PAPERS

Systematic underestimation of association between serum cholesterol concentration and ischaemic heart disease in observational studies: data from the BUPA study

M R Law, N J Wald, T Wu, A Hackshaw, A Bailey

Abstract

Objective—To estimate the size of the association between serum concentration of low density lipoprotein cholesterol and mortality from ischaemic heart disease.

Design—Prospective study of total serum cholesterol concentration and mortality from ischaemic heart disease in 21 515 men (538 deaths) and study of total cholesterol concentration measured on two occasions an average of three years apart in 5696 men in whom low density lipoprotein cholesterol concentration was also measured on the second occasion.

Subjects—Men who attended the medical centre of the British United Provident Association (BUPA) in London between 1975 and 1982.

Main outcome measure—The difference in mortality from ischaemic heart disease for a 0·6 mmol/l difference in concentration of low density lipoprotein cholesterol after adjustment for, firstly, regression dilution bias, which arises from the random fluctuation of serum cholesterol concentration in people over time, and, secondly, the surrogate dilution effect, which arises because differences in total cholesterol concentration between people reflect smaller differences in low density lipoprotein cholesterol concentration.

Results—The observed difference in mortality from ischaemic heart disease associated with a difference of 0·6 mmol/l in total serum cholesterol concentration was 17% but increased to 24% after correction for the regression dilution bias and to 27% (95% confidence interval 21% to 33%) after adjustment for both sources of underestimation, which provides an estimate of the difference in mortality for a true difference of 0·6 mmol/l in low density lipoprotein cholesterol concentration. The association was greater at younger ages. The estimated decrease in mortality from all causes was 6% before and 10% (1% to 17%) after adjustment for the two sources of underestimation. There was no excess mortality from any cause associated with low cholesterol concentration.

Conclusions—The association between serum cholesterol concentration and ischaemic heart disease is materially stronger than directly inferred from prospective studies. This has important implications for the health benefit of achieving low cholesterol concentrations.

Introduction

There is conclusive evidence that the association between serum cholesterol concentration and ischaemic heart disease is one of cause and effect. The size of the association is uncertain, however, and needs to be quantified to assess the effects of dietary change or cholesterol lowering drugs on the risk of ischaemic heart disease. Randomised trials do this directly and have shown that the risk is reversible, but their duration may have been too short to show the full effect. Observational studies, on the other hand, show the effect of long standing differences in cholesterol concentration between people, the differences having been present for many years before recruitment. Direct analysis of data from cohort studies, however, underestimates the association between cholesterol concentration and ischaemic heart disease through two mechanisms, the regression dilution bias and an effect we call the surrogate dilution effect. We therefore derived estimates to allow for these two sources of underestimation to obtain an accurate estimate of the long term reduction in risk of ischaemic heart disease for a given change in serum cholesterol concentration.

REGRESSION DILUTION BIAS

The first source of underestimation, the regression dilution bias, has previously been described.[1-5] It affects all regression analyses in which the independent variable (plotted on the horizontal axis) is subject to random variation over time through errors in measurement and fluctuation within a person. The cohort studies of serum cholesterol concentration and ischaemic heart disease have generally measured cholesterol concentration once in each person and ranked the values to divide the cohort into subgroups, usually five equal (quintile) groups. The dose-response relation between the mean cholesterol concentrations of the five groups and the five corresponding death rates for ischaemic heart disease can then be calculated. It is the use of the same measurements both to stratify the cohort into groups and to measure the mean of each group that introduces bias.

Because serum cholesterol concentration fluctuates over time single measurements can be higher or lower than the long term average value of an individual person. The groups with the higher cholesterol concentrations will include a disproportionate number of people selected because the single measurement was by chance higher than their long term average value, so the long term mean cholesterol concentration of these groups will be overestimated. Similarly, in the groups with lower concentrations the long term mean cholesterol concentration will be underestimated. The range of concentrations across the groups will be wider if based on single measurements than on long term average values. Both are associated with the same range of death rates for ischaemic heart disease, so a plot of mortality from ischaemic heart disease against serum cholesterol concentration will be shallower when based on single measurements than on long term average values.

