Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/25760 holds various files of this Leiden University dissertation

Author: Vaarhorst, Anika
Title: Genetic and metabolomic approaches for coronary heart disease risk prediction
Issue Date: 2014-05-13
A metabolomics profile is associated with the risk of incident coronary heart disease

Anika A.M. Vaarhorst1**, Aswin Verhoeven2**, Claudia M. Weller3**, Stefan Böhringer4, Sibel Göraler5, Axel Meissner6, André M. Deelder7, Peter Henneman3, Anton P.M. Gorgels8, Piet A. van den Brandt6,7, Leo J. Schouten7, Marleen M. van Greevenbroek8, Audrey H.H. Merry6, W.M. Monique Verschuren9, Arn M.J.M. van den Maagdenberg3,10, Ko Willems van Dijk3,11, Aaron Isaacs12, Dorret Boomsma13, Ben A. Oostra12, Cornelia M. van Duijn12, J. Wouter Jukema14,15,16, Jolanda M.A. Boer9, Edith Feskens17, Bastiaan T. Heijmans1, P. Eline Slagboom1,18

*These authors contributed equally to this work

From the departments of 1) Molecular Epidemiology, 2) Center for Proteomics and Metabolomics, 3) Human Genetics, 4) Medical Statistics, 10) Neurology, 11) Endocrinology, 14) Cardiology, Leiden University Medical Center, Leiden, the Netherlands From the departments of 5) Cardiology, 8) Internal Medicine (CARIM School for Cardiovascular diseases), Maastricht University Medical Centre, Maastricht, the Netherlands From the departments of 6) Epidemiology (CAPHRI School for Public Health and Primary Care), 7) Epidemiology (GROW School of Oncology and Developmental Biology), Maastricht University, Maastricht, the Netherlands; 9) National Institute for Public Health and the Environment, Bilthoven, the Netherlands; 12) Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 13) Biological Psychology, VU University, Amsterdam, the Netherlands; 15) The Durrer Center for Cardiogenetic Research, Amsterdam, the Netherlands; 16) Interuniversity Cardiology Institute of the Netherlands (ICIN), Utrecht, the Netherlands; 17) Division of Human Nutrition, Wageningen University and Research Center, Wageningen, the Netherlands; 18) Netherlands Consortium for Healthy Ageing

Accepted for publication in American Heart Journal
Abstract

Background Metabolomics, defined as the comprehensive identification and quantification of low-molecular-weight metabolites to be found in a biological sample, has been put forward as a potential tool for classifying individuals according to their risk of coronary heart disease (CHD). Here, we investigated whether a single-point blood measurement of the metabolome is associated with and predictive for the risk of CHD.

Methods & Results We obtained proton nuclear magnetic resonance (1H-NMR) spectra in 79 cases who developed CHD during follow-up (median 8.1 years) and in 565 randomly selected individuals. In these spectra 100 signals representing 36 metabolites were identified. Applying LASSO regression, we defined a weighted metabolite score consisting of 13 1H-NMR signals that optimally predicted CHD. This metabolite score, including signals representing a lipid fraction, glucose, valine, ornithine, glutamate, creatinine, glycoproteins, citrate and 1.5-anhydrosorbitol, was associated with the incidence of CHD independent of traditional risk factors (TRFs) (HR=1.50; 95%CI=1.12-2.01). Predictive performance of this metabolite score on its own was moderate (C-index=0.75; 95%CI=0.70-0.80) but after adding age and sex the C-index was only modestly lower than that of TRFs (C-index=0.81; 95%CI=0.77-0.85 and C-index=0.82; 95%CI=0.78-0.87, respectively). The metabolite score was also associated with prevalent CHD independent of TRFs (OR=1.59; 95%CI=1.19-2.13).

Conclusion A metabolite score derived from a single-point metabolome measurement is associated with CHD and metabolomics may be a promising tool for refining and improving the prediction of CHD.
A metabolomics profile is associated with the risk of incident coronary heart disease

Introduction
Over the last 50 years, risk factors that are robustly and independently associated with coronary heart disease (CHD), including lipid levels, blood pressure, lifestyle factors, family history, sex and age, were identified.\textsuperscript{1,2} Based on these traditional risk factors (TRFs), scores have been developed to predict CHD risk for an individual.\textsuperscript{1,2} The discriminatory capabilities for these scores, as assessed by the C-index, ranges from 0.71 to 0.84.\textsuperscript{1}

Metabolomics refers to the identification and quantification of low-molecular-weight metabolites in a biological sample.\textsuperscript{3} Recent technological developments made it possible to generate metabolomics profiles of blood samples consisting of 10s to 100s metabolites in a single measurement.\textsuperscript{4} These profiles are considered to be promising tools to efficiently capture the predictive information of TRFs and may potentially contribute to further improvement of primary CHD risk prediction.\textsuperscript{3}

