BMJ

RESEARCH

Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data

C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC)

Correspondence to:
crpgenetics@phpc.cam.ac.uk

Cite this as: *BMJ* 2011;342:d548
doi:10.1136/bmj.d548

ABSTRACT

Objective To use genetic variants as unconfounded proxies of C reactive protein concentration to study its causal role in coronary heart disease.

Design Mendelian randomisation meta-analysis of individual participant data from 47 epidemiological studies in 15 countries.

Participants 194 418 participants, including 46 557 patients with prevalent or incident coronary heart disease. Information was available on four *CRP* gene tagging single nucleotide polymorphisms (rs3093077, rs1205, rs1130864, rs1800947), concentration of C reactive protein, and levels of other risk factors.

Main outcome measures Risk ratios for coronary heart disease associated with genetically raised C reactive protein versus risk ratios with equivalent differences in C reactive protein concentration itself, adjusted for conventional risk factors and variability in risk factor levels within individuals.

Results *CRP* variants were each associated with up to 30% per allele difference in concentration of C reactive protein ($P<10^{-34}$) and were unrelated to other risk factors. Risk ratios for coronary heart disease per additional copy of an allele associated with raised C reactive protein were 0.93 (95% confidence interval 0.87 to 1.00) for rs3093077; 1.00 (0.98 to 1.02) for rs1205; 0.98 (0.96 to 1.00) for rs1130864; and 0.99 (0.94 to 1.03) for rs1800947. In a combined analysis, the risk ratio for coronary heart disease was 1.00 (0.90 to 1.13) per 1 SD higher genetically raised natural log (ln) concentration of C reactive protein. The genetic findings were discordant with the risk ratio observed for coronary heart disease of 1.33 (1.23 to 1.43) per 1 SD higher circulating ln concentration of C reactive protein in prospective studies (P=0.001 for difference).

Conclusion Human genetic data indicate that C reactive protein concentration itself is unlikely to be even a modest causal factor in coronary heart disease.

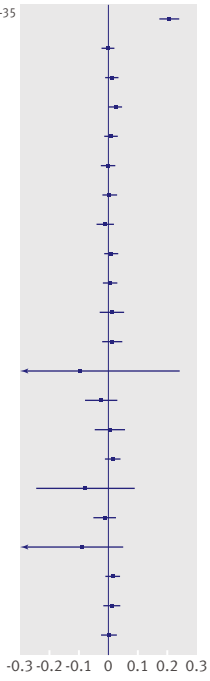
INTRODUCTION

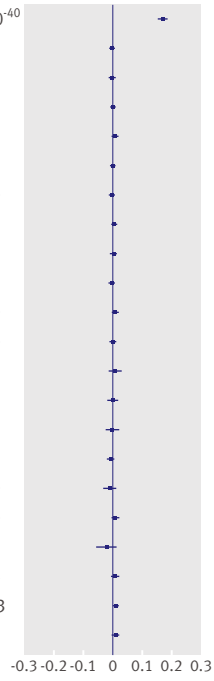
Persistent inflammation has been implicated in the pathogenesis of coronary heart disease, but causality has not been established for any specific inflammatory mediator.[1] C reactive protein, an acute phase protein produced by the liver, is the most extensively studied systemic marker of inflammation.[2] Observational epidemiological studies have shown that C reactive protein concentration is log linearly related to risk of subsequent coronary heart disease, though this association depends considerably on levels of conventional risk factors.[3] C reactive protein binds to low density lipoproteins[4] and is present in atherosclerotic plaques.[5] There is, therefore, interest in whether long term average ("usual") concentration of C reactive protein is itself causally relevant to coronary heart disease.[6 7] Randomised trials of interventions specific to C reactive protein, however, have not yet been done in relation to vascular disease outcomes.

In the absence of such trials, focused genetic studies can be used to help judge causality. This approach is known as "mendelian randomisation" because it is based on Mendel's second law, which states that alleles of different genes assort independently of one another during gamete formation. Consequently, mendelian randomisation analyses are based on Mendel's observation that inheritance of one trait should be independent of inheritance of other traits.[8] For the causal assessment of C reactive protein, a mendelian randomisation analysis should reduce confounding, provided the genetic variants used as proxies for concentration of C reactive protein are unrelated to conventional vascular risk factors and other disease markers. Such studies should also avoid distortions caused by factors occurring later in life (such as the onset of disease) because genetic variants are fixed at conception.[8] Hence, mendelian randomisation analyses should confer certain design advantages akin to those in randomised trials. When applied to other risk factors in coronary heart disease, this approach has previously confirmed the causal relevance of low density lipoprotein cholesterol,[9] increased the likelihood of causality for Lp(a) lipoprotein,[10] and reduced the likelihood of causality for fibrinogen.[11]

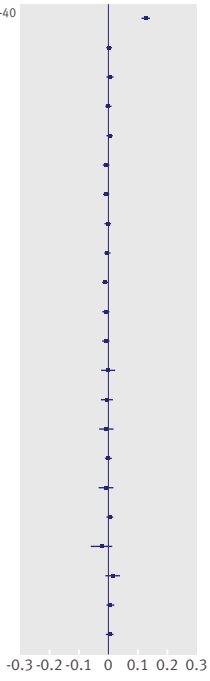
Findings from previous human genetic studies have reduced the likelihood of a major causal role for C reactive protein concentration in coronary heart disease.[12-18] As most known genetic variants related to C reactive protein account for a relatively small portion of the variability in its concentration, however, even larger and more detailed analyses are needed to