BUPA Epidemiological
Research Group,
Department of
Environmental and
Preventive Medicine,
Wolfson Institute of
Preventive Medicine, St
Bartholomew's Hospital
Medical College,
London EC1M 6BQ
M R Law, *senior lecturer*
N J Wald, *professor*
T Wu, *computer programmer*
A Hackshaw, *statistician*

BUPA Health Services,
London WC2R 3AU
A Bailey, *former medical director*

Correspondence to:
Professor Wald.

SURROGATE DILUTION EFFECT

In cohort studies a difference of 1 mmol/l in total serum cholesterol concentration might be expected to be equivalent to a difference in low density lipoprotein cholesterol of 0·67 mmol/l (since the latter is about two thirds of total). In intervention studies (trials), however, a reduction in total cholesterol concentration of 1 mmol/l corresponds to a reduction in low density lipoprotein cholesterol of 1 mmol/l because it is this component that is specifically altered by diet[6] and by most drugs. It follows that a given difference in total serum cholesterol concentration will correspond to a smaller difference in mortality from ischaemic heart disease in the cohort studies than in the trials (as the major effect on risk of ischaemic heart disease is related to low density lipoprotein cholesterol). Relative to the expected long term results from the trials the large observational studies underestimate the effect on ischaemic heart disease through measuring and classifying subjects by total cholesterol instead of low density lipoprotein cholesterol concentration. This surrogate dilution effect cannot be allowed for simply by using the 0·67:1 ratio because of the added complication that total cholesterol also includes high density lipoprotein cholesterol (about a quarter of total), which is inversely associated with low density lipoprotein cholesterol. The extent of the underestimation will therefore be less than the expected 0·67:1 ratio. Our aim was to estimate the underestimation so that a valid comparison could be made between the observational studies and the trials.

The regression dilution bias and the surrogate dilution effect are independent; if the long term average total cholesterol concentration of each person were known (abolishing regression dilution bias) those with high (or low) total cholesterol concentrations would still, on average, have high (or low) concentrations of both very low density lipoprotein and low density lipoprotein cholesterol.

CORRECTING FOR UNDERESTIMATION

A simple procedure corrects for both sources of underestimation. To correct for the regression dilution bias it is not necessary to know each person's long term average cholesterol concentration (that would require many repeat measurements). It is sufficient to know the long term mean cholesterol concentration of each subgroup (fifth) of the ranked cholesterol distribution, which can be determined by retesting a sample of men from each group once, provided that the interval between the two measurements is long enough to overcome the cyclical fluctuation in cholesterol concentrations. Measuring low density lipoprotein cholesterol concentration at a separate visit in a sample of all the men will estimate the mean low density lipoprotein cholesterol concentration of each group (table I) and simultaneously allow for both the regression dilution bias and the surrogate dilution effect. A

statistically more powerful approach that follows the same principle is to regress low density lipoprotein cholesterol concentration measured at a second visit on total cholesterol concentration measured at the first visit. The slope of the regression estimates the combined underestimation effect.

Subjects and methods

The BUPA study is a prospective study of 21 520 professional men and businessmen aged 35-64 years who attended the medical centre of the British United Provident Association (BUPA) in London for a comprehensive medical examination (including a measurement of total cholesterol concentration) between 1975 and 1982.[7 8] The study was restricted to those whose NHS records could be flagged so that the Office of Population Censuses and Surveys could inform us of all deaths and their certified causes. Further information was then sought from the doctor who certified death. Five men did not have their serum cholesterol concentration measured, leaving 21 515 in the study. Follow up was complete to the end of 1991 (276 500 man years, a mean follow up of 12·9 years). We subtracted 0·43 mmol/l from cholesterol values obtained before 1979 to allow for the change in the method of measurement described previously.[7]

STATISTICAL METHODS

The cohort was divided into fifths of the ranked distribution of total cholesterol concentration, and the death rates for ischaemic heart disease (in logarithms) of each group (weighted by the number of deaths from ischaemic heart disease) were regressed linearly on serum cholesterol concentration. The data fitted the log linear model reasonably well; a quadratic model was less satisfactory. The slope (or regression coefficient) was expressed as the age adjusted percentage difference in mortality from ischaemic heart disease for a difference in total cholesterol concentration of 0·6 mmol/l (about 10%). Smoking, blood pressure and the effect of any treatment for high blood pressure or high cholesterol concentration had little effect on the results.

