

High Density Lipoprotein pathway as a therapeutic target for coronary heart disease: individual participant meta-analysis in 28,597 individuals with 4197 coronary events.

**Authors:** Amy R Mulick<sup>\*1</sup> MSc, David Prieto-Merino<sup>1,2</sup> PhD, Therese Tillin<sup>3</sup> MSc, Aki Havulinna<sup>4</sup> DSc(Tech), Martin Shipley<sup>5</sup> MSc, Desiree Valera-Gran<sup>6</sup> PhD, Aleksandra Gentry-Maharaj<sup>7</sup> PhD, Andy Ryan<sup>7</sup> PhD, Meena Kumari<sup>8</sup> PhD, J Wouter Jukema<sup>9</sup> MD PhD FESC FACC, Alex McConnachie<sup>10</sup> PhD, Veikko Salomaa<sup>4</sup> MD PhD Research Professor, Nish Chaturvedi<sup>3</sup> MD, Goya Wannamethee<sup>11</sup> PhD FFPH, Usha Menon<sup>7</sup> MD, Barbara Jefferis<sup>11</sup> PhD, Mika Kivimaki<sup>5</sup> FMedSci, Chris J Packard<sup>12</sup> DSc, Naveed Sattar<sup>12</sup> FMedSci, John Whittaker<sup>13</sup> PhD, Aroon Hingorani<sup>3</sup> PhD, George B Ploubidis<sup>14</sup> PhD, JP Casas<sup>\*15,16</sup> PhD

- (1) Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
- (2) Applied Statistical Methods in Medical Research Group, Universidad Católica San Antonio de Murcia, Spain
- (3) Institute of Cardiovascular Science, University College London, London, UK
- (4) National Institute for Health and Welfare, Helsinki, Finland
- (5) Department of Epidemiology and Public Health, University College London, London, UK
- (6) Department of Pathology and Surgery, Miguel Hernández University, San Juan de Alicante, Spain
- (7) MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, University College London, London, UK
- (8) Institute for Social and Economic Research, University of Essex, Essex, UK
- (9) Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands
- (10) Institute of Health and Wellbeing, Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK

(11)Department of Primary Care & Population Health, University College London, Royal Free

Campus, London, UK

(12)Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

(13) Target Sciences, Glaxo-Smith-Kline, Stevenage, UK

(14) Centre for Longitudinal Studies, UCL Institute of Education, University College London, London,

UK

(15) Institute of Health Informatics, University College London, UK.

(16) Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston

Healthcare System, Boston, MA, USA.

\*authors for correspondence: Amy Mulick, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK [amy.mulick@lshtm.ac.uk](mailto:amy.mulick@lshtm.ac.uk) +44(0)207 299 4803; JP Casas, MAVERIC, VA Boston Healthcare System, 150 S Huntington Av, Jamaica Plain MA 02130, USA  
[juan.casasromero@va.gov](mailto:juan.casasromero@va.gov)

## Key Points

**Question:** Can investigation of sub-components of the high-density lipoprotein (HDL) pathway, measured through nuclear magnetic resonance spectroscopy, point to specific therapeutic targets for prevention of coronary heart disease (CHD)?

**Findings:** Using individual-level data from seven longitudinal studies including 28,597 participants and 4197 CHD events, we identified two components of the HDL pathway that were associated with reduced, and one that was associated with increased, risk of CHD.

**Meaning:** These sub-components of the HDL pathway, if causally related to atherosclerosis, offer a route to more precise therapeutic targets for prevention of CHD.

## Abstract

**Importance:** Cholesterol content in high-density lipoprotein particles (HDL-C) is associated inversely with coronary heart disease (CHD), but findings from Mendelian randomization studies and randomized trials of HDL-C raising drugs have questioned whether this link is causal. However, these analyses do not exclude a causal role for specific HDL sub-fractions of different density, mobility, size and composition.

**Objective:** To determine whether sub-components of the HDL pathway exhibit differing relationships with CHD risk.

**Design:** In seven longitudinal studies, we used factor analysis to reduce 21 measures of HDL particle size and lipid content to a smaller number of factors representing different components of the HDL pathway. We constructed factor scores and modelled their associations on CHD risk in adjusted Cox regression analyses. We pooled results using random-effects meta-analysis.

**Setting:** Seven population-, individual-, occupational- or community-based longitudinal studies in the UK and Finland.

**Participants:** 28,597 participants (49% female, mean age 59.6 years) contributed to the analysis.

**Exposures:** Sub-components of the HDL pathway, characterized by 21 measures of HDL size and lipid content based on nuclear magnetic resonance spectroscopy.

**Main Outcomes:** Incident fatal or non-fatal CHD.

**Results:** We identified 4 HDL components with highly replicable across studies; 3 were indices of particle size/composition (extra-large (XL), large (L) and medium/small (MS)), and the other an index of triglycerides (TG) carried in HDL of all sizes. After up to 17 years of follow-up, 4179 incident CHD cases occurred. After adjusting for age, sex, ethnicity, smoking, systolic blood pressure, body mass index, diabetes and LDL-C, higher levels of the XL and MS factors were linked to a reduced risk of CHD (hazard ratio per 1 standard deviation (SD) increase 0.88 [95% CI 0.85, 0.92] and 0.91 [0.87,

0.94]). In contrast, a SD increase in the level of the TG factor was associated with increased risk of CHD (1.10 [1.07, 1.14]).

**Conclusions and Relevance:** We found qualitative differences between sub-components of the HDL pathway and the risk of developing CHD. Discovery of the biological determinants of these components, possibly through genetic analysis, will facilitate selection of drug targets and inform trial design.

## Introduction

There is extensive epidemiological evidence that high levels of cholesterol in high density lipoproteins (HDL-C) is associated with reduced risk of cardiovascular events<sup>1</sup>. These observations led to a drug-discovery pipeline which employed HDL-C as target, and this biomarker informed compound selection and trial design<sup>2</sup>. As a result, several drugs that raise HDL-C by the modulation of several protein-targets were identified that were expected to decrease the risk of cardiovascular events. However, randomized trials of drugs to raise HDL-C yielded inconsistent results<sup>3-6</sup>; for example the CETP inhibitor anacetrapib<sup>4</sup> was shown to reduce coronary heart disease (CHD) risk, while niacin<sup>5</sup>, which targets a different protein, had no discernible effect. This suggests that HDL-C is incapable of distinguishing which protein is a valid therapeutic target for CHD prevention.

Some<sup>7</sup>, but not all<sup>8</sup>, Mendelian randomization studies designed to ascertain causality of HDL-C (instead of the protein target of HDL-C raising drugs), suggest that HDL-C is unlikely to be causally related to coronary atherosclerosis. However, evidence from genome-wide association studies indicates that variants at certain HDL-C raising loci do exhibit associations with CHD<sup>9, 10</sup>, whilst others do not, again suggesting that HDL-C is incapable of distinguishing which gene-encoded proteins are valid therapeutic targets for CHD prevention.

Measurement of cholesterol content in HDL particles alone may not capture the full complexity of the HDL pathway, which is composed of a heterogeneous family of lipoprotein particles of different density, mobility, size and composition that are under the control of different enzymatic processes<sup>11</sup>, which may have diverse biological functions and hence a differential role in the origins of cardiovascular events<sup>12</sup>. In support of this is the finding that rare genetic disorders of enzymes that control HDL metabolism seem to affect different sub-classes of HDL particles and this is associated with different phenotypes<sup>13</sup>. Moreover, some recent assays that measure HDL-C efflux capacity have shown this to be associated with CHD, independently of HDL-C<sup>14, 15</sup>.