Variable	rs3093077 (frequency of risk allele: G = 0.06)				rs1205 (frequency of risk allele: C = 0.67)			
	No of studies/ participants	P value	SD (95% CI) change in biomarker per allele change in SNP	SD (95% CI) change in biomarker per allele change in SNP	No of studies/ participants	P value	SD (95% CI) change in biomarker per allele change in SNP	SD (95% CI) change in biomarker per allele change in SNP
Ln C reactive protein (mg/L)	15/70 117	5.44×10^{-35}		0.207 (0.174 to 0.239)	30/105 476	1.00×10^{-40}		0.169 (0.153 to 0.185)
Age at survey (years)	18/81 648	0.83		-0.002 (-0.024 to 0.019)	37/129 717	0.49		-0.003 (-0.011 to 0.005)
BMI (kg/m^2)	16/73 663	0.34		0.011 (-0.012 to 0.034)	33/116 191	0.64		-0.003 (-0.013 to 0.008)
Systolic BP (mm Hg)	16/74 309	0.04		0.024 (0.001 to 0.047)	31/116 464	0.98		0 (-0.009 to 0.009)
Diastolic BP (mm Hg)	16/74 292	0.46		0.009 (-0.015 to 0.032)	31/116 439	0.22		0.006 (-0.004 to 0.017)
Total cholesterol (mmol/L)	16/72 938	0.91		-0.001 (-0.026 to 0.023)	33/111 422	0.75		-0.001 (-0.010 to 0.007)
Non-HDL cholesterol (mmol/L)	16/70 969	0.71		0.004 (-0.019 to 0.028)	33/109 362	0.40		-0.004 (-0.013 to 0.005)
HDL cholesterol (mmol/L)	16/70 971	0.44		-0.011 (-0.040 to 0.017)	33/109 404	0.32		0.005 (-0.004 to 0.014)
Ln triglycerides (mmol/L)	16/70 476	0.42		0.01 (-0.014 to 0.033)	32/103 906	0.78		0.002 (-0.010 to 0.013)
LDL cholesterol (mmol/L)	16/68 247	0.69		0.005 (-0.019 to 0.029)	32/101 308	0.39		-0.004 (-0.013 to 0.005)
Apolipoprotein A I (g/L)	8/58 678	0.57		0.012 (-0.029 to 0.053)	14/74 525	0.20		0.007 (-0.004 to 0.018)
Apolipoprotein B (g/L)	8/58 841	0.45		0.013 (-0.020 to 0.046)	16/76250	0.90		-0.001 (-0.011 to 0.010)
Albumin (g/L)	1/2436	0.57		-0.097 (-0.437 to 0.242)	10/21 480	0.51		0.007 (-0.015 to 0.029)
Lp(a) lipoprotein (mg/dL)	3/16 577	0.37		-0.025 (-0.079 to 0.029)	13/26 953	0.93		-0.001 (-0.019 to 0.017)
Ln interleukin 6 (ng/L)	6/13 274	0.83		0.006 (-0.045 to 0.056)	13/21 810	0.87		-0.002 (-0.024 to 0.021)
Fibrinogen (µmol/L)	13/64 190	0.30		0.014 (-0.013 to 0.041)	23/81193	0.22		-0.007 (-0.019 to 0.005)
Ln leucocyte count (×10^9/L)	2/2938	0.36		-0.078 (-0.246 to 0.089)	9/18 332	0.30		-0.012 (-0.034 to 0.010)
Glucose (mmol/L)	12/60 961	0.48		-0.014 (-0.051 to 0.024)	23/83 707	0.20		0.008 (-0.004 to 0.021)
Smoking amount (pack years)	2/926	0.14		-0.151 (-0.350 to 0.048)	9/7534	0.21		-0.021 (-0.055 to 0.012)
Weight (kg)	14/68 760	0.21		0.015 (-0.009 to 0.038)	29/106 574	0.36		0.006 (-0.007 to 0.020)
Height (cm)	14/70 385	0.44		0.011 (-0.017 to 0.040)	29/108 939	0.003		0.013 (0.004 to 0.022)
Waist:hip ratio	8/62 358	0.97		0.001 (0.025 to 0.025)	20/91 199	0.15		0.009 (-0.003 to 0.020)





Variable	rs1130864 (frequency of risk allele: T = 0.30)				rs1800947 (frequency of risk allele: G = 0.94)			
	No of studies/ participants	P value	SD (95% CI) change in biomarker per allele change in SNP	SD (95% CI) change in biomarker per allele change in SNP	No of studies/ participants	P value	SD (95% CI) change in biomarker per allele change in SNP	SD (95% CI) change in biomarker per allele change in SNP
Ln C reactive protein (mg/L)	30/98 411	1.00×10^{-40}		0.127 (0.113 to 0.141)	19/38 573	1.00×10^{-40}		0.232 (0.202 to 0.261)
Age at survey (years)	35/118 917	0.37		0.004 (-0.005 to 0.013)	25/58 385	0.17		-0.019 (-0.045 to 0.008)
BMI (kg/m^2)	32/109 149	0.23		0.007 (-0.004 to 0.017)	22/47 509	0.28		-0.015 (-0.042 to 0.012)
Systolic BP (mm Hg)	31/109 799	0.97		0 (-0.009 to 0.009)	20/48 515	0.98		0 (-0.032 to 0.031)
Diastolic BP (mm Hg)	31/109 780	0.30		0.005 (-0.004 to 0.014)	20/48 502	0.90		0.002 (-0.025 to 0.028)
Total cholesterol (mmol/L)	32/104 375	0.14		-0.007 (-0.016 to 0.002)	22/42 737	0.49		0.013 (-0.023 to 0.049)
Non-HDL cholesterol (mmol/L)	31/101 822	0.15		-0.007 (-0.016 to 0.002)	22/41 504	0.99		0 (-0.043 to 0.043)
HDL cholesterol (mmol/L)	31/101 845	0.99		0 (-0.009 to 0.009)	22/41 545	0.10		0.024 (-0.005 to 0.053)
Ln triglycerides (mmol/L)	31/96 995	0.55		-0.003 (-0.013 to 0.007)	22/42 262	0.52		-0.014 (-0.057 to 0.029)
LDL cholesterol (mmol/L)	30/93 902	0.05		-0.010 (-0.020 to 0)	22/40 806	0.64		0.009 (-0.029 to 0.047)
Apolipoprotein A I (g/L)	14/70 707	0.22		-0.007 (-0.018 to 0.004)	9/19 703	0.88		-0.003 (-0.045 to 0.039)
Apolipoprotein B (g/L)	16/72 548	0.16		-0.008 (-0.019 to 0.003)	11/21 272	0.65		0.010 (-0.034 to 0.054)
Albumin (g/L)	8/17 970	0.99		0 (-0.023 to 0.023)	9/18 078	0.89		0.003 (-0.040 to 0.046)
Lp(a) lipoprotein (mg/dL)	11/23 485	0.66		-0.004 (-0.024 to 0.015)	9/10 702	0.10		-0.050 (-0.109 to 0.010)
Ln interleukin 6 (ng/L)	12/17 694	0.57		-0.007 (-0.029 to 0.016)	8/13 970	0.93		-0.002 (-0.050 to 0.046)
Fibrinogen (µmol/L)	24/79 166	0.93		0 (-0.010 to 0.011)	12/17 892	0.96		0.001 (-0.042 to 0.044)
Ln leucocyte count (×10^9/L)	8/15 731	0.53		-0.008 (-0.032 to 0.016)	6/10 550	0.97		0.001 (-0.054 to 0.056)
Glucose (mmol/L)	22/80 236	0.26		0.006 (-0.005 to 0.017)	15/25 563	0.92		0.002 (-0.034 to 0.038)
Smoking amount (pack years)	10/7675	0.21		-0.022 (-0.057 to 0.012)	6/6477	0.36		-0.036 (-0.113 to 0.041)
Weight (kg)	28/99 410	0.21		0.007 (-0.004 to 0.019)	20/42 650	0.30		-0.018 (-0.052 to 0.016)
Height (cm)	28/101 720	0.36		0.005 (-0.006 to 0.017)	20/44 760	0.88		-0.002 (-0.030 to 0.026)
Waist:hip ratio	18/86 541	0.22		0.015 (-0.009 to 0.038)	11/26 347	0.89		0.003 (-0.034 to 0.039)