ADJUSTMENT FOR TWO SOURCES OF UNDERESTIMATION

We identified a group of 5696 men who attended twice an average of three years apart. Total cholesterol concentration was measured on both occasions and on the second occasion high density lipoprotein cholesterol and triglyceride concentrations (after overnight fasting) were also measured so that the concentration of low density lipoprotein cholesterol could be estimated by using the method of Friedewald et al.[9] (Use of the modification to the Friedewald equation proposed by DeLong et al[10] yielded the same estimate of the surrogate dilution effect.) Such estimates of low density lipoprotein cholesterol concentration in fasting subjects are close to values obtained by ultracentrifugation, the standard method ($r = 0.93$).[10 11] The fasting status of the 5696 men was confirmed by the mean (SD) serum triglyceride concentration of 1·34 (0·87) mmol/l, which was close to the reference fasting value.[12] The 5696 men were representative in that the mean of their original measurements of total cholesterol concentration (6·3 (1·1) mmol/l) was the same as that in the cohort of 21 515 men. As outlined above, the low density lipoprotein cholesterol measurements at the second visit were regressed on the total cholesterol measurements at the first visit.

Results

Table I shows, firstly, the mean total cholesterol concentration of each fifth of the distribution based on

TABLE I—Mean serum cholesterol by fifths of distribution of ranked measurements of total cholesterol concentration in cohort of 21 515 men, mean total and low density lipoprotein cholesterol concentrations from repeat measurements, and mortality from ischaemic heart disease

| Fifth of distribution of ranked serum cholesterol | Serum cholesterol (mmol/l) | | | | Mortality from ischaemic heart disease (entire cohort) | |
| | Entire cohort | | Repeat measurement in group of 5696 men | | | |
	No of men	Original total cholesterol (mean (range))	Mean total cholesterol	Mean (SE) low density lipoprotein cholesterol	No of deaths	Age adjusted rate per 1000 man years
1	4420	4·8 (<5·31)	5·0	3·3 (0·02)	57	1·11
2	4536	5·6 (5·31 to 5·94)	5·6	3·9 (0·02)	89	1·57
3	4108	6·2 (5·95 to 6·46)	6·1	4·3 (0·02)	101	1·86
4	4336	6·8 (6·47 to 7·11)	6·5	4·7 (0·02)	118	2·01
5	4115	7·9 (>7·11)	7·2	5·3 (0·02)	173	3·11
Total	21 515	6·3	6·1	4·3 (0·01)	538	1·95

the original measurements in the cohort of 21 515 men (the group of 5696 men had the same original values); secondly, the mean total cholesterol concentration; thirdly, the mean low density lipoprotein cholesterol concentration of each fifth based on the repeat measurements in the group of 5696 men; and fourthly, the age adjusted death rates for ischaemic heart disease for the five groups. The threefold difference in mortality from ischaemic heart disease across the groups (1·11 to 3·11 deaths per 1000 man years) corresponded to a difference in original total cholesterol concentration of 3·1 (from 4·8 to 7·9) mmol/l but a smaller difference in repeat total cholesterol concentration of 2·2 (5·0 to 7·2) mmol/l; this is the regression dilution bias. The threefold difference in mortality from ischaemic heart disease corresponded to an even smaller difference in low density lipoprotein cholesterol concentration of 2·0 (3·3 to 5·3) mmol/l; this is the surrogate dilution effect.

Regression of the mortality from ischaemic heart disease against the mean cholesterol concentration of each fifth of the distribution by using the original measurements of total cholesterol concentration yielded the estimate that a difference of 0·6 mmol/l in total cholesterol concentration corresponded to a difference in mortality from ischaemic heart disease of