Several studies have attempted to use a metabolomics approach to diagnose prevalent CHD\textsuperscript{5,6} or to predict incident CHD events in individuals free from cardiovascular disease\textsuperscript{7-9} or diagnosed with CHD\textsuperscript{10-13} or diabetes.\textsuperscript{14} According to some studies, a metabolomics approach might improve CHD risk prediction.\textsuperscript{8-13} For example, in one study, \textsuperscript{1}H-NMR spectroscopy improves the prediction of subclinical atherosclerosis in comparison to conventional lipid testing.\textsuperscript{8} Other studies found that a baseline metabolomics profile based on mass-spectroscopy was associated with incident cardiovascular events in patients diagnosed with CHD\textsuperscript{11} or in patients with suspected CHD.\textsuperscript{10,12,13} However, none of these studies investigated if a metabolite profile based on low-molecular-weight molecules identified by proton nuclear magnetic resonance (\textsuperscript{1}H-NMR) spectroscopy could predict incident CHD, defined as an acute myocardial infarction (MI), unstable angina pectoris (UAP) or dead because of CHD, in individuals free from cardiovascular disease. Therefore, we studied the association of an \textsuperscript{1}H-NMR based metabolite profile with incident CHD in a prospective case-cohort study. Subsequently, a second study was performed in an independent population to test if the selected metabolites were also relevant to classify prevalent CHD.

Material and Methods

Study populations

Primary study: We conducted a prospective case-cohort study within the Monitoring project on chronic disease risk factors (MORGEN-Project) 1993-1997,\textsuperscript{15} one of the two monitoring studies that were included in the Cardiovascular Registry Maastricht study.\textsuperscript{16} In total 6459 men and women, between 20-59 years old at the moment of inclusion, had given informed consent to retrieve information from the municipal registries and from the
Chapter 2

general practitioner and specialist. The study complied with the Declaration of Helsinki and was approved by the Medical Ethics Committee of TNO (Dutch Organization for Applied Scientific Research). We excluded participants who were younger than 30 years at baseline (n=1301), who had had an acute MI, UAP, a coronary artery bypass graft or a percutaneous transluminal coronary angioplasty before baseline (n=69), or were lost to follow-up (n=15), resulting in an eligible cohort of 5074 participants.

Sub-cohort selection: From the eligible cohort, a sub-cohort of 738 participants was randomly drawn. This took place before cardiologic follow-up. EDTA plasma was unavailable for 92 participants and 1H-NMR analysis failed in 19 participants. For 62 participants information on TRFs (i.e. total cholesterol [TC], high-density lipoprotein cholesterol [HDL-C], systolic blood pressure [SBP], current smoking, body mass index [BMI], current diabetes status and a parental history of MI) was incomplete, resulting in a sub-cohort of 565 participants.

Cardiologic follow-up: The cardiologic follow-up has been described in detail earlier and ended on 31 December 2003 with a median follow-up of 8.1 years (range=0.2-10.9 years). During follow-up 125 participants developed CHD (acute MI n=55, UAP n=51, dead due to CHD n=19, [ICD-9 410-414 and ICD-10 I20-I25]). For 31 patients EDTA plasma for 1H-NMR analysis was unavailable and in one patient 1H-NMR analysis failed. For 14 patients information on TRFs was incomplete, resulting in 79 patients.

Determination of TRFs: At baseline, participants filled in a questionnaire about medical history (including self-reported diabetes), parental history of MI (defined as no parents with MI, one parent with MI or both parents with MI), and lifestyle factors (including smoking). During a medical examination, information on SBP and BMI was collected and non-fasting EDTA blood samples were taken. The blood was centrifuged for 10 minutes at 1000 rpm at 4°C and EDTA plasma aliquots were stored at -80°C in tubes of 0.5 mL for future analysis, or at -20°C for cholesterol determinations. HDL-C and TC levels were determined in the plasma samples stored at -20°C using a CHOD-PAP method.

Secondary study: The Erasmus Rucphen Family study is a population-based study in a genetically isolated community in the Southwest of the Netherlands and includes 3465 individuals, who are living descendants of 22 couples that had at least six children baptized in the community church between 1850 and 1900. Details are described elsewhere. The study was approved by the Medical Ethical Committee of the Erasmus Medical Center in Rotterdam and all participants gave written informed consent.

Determination of TRFs: All participants filled in questionnaires about lifestyle, personal and family medical history. During personal interviews performed by study physicians, information on lifestyle factors, medication use, personal and family medical history was collected. Physical examinations were performed, including measurements of SBP and BMI.
In addition, fasted blood samples were taken. An electrocardiogram (ECG) was performed and scored by an experienced cardiologist. Plasma concentrations of HDL-C and TC were determined according to standard procedures. Diabetes was defined as the use of blood glucose-lowering medication and/or fasting glucose levels of ≥7 mmol/L.

**Diagnosis of CHD in the secondary study**: Participants were classified as CHD cases if they indicated during the interview or in the questionnaire that they had experienced a MI or underwent a coronary revascularization procedure, they reported angina symptoms in the interview, and/or showed signs of MI on ECG.

From 2919 participants fasting serum samples were available. Good quality 1H-NMR spectra from 2415 participants were obtained. For 2327 of these participants data on CHD diagnosis was available and 170 were classified as having CHD.