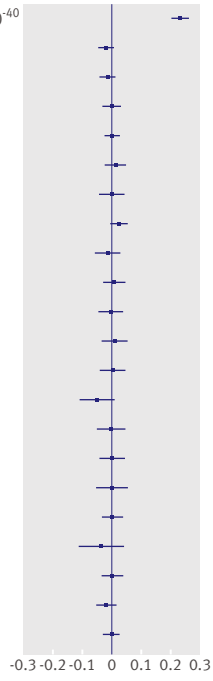


Fig 1 |Associations of four principal single nucleotide polymorphisms (SNP) related to C reactive protein with various characteristics in individuals free from known coronary heart disease at time of measurement. Estimates presented are based on random effects meta-analysis of study specific associations of each SNP with panel of risk factors, adjusted, where appropriate, for ethnicity. Per allele model corresponds to association per addition of risk allele for each SNP

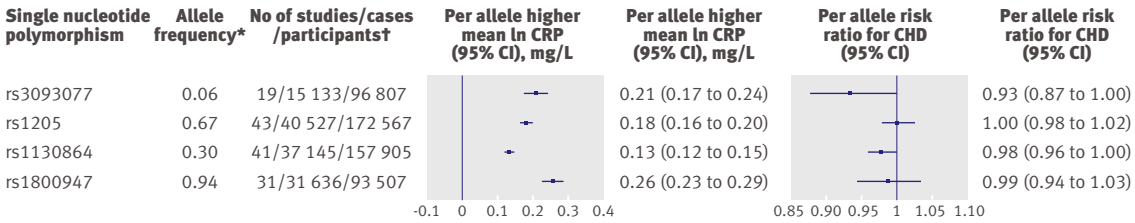


Fig 2 | Estimates of association of each single nucleotide polymorphism with ln concentrations of C reactive protein and risk of coronary heart disease (CHD). *Frequency of allele for increased concentrations of circulating ln C reactive protein (that is, risk allele). Associations presented per additional copy of risk allele. †For associations between single nucleotide polymorphism and coronary heart disease, studies with <10 cases or <50 participants were excluded from analyses. Study specific estimates stratified, where appropriate, by sex, ethnicity, and trial arm and combined with random effects models. Maximum available data on genetic variants, circulating C reactive protein, and coronary heart disease used for analyses; sensitivity analyses restricted to participants with data on C reactive protein single nucleotide polymorphisms, circulating C reactive protein, and coronary heart disease did not differ from current analyses. Fig C in appendix 3 on bmj.com shows study specific associations between single nucleotide polymorphism and C reactive protein and coronary heart disease for each single nucleotide polymorphism

reliably assess the possibility of any moderate causal role. We report such an analysis based on individual data from 194 418 participants in 47 epidemiological studies. We studied these genetic variants in relation to C reactive protein concentration, other risk factors, and risk of coronary heart disease.

METHODS

Design and rationale

The study had five inter-related components. Firstly, we selected, a priori, a panel of four single nucleotide polymorphisms that explain 98% of the variation in the *CRP* gene in populations of European descent. These variants have been shown to influence concentration of C reactive protein without affecting its protein sequence.[19] Secondly, we studied whether these polymorphisms are likely to be exclusively associated with C reactive protein concentration by evaluating them in relation to a range of other risk factors. Thirdly, we calculated risk ratios for coronary heart disease with genetically raised concentration using information on these *CRP* variants. Fourthly, we calculated risk ratios for coronary heart disease with circulating C reactive protein concentration after adjustment for conventional risk factors and variability in risk factors within individuals. Fifthly, we compared risk ratios for coronary heart disease with genetically raised concentration of C reactive protein versus risk ratios seen with equivalent differences in circulating concentrations.

Genetic variants

We used detailed information about the composition of the *CRP* gene to select a parsimonious set of "tagging" single nucleotide polymorphisms (rs3093077, rs1205, rs1130864, and rs1800947) that fully covers the common variations of this gene in populations of European descent (that is, minor allele frequency ≥0.05 and an r^2 threshold of ≥0.8).[19 20] Data available on 36 further single nucleotide polymorphisms enabled use of proxy variants when principal polymorphisms were not measured. To enhance power, we also studied combinations of alleles inherited together, or "haplotypes" (see table A in appendix 1 on bmj.com).