17% (table II). Regression of the repeat measurements on the original measurements in the 5696 men estimated that this difference in cholesterol concentration of 0·6 mmol/l based on original measurements corresponded to a true difference of only 0·424 mmol/l based on the repeat measurements of total cholesterol concentration and to a difference of 0·373 mmol/l based on the repeat measurements of low density lipoprotein cholesterol concentration. A true difference of 0·6 (rather than 0·424) mmol/l in total cholesterol concentration then corresponded to a difference in mortality from ischaemic heart disease of 24% rather than 17%, and a true difference of 0·6 (rather than 0·373) mmol/l in low density lipoprotein cholesterol concentration corresponded to a difference in mortality from ischaemic heart disease of 27% (95% confidence interval 21% to 33%). The slope of the regression line of mortality from ischaemic heart disease on serum cholesterol concentration (b, the linear regression coefficient) was increased by 61% (0·6/0·373=1·61) to correct for the regression and the surrogate dilution or by 41% (0·6/0·424=1·41) for the first and by 14% (0·424/0·373=1·14) for the second. This adjustment procedure had high precision with a 95% confidence interval about the final estimate of a 27% decrease in the risk of ischaemic heart disease of 26·5% to 27·5%. The estimate of 61% for the increase of the slope did not vary with age and applied to men of all ages. The decreasing association between serum cholesterol concentration and risk of ischaemic heart disease with age is shown in table II.

Mortality from all causes was also significantly associated with concentration of low density lipoprotein cholesterol; over all ages a difference in concentration of low density lipoprotein cholesterol of 0·6 mmol/l was associated with a difference in mortality of 10% (1% to 17%; table II). There was no suggestion of any increase in mortality from causes other than ischaemic heart disease with low cholesterol concentration (table III), though as reported before there was an excess of mortality from cancer in the lowest fifth in the first two years of follow up, attributable to preclinical cancer lowering cholesterol concentration.[7]

Discussion

Allowance for the two sources of underestimation increased the regression coefficient, b, by 61% (41% for the regression dilution bias and 14% for the surrogate dilution effect, 1·41×1·14=1·61). The estimate of the difference in mortality from ischaemic heart disease for a difference in cholesterol concentration of 0·6 mmol/l increased from 17% to a final estimate of 27% for the association with long term average values of low density lipoprotein cholesterol. The adjustment procedure was precise, with a 95% confidence interval of 26·5% to 27·5%.

The correction factor for regression dilution bias of

TABLE II—*Estimated difference in mortality from ischaemic heart disease and all causes per 0·6 mmol/l (about 10%) difference in serum cholesterol concentration, showing effect of adjusting for regression dilution bias and surrogate dilution effect*

Age at entry (mean age at death) (years)	No of deaths	Estimated decrease in mortality per 0·6 mmol/l decrease in serum cholesterol (%)		
		By using original total cholesterol measurements	By using repeat total cholesterol measurements	By using repeat low density lipoprotein cholesterol measurements (95% confidence interval)
Mortality from ischaemic heart disease:				
35-44 (49)	74	31	41	44 (31 to 55)
45-54 (58)	196	17	24	26 (16 to 36)
55-64 (68)	268	12	17	19 (10 to 28)
35-64 (62)	538	17	24	27 (21 to 33)
All cause mortality (all ages)	1543	6	9	10 (1 to 17)

TABLE III—*Age adjusted death rates per 1000 man years by category according to fifths of ranked distribution of serum cholesterol concentration*

Cause of death*	No of deaths	Serum cholesterol group (mmol/l)				
		<5·31	5·31 to 5·94	5·95 to 6·46	6·47 to 7·11	>7·11
Ischaemic heart disease (codes 410-14)	538	1·11	1·57	1·86	2·01	3·11
Other circulatory diseases (codes 390-405, 415-459, 557, 745-47)	167	0·50	0·79	0·37	0·67	0·67
Cancer (codes 140-239)	593	1·96	2·50	2·12	2·16	1·96
Accidents and suicide (codes 800-995)	70	0·22	0·26	0·21	0·28	0·30
Other causes	175	0·59	0·76	0·66	0·57	0·57
All causes except ischaemic heart disease	1005	3·27	4·31	3·36	3·69	3·49
All causes including ischaemic heart disease	1543	4·38	5·88	5·22	5·70	6·60

*According to *International Classification of Diseases*, ninth revision.

