesterol &lt;5.5 mmol/l</td>
<td>Usual 20th centile</td>
<td>42%(33%,50%)</td>
</tr>
<tr>
<td>Serum total cholesterol &lt;5.5 mmol/l</td>
<td>Intervention</td>
<td>70%(63%,76%)</td>
</tr>
<tr>
<td>Serum total cholesterol &lt;5.5 mmol/l</td>
<td>Non smoker</td>
<td>70%(63%,76%)</td>
</tr>
</tbody>
</table>

\(^a\)All PPARF estimates are adjusted for age, serum total cholesterol, diastolic blood pressure and cigarette smoking status. The right hand column also corrects for imprecision in serum total cholesterol and diastolic blood pressure.

\(^b\)See methods for definition of clinical intervention targets.

Fig. 3  PPARF estimates for serum total cholesterol, diastolic blood pressure and cigarette smoking considered simultaneously before (thin line) and after (thick line) adjusting for regression dilution. The three graphs correspond to threshold criteria for diastolic blood pressure based on (A) 20th centiles, (B) 10th centiles and (C) current clinical intervention thresholds. The dotted line in each graph corresponds to these criteria for serum total cholesterol.

Analysis were based on repeat blood pressure and cholesterol measurements taken 4 years apart, and were restricted to individuals with no previous history of CHD. Though the subjects were older at the time of the repeat measurements than at baseline, estimates of regression dilution do not vary markedly with age.\(^{14}\) Moreover, our 4-year regression dilution ratios of 0.72 and 0.52 for serum total cholesterol and diastolic blood pressure respectively were very similar to those derived from the Framingham Study over the same period (0.73 and 0.57 respectively).\(^{32}\) Separation of low and high risk groups on the basis of blood cholesterol and blood pressure levels is arbitrary; evidence from other study populations suggests that the relations between these factors and CHD...
have no threshold and continue below the levels of the bottom tenths of our study population.\textsuperscript{5,7,33} Thus, even in our low-risk group the effects of cholesterol and blood pressure exposure on CHD risk are likely to be appreciable, compared for example with the Japanese cohorts of the Seven Countries Study, in whom both mean serum total cholesterol level and CHD mortality risk were markedly lower than was the case even in our low-risk group.\textsuperscript{4} As a result, these analyses are conservative, tending to underestimate rather than overestimate the potential contribution of these factors.

Implications for prevention of coronary heart disease

Our results suggest that effective early primary prevention of CHD focused on three factors alone — blood cholesterol, blood pressure and cigarette smoking could substantially reduce CHD, though the effectiveness of primary prevention initiated in middle life would be limited by the extent of reversibility of risk at this stage.\textsuperscript{34,35} How generalisable are our findings? Although they are based in men in a high income country with a high prevalence of risk factors and CHD rates, they may be more widely applicable. In particular, they may apply to women, in whom similar population distributions and relative risks apply.\textsuperscript{12,17} They may also be of considerable relevance to the epidemic of coronary heart disease now emerging in the less affluent countries of the world, in which rising levels of blood cholesterol and blood pressure, and marked increases in the prevalence of cigarette smoking are prominent.\textsuperscript{2}

Implications for new risk factors for CHD

In recent years other potential risk markers — nutritional, genetic, social, infective, inflammatory and foetal growth, have been implicated in the causation of coronary heart disease.\textsuperscript{10,36–41} While our data do not rule out the existence of other independent risk factors for CHD (the evidence for involvement of homocysteine is particularly persuasive),\textsuperscript{42,43} such factors would need to be widely distributed and strongly related to CHD to make a substantial contribution to the CHD epidemic. Once cholesterol, blood pressure and cigarette smoking are accounted for however, the scope for other risk factors to have a large contribution towards attributable CHD risk is limited. The results suggest that greater emphasis on the control of these established CHD risk factors and improved understanding of the determinants of these factors are likely to be critical to the control of the current CHD epidemic.

Conclusions

Three established risk factors for CHD (high serum total cholesterol, high blood pressure and cigarette smoking) account for at least four-fifths of the attributable risk of CHD during middle age. Claims that half of CHD risk remains unexplained are completely untenable. The residual variation may be explained once changes in smoking habits and other established risk factors such as physical inactivity and obesity have been taken into account. Effective policies directed to the reduction of serum total cholesterol, blood pressure and cigarette smoking rates at the population level should control population CHD epidemics.

Acknowledgements

The British Regional Heart Study (established by Professor A.G. Shaper) is a British Heart Foundation Research Group and receives additional support from the Department of Health (England) and the BUPA Foundation. JRE is funded by a British Heart Foundation Junior Research Fellowship. Baseline serum total cholesterol was
measured at the Wolfson Research Laboratories, Birmingham (Prof. T. Whitehead). Total cholesterol measurements at follow-up examinations were carried out in the Department of Chemical Pathology, Royal Free Hospital (Dr M. Thomas). Cotinine assays were carried out at ABS Laboratories, New Cross Hospital (C. Feyerabend, A. Bryant). The views expressed in this publication are those of the authors and not necessarily those of the funding agencies.

References

12. Magnus P, Beaglesohole R. The real contribution of the major risk factors to the coronary epidemics: time to end the ‘only 50% myth’. Arch Int Med 2001;161:2657–60.