Contributing studies

Details of the C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC) have been described previously.[20] Tables B-E in appendix 1 on bmj.com provides details of contributing studies, and appendix 2 lists study acronyms. Studies were identified through registry approaches and systematic searches of the literature (see fig A in appendix 3, and appendix 4 on bmj.com). Individual data were supplied on 194 418 participants, including 46 557 with incident or prevalent coronary heart disease, in 47 studies. Studies used different genotyping platforms: 23 used TaqMan assays, three used KASPAR technology (KBioscience), three used restriction fragment length polymorphism, 10 used the ITMAT-Broad-CARe 50K SNP array, and eight used other multiplex methods. Thirty five studies measured C reactive protein with high sensitivity assays, directly or indirectly standardised on the International Reference Standard for C reactive protein immunoassay (WHO 85/506). The outcome was defined as fatal coronary heart disease (based on International Classification of Diseases codes), non-fatal myocardial infarction (using WHO criteria), or coronary stenosis (>50% narrowing of one of more coronary arteries assessed by angiography). All study participants provided written informed consent for use of their DNA for genetic testing.

Statistical analyses

Appendix 5 on bmj.com provides details of the statistical methods. Levels of C reactive protein and other positively skewed variables were natural log transformed. Principal analyses assumed additive effects (per allele associations), with subsidiary analyses of other genetic models. All analyses of circulating C reactive protein and other risk factors susceptible to reverse association biases were limited to participants without known coronary heart disease at time of blood sampling. We calculated study specific associations of baseline (and repeat) values of C reactive protein concentration and other characteristics with a linear regression model, adjusted for sex and ethnicity and combined across studies using

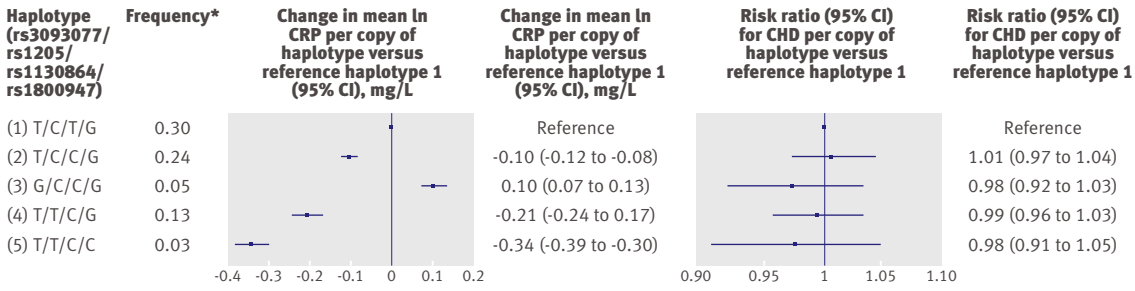


Fig 3 | Estimates of association of each haplotype with ln concentrations of CRP and risk of coronary heart disease. *Based on seven haplotypes and therefore do not add up to 1. Haplotypes 6 and 7 excluded because they represent individuals for whom it was not possible to assign to primary haplotypes because of missing data in rs3093077 and rs1800947. See table A in appendix 1, fig F in appendix 3, and appendix 5 for details of these haplotypes and explanation of methods. Additive haplotype model was used to estimate effect of each haplotype relative to two copies of haplotype 1. See appendix 5 for details of this model. Data limited to populations of European descent, for which it was possible to assign haplotypes based on linkage disequilibrium information between single nucleotide polymorphisms (see appendix 5). Studies with <10 cases or <50 participants excluded. Study specific estimates were stratified, where appropriate, by sex, ethnicity, and trial arm and combined with multivariate random effects meta-analysis. Data available on up to 25 960 cases and up to 139 251 participants of European descent from 33 studies. Fig G in appendix 3 shows study specific associations between haplotype and C reactive protein and coronary heart disease for each haplotype

random effects meta-analysis to obtain summary estimates of single nucleotide polymorphism and C reactive protein correlates.[21] We calculated risk ratios for coronary heart disease per addition of a prespecified risk allele of each single nucleotide polymorphism using a two stage approach, stratified, where appropriate, by sex, ethnicity, and adjustment for age. We calculated natural log (ln) risk ratios separately within each study before pooling them across studies by random effects meta-analysis.[3] For prospective cohort studies, we used stratified Cox proportional hazard regression models to calculate hazard ratios for incident coronary heart disease risk. For retrospective studies (and for prevalent coronary heart disease cases in prospective studies), we used either conditional or unconditional logistic regression as appropriate to calculate odds ratios. Hazard ratios and odds ratios were assumed to approximate the same risk ratios.

To correct risk ratios for coronary heart disease for potential bias caused by measurement error and variability in risk factors within an individual ("regression dilution bias"), we made allowance for regression dilution bias in both C reactive protein concentration and potential confounding factors. Regression dilution ratios were obtained by regressing 93 998 serial measurements (mean interval 2.9 (SD 1.9) years) from 35 023 participants on baseline levels of the relevant characteristic, adjusted for conventional risk factors plus baseline ln C reactive protein concentration and duration of follow-up.[22] Risk ratios for coronary heart disease, adjusted for conventional risk factors, were calculated per 1 SD higher usual ln concentration in participants. This difference corresponds to an approximately threefold difference on the original scale of C reactive protein measurement, as the pooled standard deviation for baseline ln C reactive protein is 1.05 (that is, $e^{1.05}$). To estimate the effect of genetically predicted C reactive protein on coronary heart disease, we used a hierarchical Bayesian meta-analysis framework.[23] Using information on baseline ln C reactive protein concentration within each genotype or haplotype group, we obtained a summary estimate from meta-analysis of study specific risk ratios for coronary heart disease per 1 SD higher genetically predicted C reactive protein concentration. We used I^2 statistic to assess heterogeneity.[24] Diversity at the study level was investigated by grouping studies by relevant characteristics and by meta-regression. Effect modification by several prespecified variables was investigated by formal tests of interaction. Conventional analyses were conducted in Stata v 11.0 and Bayesian analyses in WinBUGS.