TABLE IV—*Estimates of within person, between person, and total standard deviation (SD) of serum cholesterol concentration in men from 11 studies according to interval between initial and repeat cholesterol measurements*

Study	Mean interval between first and second measurement of cholesterol concentration	SD of serum cholesterol (mmol/l)			Ratio of total to between person variance (t^2/b^2)
		Within person (w)	Between person (b)	Total (t) ($t^2=w^2+b^2$)	
National diet heart study[13] (United States; n=962)	3 Weeks	0·32	0·94	0·99	1·1
Central Sweden[4] (n=7616)	6 Weeks	0·62	1·05	1·22	1·3
San Francisco[14] (n=621)	8 Weeks	0·42	0·86	0·96	1·2
Lipid research clinics program[5 15] (United States; n=2492)	10 Weeks	0·41	0·84	0·93	1·2
Medical Research Council[16] (Britain; n=7616)	1 Year	0·46	0·91	1·02	1·3
Central Sweden[4] (n=7349)	2 Years	0·69	1·00	1·22	1·5
World Health Organisation[1]	2 Years	0·65	1·03	1·22	1·4
Western Electric[17] (United States; n=1900)	1 Year	0·83	1·13	1·41	1·5
United States naval[18] (n=381)	6 Years	0·51	0·96	1·09	1·3
Framingham[19] (United States; n=620)	8 Years	0·57	0·86	1·03	1·4
Present study (n=5696)	3 Years	0·56	0·93	1·09	1·4

1·41 is equivalent to the ratio of total variance (that is, within person plus between person) to between person variance of total cholesterol concentration.[3] Table IV compares our estimate with other published estimates of within and between person standard deviation in total cholesterol concentration (or data from which these parameters could be calculated). The studies are ranked in the table according to the interval between the two cholesterol measurements. As would be expected, the within person standard deviation tended to be lower in studies in which the interval between the two measurements was only a few weeks, while the between person standard deviation was similar in all the studies. Our estimate of the ratio of total to between person variance (the regression dilution correction factor) is similar to that from the seven other studies with an interval of one year or more between the two measurements. A smaller correction factor based on repeat measurements taken a few weeks apart[5] is likely to have underestimated the effect.

The surrogate dilution effect has not previously been estimated. Some cohort studies have measured the concentration of low density lipoprotein cholesterol directly in all subjects (avoiding the need for the correction), but these studies have recorded relatively few deaths from ischaemic heart disease and lack published data on the appropriate correction for the regression dilution bias for low density lipoprotein cholesterol concentration. Our indirect estimate of the relation between low density lipoprotein cholesterol concentration and ischaemic heart disease has been validated by measurements of apolipoprotein B (the protein component of low density lipoprotein) in men in the BUPA study, which yielded age specific estimates for the association with mortality from ischaemic heart disease (corrected for regression dilution bias) that were close to our present estimates for low density lipoprotein cholesterol concentration.[8]

In the introduction we indicated that the surrogate dilution effect would be less than expected because the concentrations of high and low density lipoprotein cholesterol were inversely associated. In view of this it is perhaps surprising that high density lipoprotein cholesterol is independent of total cholesterol concentration[20] (our data showed that a 1·00 mmol/l difference in the concentration of total serum cholesterol was associated with a 0·01 mmol/l (95% confidence interval 0·00 to 0·02) difference in high density lipoprotein cholesterol). This independence arises because the tendency for the concentrations of high density lipoprotein cholesterol and total cholesterol to be positively associated (because total cholesterol concentration will

be classified as higher with raised concentration of high density lipoprotein cholesterol) is exactly offset by the modest inverse association between high and low density lipoprotein cholesterol. In practice therefore high density lipoprotein cholesterol does not contribute to the surrogate dilution effect. This effect is entirely due to very low density lipoprotein cholesterol, which is positively associated with total cholesterol but unrelated to the risk of ischaemic heart disease.[21] The size of the surrogate dilution effect is small because very low density lipoprotein cholesterol accounts for only about 10% of total cholesterol.

We have shown therefore that the results of cohort studies that use single measurements of total cholesterol concentrations can be adjusted to produce unbiased estimates of the underlying true relation of low density lipoprotein cholesterol concentration with ischaemic heart disease. A dietary change that lowers total cholesterol concentration by 0·6 mmol/l lowers low density lipoprotein cholesterol concentration by a similar absolute amount,[7] and in middle aged men the corresponding decrease in the risk of ischaemic heart disease is 25-30%.