To quantify the associations of specific components of the HDL pathway in the development of CHD, we used Nuclear Magnetic Resonance (NMR) spectroscopy<sup>16</sup> to measure 21 biomarkers encompassing lipoprotein size and composition in seven longitudinal studies with 28,597 participants and 4197 coronary events.

## Methods

### *Studies contributing data to this analysis*

Data from the British Regional Heart Study (BRHS), British Women's Heart and Health Study (BWHHS), FINRISK, PROspective Study of Pravastatin in the Elderly at Risk (PROSPER), Southall And Brent REvisited (SABRE), United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) and the Whitehall-II study (WHII) were used for this investigation. Study details have previously been described<sup>17-23</sup> and are also summarized in **Online-only material**. **Figure 1** presents the flow of participants selected for these analyses. Study design, outcome definition, and time of blood sampling used for NMR spectroscopy is described in **eTable 1**.

### *HDL particles analysis using NMR spectroscopy*

Serum blood samples were collected and stored after fasting in all studies according to study protocols, except those in FINRISK, which were after absorption was likely to be complete (median 5 hrs; 5.2% <4 hrs after last meal), and UKCTOCS, which were not taken in the fasting state.

Biomarkers were quantified using high-throughput NMR spectroscopy (Nightingale Health Ltd, Helsinki, Finland), which provides simultaneous quantification of routine lipids, lipoprotein subclass profiling with lipid concentrations within 14 subclasses, fatty acid composition, and various low-molecular weight metabolites including amino acids, ketone bodies and glycolysis-related metabolites in molar concentration units. Details of the experimentation and epidemiological applications of the NMR platform are described elsewhere<sup>16, 24</sup>.

We selected 21 HDL biomarkers from the NMR output for this analysis, including lipoprotein components (free cholesterol (FC), cholesterol esters (CE), phospholipids (PL) and triglycerides (TG)) in each size class (extra-large (XL), large (L), medium (M) and small (S)), the concentration of particles circulating in each size class and the mean diameter of all HDL particles (**eTable 2**). We excluded

HDL biomarkers that represent the direct sum of other measured lipids and derived biomarkers such as fractions.

#### *Ascertainment of coronary heart disease*

Incident CHD events were defined as non-fatal myocardial infarction (MI) and fatal cardiac death in all studies. Some studies additionally included coronary artery by-pass (CABG), percutaneous transluminal coronary angioplasty (PTCA), unstable angina and definite angina. Outcome definitions used in each study are in **eTable 1**.

#### *Statistical analysis*

We employed a two-stage analysis to investigate the association of specific components of the HDL pathway with incident CHD. This approach was applied independently in each study using individual patient data; see **Online-only material** for full technical details.

In the first stage we used unsupervised factor analysis to estimate 'latent' (i.e. unobserved but inferred through statistical modelling from observed variables) HDL factors as summary measures that attempt to capture systematic information from the correlation matrix of the 21 biomarkers.

We fitted a saturated 15-factor model and, to determine the optimal number of factors, looked for eigenvalue  $>1$  (**Figure 2**) and a threshold that provided a consistent solution across studies and was supported by plausible biological interpretations of the factors generated. We used maximum likelihood estimation followed by oblique quartimin rotation and defined resultant loadings  $\geq 0.4$  (in absolute value) as salient<sup>25</sup>. We assessed consistency of between-study solutions with a coefficient of variation (CV). From the optimal solution we predicted factor scores for individuals using the regression method<sup>25</sup>.

In the second stage we explored the relationship between HDL factor scores and incident CHD in participants free of CHD at baseline in a complete-case analysis with Cox regression. Scores from

each HDL factor were standardized (mean=0, SD=1). We evaluated factors jointly in a model adjusted for age, then progressively introduced confounders of CHD risk: sex, ethnicity, smoking, systolic blood pressure (SBP), body mass index (BMI), prevalent Type-2 diabetes mellitus, and LDL-C where available. We further adjusted for TG not carried by HDL as a sensitivity analysis; we use non-HDL-TG rather than total-TG to avoid collinearity with the HDL-TG present in both. We did not adjust for HDL-C, as its components (free-cholesterol and cholesterol-ester within each particle size) are already included in factor scores. In each study we excluded covariates with >10% missingness from modelling.

We pooled study-specific regression results in a random-effects meta-analysis using the DerSimonian-Laird estimator of between-study variability, and quantified heterogeneity with the  $I^2$  statistic<sup>26</sup>.

We used R (R Foundation for Statistical Computing, Vienna, Austria; <http://www.r-project.org>) for all statistical analyses; version 3.0.2 for meta-analysis.

## Results

A total of 33,364 subjects (47% women) with a mean age of 59.6 years contributed data to the factor analyses. In the regression analysis of the associations between factor scores and incident CHD, we excluded 1632 (4.3%) due to missing covariate or outcome data and 3135 (9.4%) due to history of CHD at or prior to baseline. Baseline characteristics of the participants and proportion of missingness for data used in both analyses are shown in **Table 1** and **eTables 3 and 4**. After a follow-up for CHD that ranged from a mean of 2.7 years in PROSPER to 17.3 years in SABRE, 4179 incident CHD events were accrued among the 28,597 participants with complete data and no history of CHD at baseline.

### *NMR spectroscopy vs. enzymatic HDL-C measurements*

We found the degree of correlation within studies between levels of HDL-C determined using enzymatic methods versus those determined using NMR spectroscopy to be between 0.76 and 0.90. There was good agreement between each one's association with CHD in all studies with the two HRs varying by <15% (**eTable 5**).

### *HDL factors*

The pattern of correlation among HDL biomarkers was consistent across studies (**eFigure 1**) as was the distribution of HDL-C subfractions across total HDL-C (**eFigure 2**). Screeplots for all seven studies suggested the presence of 3 to 5 HDL factors (**Figure 2**). Content from the 4-factor solution was the most consistent between studies (**Figure 3** and **eFigure 3**). Each study estimated: a factor denoted extra-large (XL) that, when factor loadings were averaged, loaded highly on the 5 XL-HDL biomarkers (particle concentration, phospholipids, triglycerides, and free and esterified cholesterol) and the mean HDL particle diameter; a large (L) factor that loaded highly on the 5 L-HDL biomarkers and the mean HDL particle diameter; a medium/small (MS) factor that loaded highly on the 5 M-HDL

biomarkers and 4 S-HDL biomarkers (excluding triglycerides); and a triglyceride (TG) factor that loaded highly only on triglycerides in all particle sizes.

The pattern of factor loadings was highly consistent between the seven studies: the median CV within each factor was 5.6, 10.1, 10.2 and 13.1% in XL, L, MS and TG respectively (**eTable 6**). The solution was also highly specific, with 17 out of 21 HDL biomarkers being associated with only one factor, even though in an unsupervised factor analysis they could load to all factors. Exceptions were the mean HDL diameter (associated with the XL and L factors) and triglycerides components in XL-, L- and M-HDL particles (associated with their relevant size factor and with the TG factor).