RESULTS

Mean age at entry was 59 years (SD 10), 89% of participants were of European descent, and 44% were women. Minor allele frequencies ranged from 0.06 to 0.33 for the principal single nucleotide polymorphisms. For any given polymorphism, minor allele frequencies were similar across studies. Mean baseline concentrations of C reactive protein varied across studies, though standard deviations were broadly similar (see fig B in appendix 3 on bmj.com), with an overall median of 1.7 (5th, 95th centile 0.3, 12.7) mg/L. The regression dilution ratio of ln C reactive protein, adjusted for age and sex, was 0.57 (95% confidence interval 0.51 to 0.64), similar to those observed here for systolic blood pressure (0.51, 0.48 to 0.54), and total cholesterol (0.55, 0.52 to 0.60).

CRP variants, C reactive protein concentration, and levels of other risk factors

Each of these *CRP* variants was associated with baseline C reactive protein, with per allele differences in C reactive protein concentration of 23% (95% confidence interval 19% to 27%) for rs3093077, 19% (17% to 21%) for rs1205, 14% (12% to 16%) for rs1130864, and 30% (26% to 34%) for rs1800947 (P<0.001; fig 1) (see also fig C in appendix 3 and table F in appendix 1 on bmj.com). These associations were consistent over

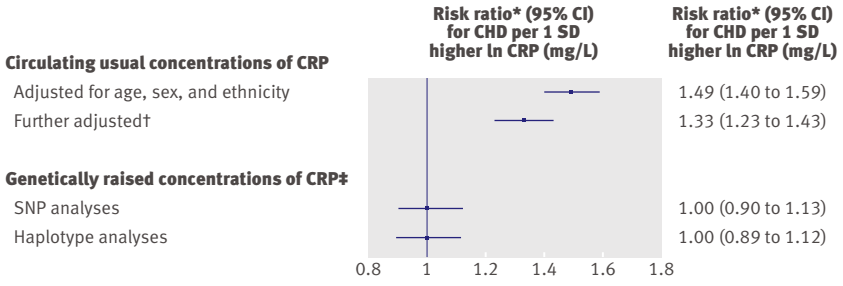


Fig 4 |Estimates of association of circulating concentrations and genetically raised concentrations of C reactive protein (CRP) with risk of coronary heart disease (CHD). *Corrected for regression dilution in C reactive protein and potential confounding factors.

multiple repeat measures of circulating C reactive protein taken several years apart (see fig D in appendix 3 on bmj.com). With the exception of ethnicity (with which the principal variants were strongly correlated; P<0.001), single genetic variants associated with C reactive protein were not associated with conventional risk factors or other inflammatory markers (fig 1). With haplotype 1 as a reference, *CRP* haplotypes were associated with differences in ln C reactive protein concentration ranging from 0.1 to 0.34 mg/L and were not associated with conventional risk factors and other characteristics (see figs E-G in appendix 3 and table F in appendix 1 on bmj.com). By contrast, baseline C reactive protein concentration itself was associated with conventional risk factors, inflammatory markers, and other characteristics (see table G in appendix 1 on bmj.com).

Risk ratios for coronary heart disease with *CRP* genetic variants

Risk ratios for coronary heart disease per addition of a risk ("C reactive protein raising") allele were 0.93 (0.87 to 1.00) for rs3093077, 1.00 (0.98 to 1.02) for rs1205, 0.98 (0.96 to 1.00) for rs1130864, and 0.99 (0.94 to 1.03) for rs1800947 (fig 2). There was modest heterogeneity in these estimates (I^2 values for those risk ratios were 0%, 26%, 0%, and 0%, respectively, see fig C in appendix 3 on bmj.com) with similar findings under a range of circumstances (see table H in appendix 1 on bmj.com). For haplotype analyses, risk ratios for coronary heart disease per copy of haplotype (relative to two copies of haplotype 1) were 1.01 (0.97 to 1.04) with haplotype 2, 0.98 (0.92 to 1.03) with haplotype 3, 0.99 (0.96 to 1.03) with haplotype 4, and 0.98 (0.91 to 1.05) with haplotype 5 (fig 3) (see also fig F in appendix 3 on bmj.com). There was little heterogeneity in these risk ratios (see fig G in appendix 3 on bmj.com). Data were insufficient to investigate effects in different ethnic groups.

Risk ratios for coronary heart disease with usual concentrations of C reactive protein

In analyses restricted to 27 long term prospective studies comprising 124 931 participants and 10 981 incident cases of coronary heart disease, there was an approximately log linear association between C reactive protein concentration and risk of coronary heart disease (see fig H in appendix 3 on bmj.com). In these studies, the risk ratio for coronary heart disease, adjusted for age, sex, and ethnicity only, was 1.49 (1.40 to 1.59) per 1 SD higher "usual" ln C reactive protein concentration (that is, a risk ratio that has made allowance for regression dilution) (fig 4). The risk ratio for coronary heart disease was 1.33 (1.23 to 1.43) after further adjustment for smoking status, history of diabetes mellitus, and usual levels of systolic blood pressure, body mass index (BMI), non-high density lipoprotein cholesterol, high density lipoprotein cholesterol, and triglyceride concentrations. There was little evidence of heterogeneity (see fig I in appendix 3 on bmj.com). Risk ratios with higher C reactive protein concentration were broadly similar under a range of circumstances (see fig J in appendix 3 on bmj.com). Multivariable adjusted risk ratios for coronary heart disease with C reactive protein concentration weakened after further adjustment for fibrinogen or interleukin 6 (see table I in appendix 1 on bmj.com).