1 Gardner MJ, Heady JA. Some effects of within-person variability in epidemiological studies. Journal of Chronic Diseases 1973;26:781-95.
2 Chen Z, Peto R, Collins R, MacMahon S, Lu J, Li W. Serum cholesterol concentration and coronary heart disease in a population with low cholesterol concentrations. BMJ 1991;303:276-82.
3 Frost CD, Law MR, Wald NJ. By how much does salt restriction lower blood pressure? II. An analysis of observational data within populations. BMJ 1991;302:815-8.
4 Törnberg S, Jakobsson KFS, Eklund GA. Stability and validity of a single serum cholesterol measurement in a prospective cohort study. Int J Epidemiol 1988;17:797-803.
5 Davis CE, Rifkind BM, Brenner H, Gordon DJ. A single cholesterol measurement underestimates the risk of coronary heart disease. JAMA 1990;264:3044-6.
6 Grundy SM, Nix D, Whelan MF, Franklin L. Comparison of three cholesterol-lowering diets in normolipidemic men. JAMA 1986;256: 2351-5.
7 Wald NJ, Thompson SG, Law MR, Densem JW, Bailey A. Serum cholesterol and subsequent risk of cancer: results from the BUPA study. Br J Cancer 1989;59:936-8.
8 Wald NJ, Law M, Watt H, Wu T, Bailey A, Johnson M, et al. Apolipoproteins and ischaemic heart disease: implications for screening. Lancet 1994;343: 75-9.
9 Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972;18:499-502.
10 DeLong DM, DeLong ER, Wood PD, Lippel K, Rifkind BM. A comparison of methods for the estimation of plasma low- and very low-density lipoprotein cholesterol. JAMA 1986;256:2372-7.
11 Rao A, Parker AH, El-Sheroni N, Babelly MM. Calculation of low-density lipoprotein cholesterol with use of triglyceride/cholesterol ratios in lipoproteins compared with other calculation methods. Clin Chem 1988;34: 2532-4.
12 Lentner C, ed. Geigy scientific tables. Vol 3. Physical chemistry, composition of blood. 8th ed. Basle: Ciba-Geigy, 1984.
13 National Diet-Heart Study Research Group. The national diet-heart study final report. Circulation 1968;37-8(I suppl):1-428.
14 Williams GZ, Widdowson GM, Penton J. Individual character of variation in time-series studies of healthy people. Differences in values for clinical chemical analytes in serum among demographic groups, by age and sex. Clin Chem 1978;24:313-20.
15 Jacobs DR, Barrett-Connor E. Retest reliability of plasma cholesterol and triglyceride. Am J Epidemiol 1982;116:878-85.
16 Thompson SG, Pocock SJ. The variability of serum cholesterol measurements: implications for screening and monitoring. J Clin Epidemiol 1990;43:783-9.
17 Shekelle RB, Shryock AM, Paul O, Lepper M, Stamler J, Liu S, et al. Diet, serum cholesterol, and death from coronary heart disease. The Western Electric study. N Engl J Med 1981;304:65-70.
18 Harlan WR, Oberman A, Mitchell RE, Graybiel A. Constitutional and environmental factors related to serum lipid and lipoprotein levels. Ann Intern Med 1967;66:540-55.
19 Castelli WP, Garrison RJ, Wilson PWF, Abbott RD, Kalousdian S, Kannel WB. Incidence of coronary heart disease and lipoprotein cholesterol levels: the Framingham study. JAMA 1986;256:2835-8.
20 American Health Foundation. Conference on the health effects of blood lipids: epidemiological section. Prev Med 1979;8:625-7.
21 Cremer P, Elster H, Labrot B, Kruse B, Muche R, Seidel D. Incidence rates of fatal and nonfatal myocardial infarction in relation to the lipoprotein profile: first prospective results from the Gottingen risk, incidence, and prevalence study (GRIPS). Klin Wochenschr 1988;66(XI suppl):42-9.

Public health implications

- The association between serum cholesterol concentration and death from ischaemic heart disease is stronger than directly inferred from prospective (cohort) epidemiological studies because of two sources of underestimation that affect these studies

- Correction for the underestimation makes the association about half as strong again: a 30% reduction in ischaemic heart disease at age 60, instead of 20%, for a 10% reduction in serum cholesterol concentration

- The effect of underestimation was quantified in this study and used to correct the results from other prospective studies

- No excess mortality from any cause was apparent in men with low cholesterol concentrations