We observed moderate correlations between the XL and L factors (mean  $r=0.58$ ) and between the L and MS factors (mean  $r=0.36$ ). All other correlations among factors were  $<0.2$ ; in particular, the MS factor did not correlate with XL (mean  $r=0.06$ ) or TG (mean  $r=0.02$ ). Correlation patterns were highly consistent across studies, with all CVs  $<12\%$  (**eTable 7**).

#### HDL factors vs. HDL-C

HDL-C was positively correlated with the XL (mean  $r= 0.67$ ), L ( $r=0.79$ ), and MS ( $r=0.63$ ) HDL factors. These were driven by the correlations of HDL-C sub-fractions with their respective factors. HDL-C was negatively correlated with the TG factor ( $r=-0.32$ ) (**eFigure 4**). Despite these moderate to strong correlations, HDL cholesterol sub-fractions make up less than half of the HDL factor scores: 42% of XL; 32% of L; 42% of MS and 25% of TG (**eFigure 5 and eTable 6**).

Where data was available we examined correlations between the HDL factors and HDL<sub>2</sub>-C, HDL<sub>3</sub>-C and apolipoprotein-B (**eTable 8**). HDL<sub>2</sub>-C was strongly correlated with the XL, L and MS factors (mean  $r=0.56$ , 0.76 and 0.71) and HDL<sub>3</sub>-C moderately so ( $r=0.45$ , 0.49 and 0.36). With the TG factor, they were correlated in opposite directions (HDL<sub>2</sub>-C: -0.39, HDL<sub>3</sub>-C: 0.24). These findings indicate a lack of specificity between HDL<sub>2</sub>-C and HDL<sub>3</sub>-C and the HDL factors. Apolipoprotein-B was uncorrelated with the XL and MS factors ( $r<0.01$  and  $r=-0.02$ , respectively), but moderately negatively correlated with the L factor ( $r=-0.23$ ) and positively correlated with the TG factor ( $r=0.57$ ).

### *Association between HDL factors and CHD*

In 5 out 7 studies all potential confounders were available. Due to unavailability or missingness >10% UKCTOCS did not adjust for smoking, SBP and diabetes and WH-II did not adjust for BMI. Details of how we accounted for this in the meta-analysis are available in the **Online-only Material**. UKCTOCS did not record whether participants had a history of CHD prior to the time of serum sampling, so such participants (if any) could not be excluded or adjusted for and were included in the analysis.

**Figure 4** describes the summary results of the progressively adjusted Cox models on the four HDL factors. After adjusting for age and the other three factors an SD increase in the XL factor was associated with a 14% reduction in the risk of CHD (HR 0.86, 95% CI [0.81,0.91]), in the MS factor with an 8% reduction (0.92 [0.88,0.96]) and in the TG factor with a 15% increase (1.15 [1.11,1.19]). The progressive adjustment of established confounders for CHD (sex, ethnicity, smoking, SBP, BMI, diabetes, LDL-C) had minimal effect on these estimates and reduced between-study heterogeneity. After all adjustments there was evidence of 12% (0.88 [0.85, 0.92]) and 9% (0.91 [0.87, 0.94]) reduction in CHD risk per SD increase in the XL and MS factors and a 10% (1.10 [1.07, 1.14]) increase in risk per SD increase in the TG factor, with little heterogeneity between studies (all  $I^2 \leq 12\%$ ). There was no evidence of an association between the L factor and CHD (1.00 [0.92, 1.08];  $I^2=66\%$ ). **eFigure 6** shows study-specific HRs for the minimally- and maximally-adjusted models.

Associations of the XL, L and MS factors with CHD did not change after further adjustment by non-HDL triglycerides. However, the TG factor lost a substantial amount of precision, likely due to collinearity between this factor and TG in other lipoproteins, which were strongly correlated in most studies (range: 0.39, 0.88). Our results were unaltered by excluding data from the statin arm of PROSPER.

## Discussion

In this analysis including up to 33,364 individuals and 4,197 coronary events, we provide evidence indicating heterogeneous associations between components of the HDL-pathway, proxied through four HDL factors, with the risk of CHD. The HDL components represented by HDL biomarkers covering size and composition on extra-large and medium/small particles were inversely associated with CHD, while the triglyceride component was directly associated after adjustment for potential confounders including LDL-C. No association was observed for the component represented by biomarkers covering size and composition of large particles.

The potential benefits of the HDL-components we identified, compared with HDL-C, can be illustrated through an analogy on total-cholesterol versus LDL-C. Initial technology suitable for population-based studies was capable of capturing only total-cholesterol, and this measurement, in studies like the Framingham Heart Study<sup>27</sup>, showed a harmful association with CHD. Improvements in technology made the measurement of major cholesterol sub-fractions possible and revealed opposite associations: LDL-C was harmful but HDL-C was protective. In this study, further improvements in technology through NMR spectroscopy allowed us to uncover heterogeneous associations within the HDL pathway.

The four HDL components described in this analysis were highly reproducible across seven studies, despite important differences in study design and participant characteristics. This high reproducibility suggests these factors are very likely to have distinctive biological determinants. We anticipate that discovery of their biological determinants using genomics, proteomics and functional assays could help refine selection of intervention targets. Given the qualitatively different associations with CHD, genomic studies aimed for target selection should consider the potential associations that each gene-encoded protein may have on four HDL components modelled jointly.

Until now, target-selection, compound development, early pre-clinical and randomized trials have used HDL-C as the sole metric to drive the drug-discovery process. With this approach, if two distinct

drug targets (and associated compounds) had a similar effect on HDL-C concentration, they would be expected to have similar therapeutic efficacy against CHD. However, evidence from large-scale phase-3 trials, other observational studies showing that HDL-C protection disappears with adjustment for other HDL particles<sup>28, 29</sup>, and the results we present in this paper do not support this hypothesis.

Additional support for the HDL factors we identified comes from the good agreement with the five HDL subclasses (extra-large, large, medium, small and very small) proposed by Rosenson et al<sup>13</sup>.

These categories were proposed as a new nomenclature to capture the diverse data around HDL physical properties derived from various HDL measurement techniques, but not including the specific NMR platform used in this report. Although previous studies<sup>28-30</sup> have attempted to uncover the impact of HDL sub-classes on CHD using NMR, our study provide novel lines of evidence. First, previous studies only measured HDL particle size and not composition within HDL sub-classes.

Therefore, their results do not truly capture the impact of HDL as a pathway. An example of this is that role of triglycerides carried only within HDL particles was not assessed in previous studies.

Second, our study covering seven cohorts includes ten times more CHD cases than any of previous studies, refuting any concern on inconsistency of findings. Taking into consideration those differences, our findings also mirror those from the VA-HIT trial, which found that medium and small HDL particles were protective against CHD while the large particles were not.

Our analysis has several strengths. First, we observed reproducible estimates for the HDL components in independent, unsupervised factor analyses across seven diverse study populations.

Second, despite using an unsupervised strategy that allows each biomarker to associate with all factors, the final composition of the factors had high specificity (17 of 21 loading on a single factor) and strong biological plausibility. For example, the biomarkers that formed the XL, L and MS factors were largely represented by particle concentrations, phospholipids, free-cholesterol and cholesterol esters from their respective HDL-subclasses. Third, we observed minimal between-study heterogeneity in regression estimates for three of the four factors. Fourth, several publications<sup>31-33</sup>

have shown the validity of the NMR platform used in this analysis. For example, genetic determinants of major lipid fractions measured through NMR are consistent with those discovered by enzymatic methods<sup>31</sup>. We also showed strong correlation between NMR HDL-C and enzymatic HDL-C and concordant CHD protective associations across all studies.