Integration of data on *CRP* variants and C reactive protein concentration

Risk ratios for coronary heart disease were 1.00 (0.90 to 1.13) per 1 SD higher genetically raised C reactive protein as determined by all four principal single nucleotide polymorphisms (163 174 participants, 37 736 cases, 44 studies). This corresponds to a risk ratio of 1.00 (0.97 to 1.02) per 20% lower C reactive protein, which is equivalent to about 0.38 mg/L lower C reactive protein when the population mean is 1.88 mg/L. From information on all common *CRP* haplotypes in populations of European descent, the risk ratio was 1.00 (0.89 to 1.12; 152 678 participants, 33 589 cases, 39 studies; fig 4) (see also fig K in appendix 3 on bmj.com).

We observed qualitatively similar results to those reported above in analyses that used different genetic models (see fig L in appendix 3 on bmj.com); excluded variants or studies that deviated from Hardy-Weinberg equilibrium or were judged to be insufficiently strong genetic instruments[25]; used fixed effect meta-analysis models (see figs C, G, I, and K in appendix 3 on bmj.com); omitted the 11 734 participants who reported using cardiovascular drugs (including statins) at baseline; omitted the 18 198 participants in clinical trials; included only people of European descent (see fig M in appendix 3 on bmj.com); compared prospective and retrospective studies (fig K in appendix 3 on bmj.com); and compared larger studies versus smaller studies (available on request).

DISCUSSION

Implications for disease aetiology

Using the principle of mendelian randomisation, we used *CRP* genetic variants as proxies to help judge whether usual C reactive protein concentration is causally relevant to coronary heart disease. Our results indicate that genetically raised concentrations of C reactive protein are unrelated to conventional risk factors and risk of coronary heart disease. Given the power of our study, these results suggest that C reactive protein concentration is

WHAT IS ALREADY KNOWN ON THIS TOPIC

Blood concentrations of C reactive protein are strongly and continuously associated with future risk of coronary heart disease, though it is unknown whether this correlation reflects cause and effect

Genetic variants related to C reactive protein can be used as proxies for C reactive protein concentration to help judge causality ("mendelian randomisation analyses")

Previous studies have been insufficiently powerful and detailed to evaluate the possibility of any moderate causal role for C reactive protein in coronary heart disease

WHAT THIS STUDY ADDS

With individual data from almost 200 000 people (including almost 47 000 with coronary heart disease), the current study has shown that genetically raised concentration of C reactive protein is unrelated to conventional risk factors and risk of coronary heart disease

Human genetic data indicate that C reactive protein concentration itself is unlikely to be even a modest causal factor in coronary heart disease

unlikely to have even a modest causal role in coronary heart disease. Furthermore, we found no material associations between *CRP* genetic variants and fibrinogen, interleukin 6, and leucocyte count. Because these results suggest that C reactive protein concentration is unlikely to be causally relevant to these inflammatory markers, the findings imply that analyses of C reactive protein concentration and disease outcomes should adjust for these inflammatory markers (that is, treat them as genuine confounders). Indeed, risk ratios for coronary heart disease with circulating C reactive protein were further weakened when we adjusted for such inflammatory markers. This finding further decreases the likelihood that C reactive protein concentration is causally relevant.[3] Collectively, therefore, our results suggest that studies seeking to test the inflammation hypothesis in coronary heart disease should examine inflammatory mediators other than C reactive protein.[1 26 27]

Study strengths

Our analysis builds on previous studies, combining several major advantages.[12-18] Firstly, we carefully selected and assessed specific polymorphisms to capture the full range of common variability in the *CRP* gene. Variants in the *CRP* gene related to C reactive protein concentration are more likely to be exclusively associated with C reactive protein concentration than variants in other genes.[18] Secondly, our study involved over 90 000 more participants than the previous largest study, substantially enhancing statistical power. Thirdly, the availability of data from individual participants enabled a comprehensive and detailed range of analyses not possible with aggregated data (such as haplotype analyses and instrumental variables analyses[23]). For example, compared with analysis of single nucleotide polymorphisms, haplotype analysis allows more complete consideration of genetic variation at a locus. Fourthly, we had information on 94 000 serial measurements, enabling correction of risk ratios for regression dilution bias[22] and yielding results supporting the idea that *CRP* variants are related to lifetime variation in C reactive protein concentration. Finally, the generalisability of our results is supported by their consistency in 47 studies in 15 countries.[24]

Potential limitations

For studies that did not directly measure the principal *CRP* single nucleotide polymorphisms we investigated, we used alternative variants as proxies. There is the possibility of residual confounding by unrecognised effects of genotypes on other risk factors and by adaptation during early life to compensate for genetically raised concentrations of C reactive protein,[8] though there is no evidence of their impact in the current context. Further study is needed in non-white people, particularly as the genetic architecture of *CRP* differs by ethnicity.[28]

Broader context

Our results generally agree with those from previous studies that used other methods, such as the null findings of in vivo studies of atherosclerosis that have involved either injection of C reactive protein in different species or transgenic expression of C reactive protein in mice and rabbits.[29 30] Our results also agree with well designed in vitro studies, in which uncontaminated preparations of C reactive protein were used, which have generally failed to produce persuasive evidence for pro-atherothrombotic effects of C reactive protein on various cell types.[31 32] Irrespective of the causal relevance of C reactive protein itself to coronary heart disease, however, there is considerable evidence that persistent inflammation might contribute to coronary heart disease.[1 33] Hence, there is a need to identify specific genetic, biochemical, and environmental determinants responsible for these associations. Anti-inflammatory interventions have not yet been adequately studied to test the inflammation hypothesis in coronary heart disease.[1 26 27]

Furthermore, as distinct from possible associations of underlying usual C reactive protein concentration with later development of coronary heart disease, there is interest in the possibility that massively raised concentrations of C reactive protein at the time of acute ischaemic events could contribute to severity and outcome of ischaemic lesions.[34] Our findings also do not address the separate issue of the value of measurement of circulating C reactive protein in prediction of long term vascular risk.[35]