We chose not to use total HDL-C for the construction of the factors. Instead we used their components and hence did not adjust estimates for total HDL-C, as the variables that comprise it were already included in the factors. Adjustment would be desirable if the aim was to propose to use these new HDL factors as better measures of “risk prediction” for CHD. Instead our goal was to identify components of the HDL pathway relevant for drug-discovery. By analogy, total cholesterol, instead of LDL-C, is the biomarker used in “risk prediction” models like Framingham<sup>34</sup> or the Pooled Cohort Equations<sup>35</sup>, despite being of less value for drug-discovery.

Several findings indicate that the HDL factors are distinct from cholesterol carried across HDL particles of different size. First, free-cholesterol and cholesterol ester within HDL particles accounted for less than half of the value of HDL factor scores. Second, HDL<sub>2</sub>-C and HDL<sub>3</sub>-C were not strongly correlated with the HDL-factors. Third, despite HDL-C having moderate-to-strong correlations with the HDL-factors, the factors that showed associations with CHD were largely uncorrelated between themselves.

Even though we had access to a large sample and standardized analysis plan, our results need to be replicated in other large-scale surveys such as the UK-Biobank or the Mexico City Prospective study where the same NMR platform-based output should be available in the near future. This increased sample size will allow robust detection of small associations such as those we report in the current study and avoid use of heterogeneous datasets in terms of study design, outcome ascertainment and missingness and avoid the need for assumptions related to meta-analytic techniques. Additional advantages of studies like UK-Biobank<sup>36</sup> and INTERVAL<sup>37</sup>, with genomics and serum NMR

measurements, are the possibility to explore if the associations of HDL factors with incident CHD are likely to be causal.

In conclusion, our HDL factor analysis uncovered the presence of heterogeneous associations between HDL-pathway factors and CHD and suggests that the full biological effect of these pathways is unlikely to be captured by HDL-C alone. If our findings are validated in independent cohorts in future analyses using genetics and proven to be causal, this will help to precisely inform the selection of drug targets and to guide the drug discovery process from pre-clinical phase to phase-3 trials.

## Acknowledgements

The authors acknowledge and thank the participants and study teams from BRHS, BWHHS, FINRISK, PROSPER, SABRE, UKCTOCS and WH-II for data collection and support and Drs Reecha Sofat and Rui Providencia for providing valuable comments on the manuscript.

## Funding Declarations

The BRHS is supported by British Heart Foundation grant RG/13/16/30528. The BWHHS is supported by British Heart Foundation grant PG/13/66/30442. FINRISK: VS has been supported by the Finnish Foundation for Cardiovascular Research. SABRE was funded at baseline by the UK Medical Research Council and Diabetes UK. Follow-up studies have been funded by the Wellcome Trust (WT 082464), British Heart Foundation (SP/07/001/23603 and CS/13/1/30327) and Diabetes UK (who funded the metabolomics analyses: 13/0004774). NC received support from the National Institute for Health Research University College London Hospitals Biomedical Research Centre. Support has also been provided at follow-up by the North and West London and Central and East London National Institute of Health Research Clinical Research Networks. UKCTOCS was funded by Medical Research Council (G9901012 and G0801228), Cancer Research UK (C1479/A2884), and Department of Health and The Eve Appeal. UM is supported by funding from the NIHR University College London Hospitals (UCLH) Biomedical Research Centre. MK is supported by the Medical Research Council (R024227) for WH-II infrastructure and data collection; MS is supported by the British Heart Foundation for data analysis. The UCLEB Consortium is supported by the British Heart Foundation (RG/10/12/28456 and SP/13/6/30554) and by the Rosetrees and Stoneygate Trust. ADH is an NIHR Senior Investigator and funded by the UCL Hospitals NIHR Biomedical Research Centre.

### **Role of the funders**

The funders and study sponsors have played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; nor in the decision to submit the article for publication. The researchers were independent from the funders and sponsors.

### **Conflicts of interest**

VS has participated in a conference trip supported by Novo Nordisk and received a modest honorarium from the same source for participating in an advisory board meeting. JW reports employment with and ownership of shares in GSK. JPC received funding from GSK to conduct methodological work around use of electronic health records and multi-omics for drug-discovery. CJP reports honoraria or grants from MSD, Sanofi/Regeneron, Amgen, and Daiichi-Sankyo. UM has stocks in Abcodia Pvt Ltd, awarded to her by UCL. All other authors have no conflicts to declare.

## References

1. Vergeer M, Holleboom AG, Kastelein JJ, Kuivenhoven JA. The HDL hypothesis: does high-density lipoprotein protect from atherosclerosis? *Journal of lipid research*. 2010 Aug;51(8):2058-73
2. Kingwell BA, Chapman MJ, Kontush A, Miller NE. HDL-targeted therapies: progress, failures and future. *Nature Reviews Drug Discovery*. 2014 05/23/online;13:445
3. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, Komajda M, et al. Effects of torcetrapib in patients at high risk for coronary events. *The New England journal of medicine*. 2007 Nov 22;357(21):2109-22
4. Bowman L, Hopewell JC, Chen F, Wallendszus K, Stevens W, Collins R, et al. Effects of Anacetrapib in Patients with Atherosclerotic Vascular Disease. *The New England journal of medicine*. 2017 Sep 28;377(13):1217-27
5. Landray MJ, Haynes R, Hopewell JC, Parish S, Aung T, Tomson J, et al. Effects of extended-release niacin with laropiprant in high-risk patients. *The New England journal of medicine*. 2014 Jul 17;371(3):203-12
6. Lincoff AM, Nicholls SJ, Riesmeyer JS, Barter PJ, Brewer HB, Fox KAA, et al. Evacetrapib and Cardiovascular Outcomes in High-Risk Vascular Disease. *The New England journal of medicine*. 2017 May 18;376(20):1933-42
7. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. *European heart journal*. 2014 Jan 27
8. Burgess S, Freitag DF, Khan H, Gorman DN, Thompson SG. Using multivariable Mendelian randomization to disentangle the causal effects of lipid fractions. *PloS one*. 2014;9(10):e108891
9. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature genetics*. 2017 Sep;49(9):1385-91
10. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nature genetics*. 2017 Dec;49(12):1758-66
11. Li Z, McNamara JR, Ordovas JM, Schaefer EJ. Analysis of high density lipoproteins by a modified gradient gel electrophoresis method. *Journal of lipid research*. 1994 Sep;35(9):1698-711
12. Russell Warnick G, Nauck M, Rifai N. Evolution of Methods for Measurement of HDL-Cholesterol: From Ultracentrifugation to Homogeneous Assays. *Clinical chemistry*. 2001;47(9):1579-96
13. Rosenson RS, Brewer HB, Jr., Chapman MJ, Fazio S, Hussain MM, Kontush A, et al. HDL measures, particle heterogeneity, proposed nomenclature, and relation to atherosclerotic cardiovascular events. *Clinical chemistry*. 2011 Mar;57(3):392-410
14. Rohatgi A, Khera A, Berry JD, Givens EG, Ayers CR, Wedin KE, et al. HDL cholesterol efflux capacity and incident cardiovascular events. *The New England journal of medicine*. 2014 Dec 18;371(25):2383-93
15. Saleheen D, Scott R, Javad S, Zhao W, Rodrigues A, Picataggi A, et al. Association of HDL cholesterol efflux capacity with incident coronary heart disease events: a prospective case-control study. *The lancet Diabetes & endocrinology*. 2015 Jul;3(7):507-13
16. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circulation Cardiovascular genetics*. 2015 Feb;8(1):192-206
17. Lennon LT, Ramsay SE, Papacosta O, Shaper AG, Wannamethee SG, Whincup PH. Cohort Profile Update: The British Regional Heart Study 1978-2014: 35 years follow-up of cardiovascular disease and ageing. *International journal of epidemiology*. 2015 Jun;44(3):826-g
18. Lawlor DA, Bedford C, Taylor M, Ebrahim S. Geographical variation in cardiovascular disease, risk factors, and their control in older women: British Women's Heart and Health Study. *J Epidemiol Community Health*. 2003;57:134-40

19. Borodulin K, Vartiainen E, Peltonen M, Jousilahti P, Juolevi A, Laatikainen T, et al. Forty-year trends in cardiovascular risk factors in Finland. *European journal of public health*. 2015 Jun;25(3):539-46
20. Shepherd J, Gerard Jan Blauw, Michael B Murphy, Cobbe SM, Bollen ELEM, Brendan M Buckley, et al. The Design of a Prospective Study of Pravastatin in the Elderly at Risk (PROSPER). *Am J Cardiol*. 1999;84:1192-7
21. Tillin T, Forouhi NG, McKeigue PM, Chaturvedi N, Group SS. Southall And Brent REvisited: Cohort profile of SABRE, a UK population-based comparison of cardiovascular disease and diabetes in people of European, Indian Asian and African Caribbean origins. *International journal of epidemiology*. 2012 Feb;41(1):33-42
22. Menon U, Gentry-Maharaj A, Ryan A, Sharma A, Burnell M, Hallett R, et al. Recruitment to multicentre trials--lessons from UKCTOCS: descriptive study. *Bmj*. 2008 Nov 13;337:a2079
23. Marmot M, Brunner E. Cohort Profile: the Whitehall II study. *International journal of epidemiology*. 2005 Apr;34(2):251-6
24. Wurtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *American journal of epidemiology*. 2017 Nov 1;186(9):1084-96
25. Gorsuch RL. Exploratory factor analysis. In: Nesselroade JR, Cattell RB, editors. *Handbook of multivariate experimental psychology*. 2 ed. New York, NY, US: Plenum Press; 1988. p. 231-58.
26. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-60
27. Kannel WB, Castelli WP, Gordon T. Cholesterol in the prediction of atherosclerotic disease. New perspectives based on the Framingham study. *Annals of internal medicine*. 1979 Jan;90(1):85-91
28. Mackey RH, Greenland P, Goff DC, Jr., Lloyd-Jones D, Sibley CT, Mora S. High-density lipoprotein cholesterol and particle concentrations, carotid atherosclerosis, and coronary events: MESA (multi-ethnic study of atherosclerosis). *J Am Coll Cardiol*. 2012 Aug 7;60(6):508-16
29. Mora S, Glynn RJ, Ridker PM. High-density lipoprotein cholesterol, size, particle number, and residual vascular risk after potent statin therapy. *Circulation*. 2013 Sep 10;128(11):1189-97
30. Ottos JD, Collins D, Freedman DS, Shalaurova I, Schaefer EJ, McNamara JR, et al. Low-density lipoprotein and high-density lipoprotein particle subclasses predict coronary events and are favorably changed by gemfibrozil therapy in the Veterans Affairs High-Density Lipoprotein Intervention Trial. *Circulation*. 2006 Mar 28;113(12):1556-63
31. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikainen LP, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature genetics*. 2012 Jan 29;44(3):269-76
32. Wurtz P, Havulinna AS, Soininen P, Tynkkynen T, Prieto-Merino D, Tillin T, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation*. 2015 Mar 3;131(9):774-85
33. Holmes MV, Millwood IY, Kartsonaki C, Hill MR, Bennett DA, Boxall R, et al. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *J Am Coll Cardiol*. 2018 Feb 13;71(6):620-32
34. D'Agostino RB, Sr., Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008 Feb 12;117(6):743-53
35. Goff DC, Jr., Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Sr., Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014 Jul 1;63(25 Pt B):2935-59

36. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*. 2015 Mar;12(3):e1001779
37. Moore C, Sambrook J, Walker M, Tolkien Z, Kaptoge S, Allen D, et al. The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial. *Trials*. 2014 Sep 17;15:363

## Figure legends

**Figure 1:** Participant flow for each study

**Figure 2:** Screeplot of eigenvalues (solid lines, left y-axis) and cumulative proportion of variance explained (dashed lines, right y-axis) for each factor from the saturated 15-factor factor analysis solutions in seven studies. The grey and pink horizontal lines denote an eigenvalue of 1 and cumulative variance explained of 80% respectively. The vertical line indicates the optimal number of factors with most eigenvalues greater than 1 and >80% of variance explained.

**Figure 3:** HDL biomarker factor loading patterns for 4 latent factors for each of the seven studies.

Biomarkers with factor loadings outside the range of +/-0.4 factor loading (as indicated by the dotted lines) are considered 'high' loading.

**Figure 4:** Associations between HDL factors and the risk of CHD in 28 597 participants with no history of CHD at the time of serum sampling. The factors are jointly modelled and minimally adjusted for age, then cumulatively by sex, ethnicity, smoking, systolic blood pressure, body mass index, diabetes, LDL-cholesterol concentration and non-HDL triglyceride concentration. Estimates represent the effect for a 1-SD increase in the normalised factor scores.