Writing committee
Frances Wensley, Pei Gao (University of Cambridge); Stephen Burgess (MRC Biostatistics Unit, Cambridge,); Stephen Kaptoge, Emanuele Di Angelantonio (University of Cambridge); Tina Shah (University College London); James C Engert (McGill University, Canada); Robert Clarke (University of Oxford); George Davey-Smith (University of Bristol); Børge G Nordestgaard (University of Copenhagen, Denmark); Danish Saleheen (University of Cambridge); Nilesh J Samani (University of Leicester); Manjinder Sandhu (University of Cambridge); Sonia Anand (McMaster University, Canada); Mark B Pepys (University College London); Liam Smeeth, John Whittaker, Juan Pablo Casas (London School of Hygiene and Tropical Medicine, UK); Simon G Thompson (MRC Biostatistics Unit, Cambridge); Aroon D Hingorani (University College London); John Danesh (University of Cambridge).
C reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC) Investigators

AGES: G Eiriksdottir, T B Harris, L J Launer, V Gudnason; ARIC: A R Folsom, G Andrews, C M Ballantyne; BHF-FHS: N J Samani, A S Hall, P S Braund, A J Balmforth; BRHS: P H Whincup, R Morris; BWHHS: D A Lawlor, G D O Lowe, N Timpson, S Ebrahim; CAPS: Y Ben-Shlomo, G Davey-Smith, N Timpson; CCHS, CGPS, and CIHDS: B G Nordestgaard, A Tybjærg-Hansen, J Zacho; CHAOS: M Brown, M Sandhu, S L Ricketts, S Ashford; CHS: L Lange, A Reiner, M Cushman, R Tracy (see www.chs-nhlbi.org for acknowledgments); CRPHANS: C Wu, J Ge, Y Zou, A Sun; CUDAS and CUPID: J Hung, B McQuillan, P Thompson, J Beilby, N Warrington, L J Palmer; 4D: C Wanner, C Drechsler, M M Hoffmann; EAS: F G R Fowkes, G D O Lowe, I Tzoulaki; ELSA: M Kumari, M Miller, M Marmot; EPICNL: C Onland-Moret, Y T van der Schouw, J M Boer, C Wijmenga; EPICNOR: S L Ricketts, S Ashford, M Sandhu, K-T Khaw; FRAMOFF: R S Vasan, R B Schnabel, J F Yamamoto, E J Benjamin; GERMIFS: H Schunkert, J Erdmann, I R König, C Hengstenberg; GISSI-P: B Chiodini, M G Franzosi, S Pietri, F Gori; HEALTHABC: M Rudock; Y Liu, K Lohman, T B Harris; HIFMECH: S E Humphries, A Hamsten; HIMS: P E Norman, G J Hankey, K Jamrozik, L J Palmer; HPFS: E B Rimm, J K Pai; HVHS: B M Psaty, S R Heckbert, J C Bis; INTERHEART: S Yusuf, S Anand, J C Engert, C Xie; ISIS: R Collins, R Clarke, D Bennett; LOLIPOP: J Kooner, J Chambers, P Elliott; LURIC: W März, M E Kleber, B O Böhm, B R Winkelmann; MALMO: O Melander, G Berglund; MONICAKORA: W Koenig, B Thorand, J Baumert, A Peters; NHS: E B Rimm, J Manson, J K Pai; NPHSII: S E Humphries, J A Cooper, P J Talmud; NSC: P Ladenvall, L Johansson, J-H Jansson, G Hallmans; PENNCATH: M P Reilly, L Qu, M Li, D J Rader; PROCARDIS: H Watkins, R Clarke, J Hopewell; PROMIS: D Saleheen, J Danesh, P Frossard; PROSPER: N Sattar, M Robertson, J Shepherd, E Schaefer; ROTTERDAM: A Hofman, J C M Witteman, I Kardys, A Dehghan; SHEEP: U de Faire, A Bennet, B Gigante, K Leander; SPEED: Y Ben-Shlomo, G Davey-Smith, N Timpson; UCP: B Peters, A H Maitland-van der Zee, A de Boer, O Klungel; WHIOS: A Reiner, J Manson, P Greenland, J Dai, S Liu; WHITEII: M Kumari, E Brunner, M Kivimaki, M Marmot; WOSCOPS: N Sattar, D O'Reilly, I Ford, C J Packard.
We thank C Hennekens for helpful comments.
Contributors: Statistical team: S G Thompson (co-chair), J Whittaker (co-chair), S Burgess, P Gao, S Kaptoge, C Verzilli, and F Wensley (coordinator). Data management: M Walker and S Watson. Coordinating centre: F Wensley (coordinator), S Anand, S Burgess, J P Casas, E Di Angelantonio, J Engert, P Gao, S Kaptoge, MB Pepys, T Shah, L Smeeth, SG Thompson, C Verzilli, M Walker, S Watson, J Whittaker, S Yusuf, A D Hingorani (co-principal investigator), and J Danesh (co-principal investigator). P Gao, S Burgess, S Kaptoge, and E Di Angelantonio contributed equally to this work, as did AD Hingorani and J Danesh. J Danesh is guarantor.
Funding: The coordinating centre is supported by grant SP/08/007 from the British Heart Foundation. Various sources have supported recruitment, follow-up, and laboratory measurements in the contributing studies. Investigators of several of these studies have contributed to a list naming some of these funding sources, which can be found at http://ceu.phpc.cam.ac.uk/research/ccgc/studies/. AD Hingorani acknowledges support from the Rosetrees Foundation and the British Heart Foundation (FS05/125).
Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.
Ethical approval: This study was approved by the Cambridgeshire ethics review committee, and was conducted independently from its funders.
Data sharing: No additional data available.