## Tables

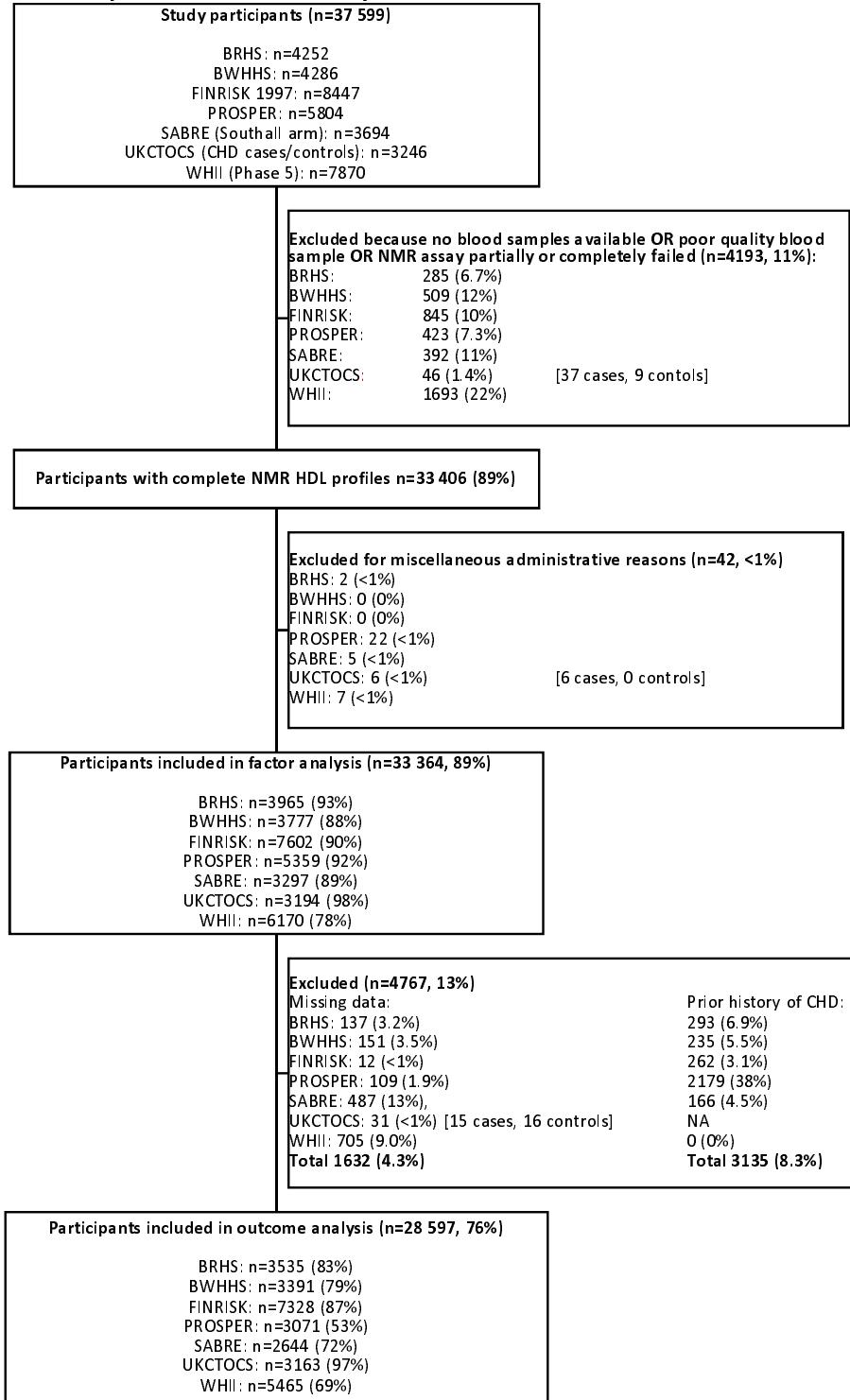
**Table 1:** Baseline subject demographics for CHD regression analyses (n=28,597). All numbers are mean (SD) or n (%).

	BRHS (n=3535)	BWHHS (n=3391)	FINRISK (n=7328)	PROSPER (n=3071)	SABRE (n=2644)	UKCTOCS (n=3163)	WHII (n=5465)
<b>Age (years)</b>	68.6 (5.49)	68.64 (5.48)	47.77 (13)	75.6 (3.32)	51.92 (7.08)	64.73 (6.2)	55.56 (6.01)
<b>Gender (female)</b>	0 (0%)	3391 (100%)	3754 (51%)	1770 (58%)	388 (15%)	3163 (100%)	1561 (29%)
<b>Ethnicity (white)</b>	3512 (99%)	3378 (>99%)	7328 (100%)	3071 (100%)	1246 (47%)	3068 (97%)	5101 (93%)
<b>Smoking (current)</b>	455 (13%)	369 (11%)	1740 (24%)	934 (30%)	588 (22%)	NA	541 (9.9%)
<b>Systolic BP (mmHg)</b>	149.7 (23.7)	153.2 (25.3)	135.7 (19.9)	155.9 (22)	124.3 (17.5)	NA	123 (16.4)
<b>BMI (kg/m<sup>2</sup>)</b>	26.84 (3.63)	27.4 (4.82)	26.61 (4.52)	29.28 (146)	26.12 (3.77)	27.02 (5.37)	26.03 (3.88)
<25	1100 (31%)	1295 (38%)	2897 (40%)	1139 (37%)	1082 (41%)	1273 (40%)	2049 (37%)
25-30	1869 (53%)	1338 (39%)	3012 (41%)	1323 (43%)	1220 (46%)	1164 (37%)	2061 (38%)
>30	565 (16%)	757 (22%)	1418 (19%)	608 (20%)	341 (13%)	740 (23%)	640 (12%)
<b>Diabetes(yes)</b>	205 (5.8%)	302 (8.9%)	393 (5.4%)	245 (8%)	303 (11%)	NA	264 (4.8%)
<b>Total cholesterol (mmol/l)*</b>	6.026 (1.07)	6.655 (1.18)	5.547 (1.06)	5.033 (1.08)	6.004 (1.1)	NA	5.942 (1.04)
<b>NMR Total cholesterol (mmol/l)</b>	4.532 (0.92)	6.011 (1.33)	5.329 (1.11)	4.218 (0.951)	3.958 (0.905)	4.94 (1.05)	5.071 (0.909)
<b>HDL cholesterol (mmol/l)*</b>	1.329 (0.341)	1.665 (0.455)	1.401 (0.362)	1.397 (0.388)	1.29 (0.365)	NA	1.464 (0.393)
<b>NMR HDL cholesterol (mmol/l)</b>	1.197 (0.306)	1.676 (0.451)	1.581 (0.389)	1.359 (0.279)	1.041 (0.234)	1.624 (0.411)	1.514 (0.329)
<b>LDL cholesterol (mmol/l)*</b>	3.893 (0.963)	4.165 (1.07)	3.485 (0.933)	3.125 (0.974)	3.92 (0.964)	NA	3.863 (0.933)
<b>NMR LDL cholesterol (mmol/l)</b>	1.815 (0.515)	2.436 (0.75)	1.919 (0.602)	1.388 (0.494)	1.42 (0.453)	1.803 (0.584)	1.895 (0.486)
<b>Triglycerides (mmol/l)*</b>	1.827 (1.07)	1.833 (0.952)	1.483 (1.03)	1.38 (0.654)	1.888 (1.39)	NA	1.355 (0.864)
<b>NMR Triglycerides (mmol/l)</b>	1.561 (0.744)	1.664 (0.84)	1.306 (0.683)	1.288 (0.56)	1.006 (0.401)	1.721 (0.895)	1.241 (0.468)
<b>Non-HDL triglycerides (mmol/l)*</b>	1.404 (0.709)	1.49 (0.798)	1.132 (0.63)	1.149 (0.527)	0.8975 (0.377)	1.529 (0.842)	1.102 (0.441)
<b>On lipid-lowering medication**</b>	192 (5.4%)	201 (5.9%)	185 (2.5%)	1503 (49%)	4 (<1%)	NA	119 (2.2%)
<b>Prior history of CHD</b>	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	NA	0 (0%)

<b>Had CHD event</b>	331 (9.4%)	248 (7.3%)	802 (11%)	217 (7.1%)	821 (31%)	1565 (49%)	195 (3.6%)
<b>Follow-up time (years)</b>	8.57 (2.64)	10.49 (2.7)	13.87 (2.79)	2.71 (0.58)	17.26 (5.48)	5.20 (2.27)	6.23 (1.24)

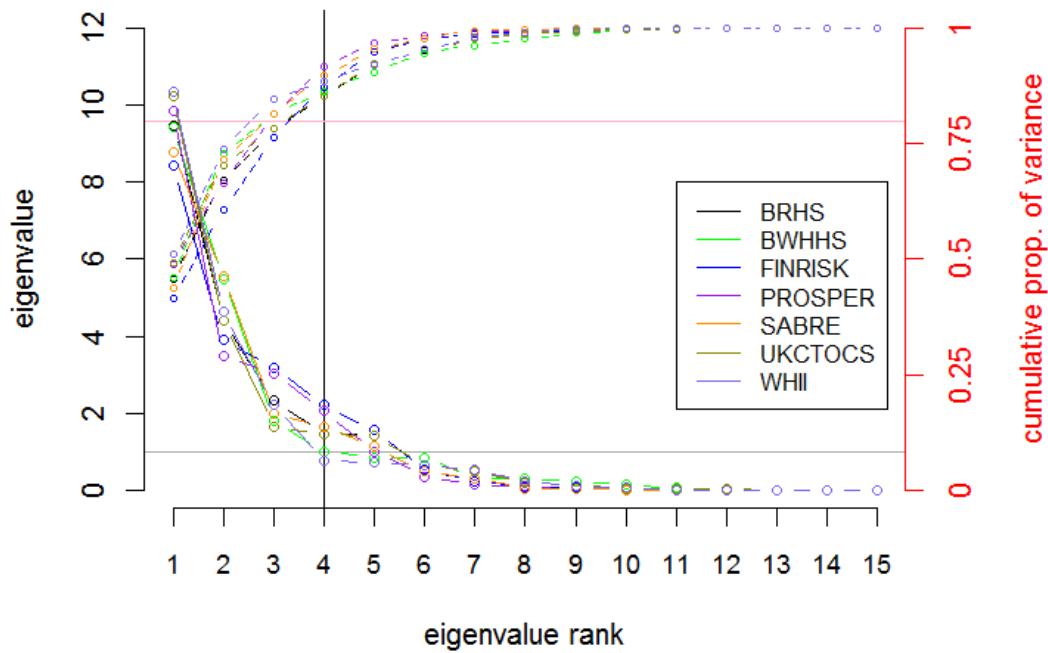
\*fasting measures; \*\*number in pravastatin arm of the PROSPER trial; NA, not available.

**Figure 1: Participant flow for each study**

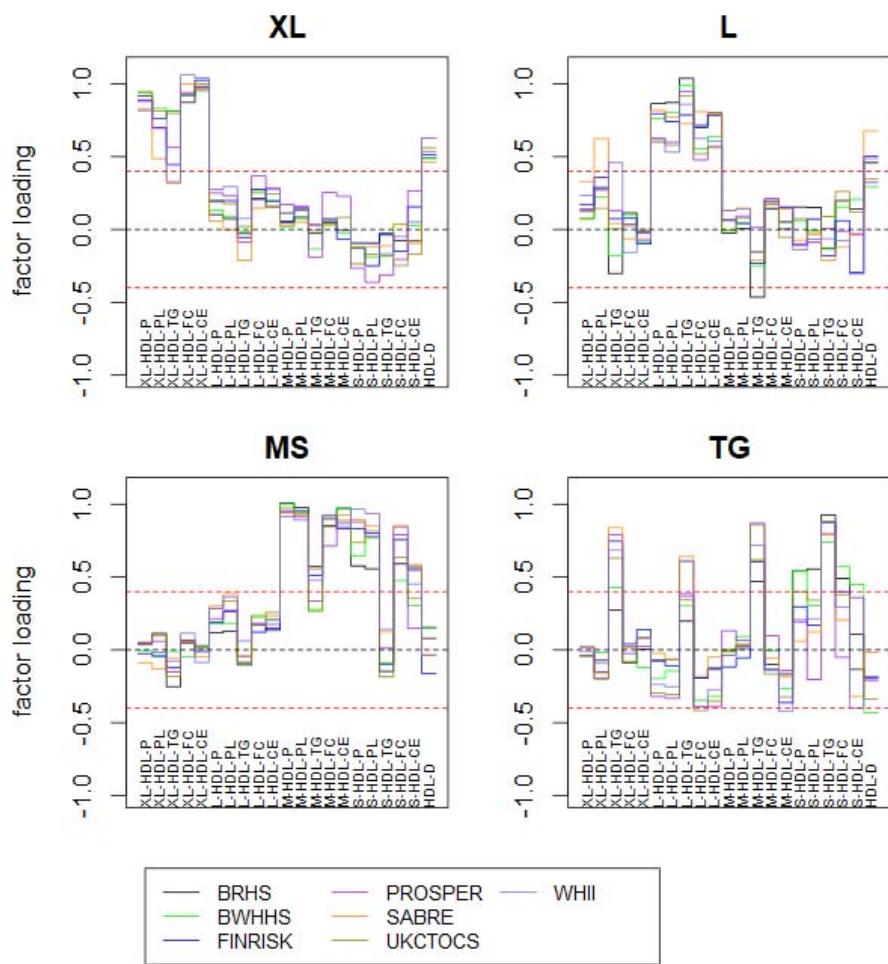


NA: not available.

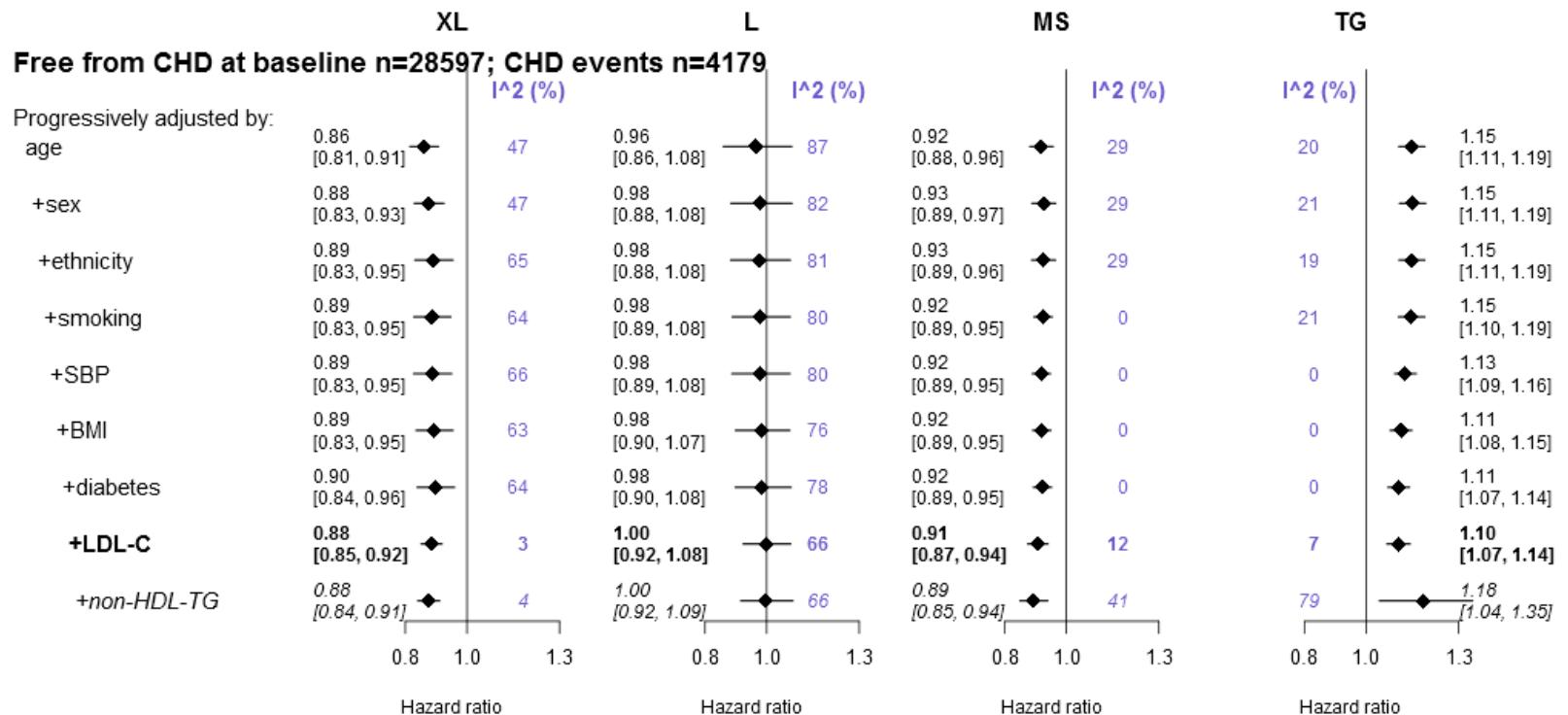
**Figure 2: Screeplot of eigenvalues (solid lines, left y-axis) and cumulative proportion of variance explained (dashed lines, right y-axis) for each factor from the saturated 15-factor factor analysis solutions in seven studies. The grey and pink horizontal lines denote an eigenvalue of 1 and cumulative variance explained of 80% respectively. The vertical line indicates the optimal number of factors with most eigenvalues greater than 1 and >80% of variance explained.**