1 Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med* 2005;352:1685-95.
2 Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J Clin Invest* 2003;111:1805-12.
3 Emerging Risk Factors Collaboration. C-reactive protein concentration and risk of coronary heart disease, stroke and mortality: an individual participant meta-analysis. *Lancet* 2010;375:132-40.
4 De Beer FC, Soutar AK, Baltz ML, Trayner IM, Feinstein A, Pepys MB. Low density lipoprotein and very low density lipoprotein are selectively bound by aggregated C-reactive protein. *J Exp Med* 1982;156:230-42.
5 Zhang YX, Cliff WJ, Schoefl GI, Higgins G. Coronary C-reactive protein distribution: its relation to development of atherosclerosis. *Atherosclerosis* 1999;145:375-9.
6 Bisoendial RJ, Boekholdt SM, Vergeer M, Stroes ES, Kastelein JJ. C-reactive protein is a mediator of cardiovascular disease. *Eur Heart J* 2010;31:2087-91.
7 Anand SS, Yusuf S. C-reactive protein is a bystander of cardiovascular disease. *Eur Heart J* 2010;31:2092-6.
8 Davey Smith G, Ebrahim S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ* 2005;330:1076-9.
9 Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 2006;354:1264-72.
10 Kamstrup PR, Tybjaerg-Hansen A, Steffensen R, Nordestgaard BG. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA* 2009;301:2331-9.
11 Keavney B, Danesh J, Parish S, Palmer A, Clark S, Youngman L, et al. Fibrinogen and coronary heart disease: test of causality by "Mendelian randomization." *Int J Epidemiol* 2006;35:935-43.
12 Kardys I, de Maat MP, Uitterlinden AG, Hofman A, Witteman JC. C-reactive protein gene haplotypes and risk of coronary heart disease: the Rotterdam Study. *Eur Heart J* 2006;27:1331-7.
13 Lange LA, Carlson CS, Hindorff LA, Lange EM, Walston J, Durda JP, et al. Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. *JAMA* 2006;296:2703-11.
14 Casas JP, Shah T, Cooper J, Hawe E, McMahon AD, Gaffney D, et al. Insight into the nature of the CRP-coronary event association using Mendelian randomization. *Int J Epidemiol* 2006;35:922-31.
15 Pai JK, Mukamal KJ, Rexrode KM, Rimm EB. C-reactive protein (CRP) gene polymorphisms, CRP levels, and risk of incident coronary heart disease in two nested case-control studies. *PLoS One* 2008;3:e1395.
16 Lawlor DA, Harbord RM, Timpson NJ, Lowe GD, Rumley A, Gaunt TR, et al. The association of C-reactive protein and CRP genotype with coronary heart disease: findings from five studies with 4,610 cases amongst 18,637 participants. *PLoS One* 2008;3:e3011.
17 Zacho J, Tybjaerg-Hansen A, Jensen JS, Grande P, Sillesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. *N Engl J Med* 2008;359:1897-908.
18 Elliott P, Chambers JC, Zhang W, Clarke R, Hopewell JC, Peden JF, et al. Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA* 2009;302:37-48.
19 Verzilli C, Shah T, Casas JP, Chapman J, Sandhu M, Debenham SL, et al. Bayesian meta-analysis of genetic association studies with different sets of markers. *Am J Hum Genet* 2008;82:859-72.
20 CRP CHD Genetics Collaboration. Collaborative pooled analysis of data on C-reactive protein gene variants and coronary disease: judging causality by Mendelian randomisation. *Eur J Epidemiol* 2008;23:531-40.
21 Fibrinogen Studies Collaboration. Associations of plasma fibrinogen levels with established cardiovascular disease risk factors, inflammatory markers, and other characteristics: individual participant meta-analysis of 154,211 adults in 31 prospective studies: the Fibrinogen Studies Collaboration. *Am J Epidemiol* 2007;166:867-79.
22 Fibrinogen Studies Collaboration. Correcting for multivariate measurement error by regression calibration in meta-analyses of epidemiological studies. *Stat Med* 2009;28:1067-92.
23 CRP CHD Genetics Collaboration. Bayesian methods for meta-analysis of causal relationships estimated using genetic instrumental variables. *Stat Med* 2010;29:1298-311.
24 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58.
25 Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997;65:557-86.
26 Crossman DC, Morton AC, Gunn JP, Greenwood JP, Hall AS, Fox KA, et al. Investigation of the effect of interleukin-1 receptor antagonist (IL-1ra) on markers of inflammation in non-ST elevation acute coronary syndromes (the MRC-ILA-HEART Study). *Trials* 2008;9:8.
27 Lachmann HJ, Lowe P, Felix SD, Rordorf C, Leslie K, Madhoo S, et al. In vivo regulation of interleukin 1beta in patients with cryopyrin-associated periodic syndromes. *J Exp Med* 2009;206:1029-36.
28 Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M, et al. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *Am J Hum Genet* 2005;77:64-77.
29 Kovacs A, Tornvall P, Nilsson R, Tegner J, Hamsten A, Bjorkegren J. Human C-reactive protein slows atherosclerosis development in a mouse model with human-like hypercholesterolemia. *Proc Natl Acad Sci USA* 2007;104:13768-73.
30 Koike T, Kitajima S, Yu Y, Nishijima K, Zhang J, Ozaki Y, et al. Human C-reactive protein does not promote atherosclerosis in transgenic rabbits. *Circulation* 2009;120:2088-94.

31 Taylor KE, Giddings JC, van den Berg CW. C-reactive protein-induced in vitro endothelial cell activation is an artefact caused by azide and lipopolysaccharide. *Arterioscler Thromb Vasc Biol* 2005;25:1225-30.
32 Casas JP, Shah T, Hingorani AD, Danesh J, Pepys MB. C-reactive protein and coronary heart disease: a critical review. *J Intern Med* 2008;264:295-314.
33 Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med* 2004;351:2611-8.
34 Pepys MB, Hirschfield GM, Tennent GA, Gallimore JR, Kahan MC, Bellotti V, et al. Targeting C-reactive protein for the treatment of cardiovascular disease. *Nature* 2006;440:1217-21.
35 Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO III, Criqui M, et al. Markers of inflammation and cardiovascular disease: application to clinical and public health practice. A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* 2003;107:499-511.

Accepted: 9 November 2010