**Figure 3: HDL biomarker factor loading patterns for 4 latent factors for each of the seven studies. Biomarkers with factor loadings outside the range of +/-0.4 factor loading (as indicated by the dotted lines) are considered 'high' loading.**



**Figure 4: Associations between HDL factors and the risk of CHD in 28 597 participants with no history of CHD at the time of serum sampling. The factors are jointly modelled and minimally adjusted for age, then cumulatively by sex, ethnicity, smoking, systolic blood pressure, body mass index, diabetes, LDL-cholesterol concentration and non-HDL triglyceride concentration. Estimates represent the effect for a 1-SD increase in the normalised factor scores.**



**Table 1: Baseline subject demographics for CHD regression analyses (n=28,597). All numbers are mean (SD) or n (%).**

	BRHS (n=3535)	BWHHS (n=3391)	FINRISK (n=7328)	PROSPER (n=3071)	SABRE (n=2644)	UKCTOCS (n=3163)	WHII (n=5465)
<b>Age (years)</b>	68.6 (5.49)	68.64 (5.48)	47.77 (13)	75.6 (3.32)	51.92 (7.08)	64.73 (6.2)	55.56 (6.01)
<b>Gender (female)</b>	0 (0%)	3391 (100%)	3754 (51%)	1770 (58%)	388 (15%)	3163 (100%)	1561 (29%)
<b>Ethnicity (white)</b>	3512 (99%)	3378 (>99%)	7328 (100%)	3071 (100%)	1246 (47%)	3068 (97%)	5101 (93%)
<b>Smoking (current)</b>	455 (13%)	369 (11%)	1740 (24%)	934 (30%)	588 (22%)	NA	541 (9.9%)
<b>Systolic BP (mmHg)</b>	149.7 (23.7)	153.2 (25.3)	135.7 (19.9)	155.9 (22)	124.3 (17.5)	NA	123 (16.4)
<b>BMI (kg/m<sup>2</sup>)</b>	26.84 (3.63)	27.4 (4.82)	26.61 (4.52)	29.28 (146)	26.12 (3.77)	27.02 (5.37)	26.03 (3.88)
<25	1100 (31%)	1295 (38%)	2897 (40%)	1139 (37%)	1082 (41%)	1273 (40%)	2049 (37%)
25-30	1869 (53%)	1338 (39%)	3012 (41%)	1323 (43%)	1220 (46%)	1164 (37%)	2061 (38%)
>30	565 (16%)	757 (22%)	1418 (19%)	608 (20%)	341 (13%)	740 (23%)	640 (12%)
<b>Diabetes(yes)</b>	205 (5.8%)	302 (8.9%)	393 (5.4%)	245 (8%)	303 (11%)	NA	264 (4.8%)
<b>Total cholesterol (mmol/l)*</b>	6.026 (1.07)	6.655 (1.18)	5.547 (1.06)	5.033 (1.08)	6.004 (1.1)	NA	5.942 (1.04)
<b>NMR Total cholesterol (mmol/l)</b>	4.532 (0.92)	6.011 (1.33)	5.329 (1.11)	4.218 (0.951)	3.958 (0.905)	4.94 (1.05)	5.071 (0.909)
<b>HDL cholesterol (mmol/l)*</b>	1.329 (0.341)	1.665 (0.455)	1.401 (0.362)	1.397 (0.388)	1.29 (0.365)	NA	1.464 (0.393)
<b>NMR HDL cholesterol (mmol/l)</b>	1.197 (0.306)	1.676 (0.451)	1.581 (0.389)	1.359 (0.279)	1.041 (0.234)	1.624 (0.411)	1.514 (0.329)
<b>LDL cholesterol (mmol/l)*</b>	3.893 (0.963)	4.165 (1.07)	3.485 (0.933)	3.125 (0.974)	3.92 (0.964)	NA	3.863 (0.933)
<b>NMR LDL cholesterol (mmol/l)</b>	1.815 (0.515)	2.436 (0.75)	1.919 (0.602)	1.388 (0.494)	1.42 (0.453)	1.803 (0.584)	1.895 (0.486)
<b>Triglycerides (mmol/l)*</b>	1.827 (1.07)	1.833 (0.952)	1.483 (1.03)	1.38 (0.654)	1.888 (1.39)	NA	1.355 (0.864)
<b>NMR Triglycerides (mmol/l)</b>	1.561 (0.744)	1.664 (0.84)	1.306 (0.683)	1.288 (0.56)	1.006 (0.401)	1.721 (0.895)	1.241 (0.468)
<b>Non-HDL triglycerides (mmol/l)*</b>	1.404 (0.709)	1.49 (0.798)	1.132 (0.63)	1.149 (0.527)	0.8975 (0.377)	1.529 (0.842)	1.102 (0.441)
<b>On lipid-lowering medication**</b>	192 (5.4%)	201 (5.9%)	185 (2.5%)	1503 (49%)	4 (<1%)	NA	119 (2.2%)
<b>Prior history of CHD</b>	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	NA	0 (0%)
<b>Had CHD event</b>	331 (9.4%)	248 (7.3%)	802 (11%)	217 (7.1%)	821 (31%)	1565 (49%)	195 (3.6%)
<b>Follow-up time (years)</b>	8.57 (2.64)	10.49 (2.7)	13.87 (2.79)	2.71 (0.58)	17.26 (5.48)	5.20 (2.27)	6.23 (1.24)

\*fasting measures; \*\*number in pravastatin arm of the PROSPER trial; NA, not available.