**1H-NMR metabolite profiling**

The stored EDTA plasma and serum samples were thawed at 4°C and were mixed by inverting the tubes 10 times. Next, samples (300 µL) were mixed with 300 µL TSP buffer (see Supplement I, supplementary Material & methods) and transferred into 5mm NMR tubes and kept at 6°C while queued for acquisition. Two-dimensional J-resolved and Carr-Purcell-Meiboom-Gill spectra were acquired on a 600 MHz Bruker Avance II spectrometer (Bruker BioSpin, Karlsruhe, Germany), operating at a sample temperature of 310 K. For details on acquisition, processing, quality control, scaling and calibration of the 1H-NMR spectra see Supplement I, supplementary material & methods.

Using the procedure described in the Supplement I, supplementary Material & methods, 100 signals were detected and quantified in the 1H-NMR spectra of every individual in the primary study. For 76 signals metabolites were assigned (see Supplement I, table S1 for all 1H-NMR signals and their assigned metabolites). These signals represented 36 different compounds (i.e. after subtracting the signals representing free EDTA). Signals representing calcium-EDTA and magnesium-EDTA complexes may give an indication of the levels of calcium and magnesium ions, respectively. Using the same procedure, 68 out of 100 signals detected in the primary study were detected in the secondary study. For 54 out of these 68 signals metabolites could be assigned, representing 28 different compounds.

**Statistical analysis**

In the primary study, Cox regression, adjusted for delayed entry, and according to the method of Prentice to adjust for the case-cohort design was performed to see whether
baseline characteristics were associated with incident CHD.\textsuperscript{23} Age in years was used as the time-scale variable.

Prior to analysis, the $^1$H-NMR signals were transformed to Z-scores. We selected a subset of the most informative signals for CHD prediction, using least absolute shrinkage and selection operator (LASSO) regression,\textsuperscript{24} and performed 10-fold cross-validation to determine the tuning parameter.\textsuperscript{24,25} This set was further reduced to signals that could be detected in both studies. The linear predictor of the Cox-model was used as a weighted metabolite score (sum of regression coefficients multiplied by the corresponding covariate values).

Cox regression, according to the method of Prentice to adjust for the case-cohort design, was used to calculate whether this metabolite score was associated with incident CHD before and after adjusting for TRFs.\textsuperscript{23}

To investigate whether the metabolite score improved risk discrimination, Harrell’s concordance index (C-index),\textsuperscript{26} the net reclassification index (NRI) and the integrative discrimination index (IDI) were calculated.\textsuperscript{27} For the NRI the following risk categories were applied: 0–<5%, 5–<10%, 10–<20%, ≥20%. Next, using ANOVA, we investigated to what extent TRFs can explain the variance in the metabolite score in the sub-cohort. Stata/SE version 11.2 was used to calculate the C-index. R-package PredictABEL version 1.2-1 was used to calculate the IDI. For all other analyses, R version 2.14.1 was used.

In the secondary study, raw $^1$H-NMR signal data were adjusted for kinship by linear regression in GenABEL.\textsuperscript{28} The residuals for all signals were transformed into Z-scores. Logistic regression was performed to assess the association of TRFs with prevalent CHD. To assess whether the metabolite signals selected by LASSO regression in the primary study were also relevant for the identification of prevalent CHD cases, we put these signals into one logistic regression model to determine their individual regression coefficients. Next, we used logistic regression analysis to test the association of this metabolite score, thus with weights based on the secondary study, with prevalent CHD. Finally, we tested to what extent TRFs can explain the variance in the metabolite score using ANOVA. For these analyses, PASW statistics version 18 (SPSS-IBM, New York, US) was used.

**Results**

*Primary study:* The TRFs were associated with incident CHD. When all TRFs were entered into one Cox proportional hazards model with age in years as the time-scale variable, HDL-C, SBP, sex and current smoking remained independently associated with incident CHD (table 1). See supplement I, table S2 for the baseline characteristics before excluding individuals
with missing data. For the association of the individual $^1$H-NMR signals with incident CHD see supplement I, table S1.

**Table 1** | TRFs and their association with incident CHD in the primary study.

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Sub-cohort</th>
<th>HR (95%CI)$^*$</th>
<th>HR (95%CI)$^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>51.9±6.1</td>
<td>44.8±8.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total cholesterol (mmol/L)</td>
<td>5.9±1.0</td>
<td>5.3±1.0</td>
<td>1.30 (1.03-1.64)</td>
<td>1.21 (0.93-1.58)</td>
</tr>
<tr>
<td>HDL cholesterol (mmol/L)</td>
<td>1.1±0.3</td>
<td>1.3±0.4</td>
<td>0.12 (0.05-0.30)</td>
<td>0.34 (0.12-0.94)</td>
</tr>
<tr>
<td>SBP (mmHg)</td>
<td>134.5±16.8</td>
<td>121.5±15.1</td>
<td>1.04 (1.02-1.05)</td>
<td>1.03 (1.01-1.05)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>27.6±4.6</td>
<td>25.4±3.8</td>
<td>1.10 (1.03-1.17)</td>
<td>1.03 (0.95-1.11)</td>
</tr>
<tr>
<td>Men</td>
<td>79.6% (63)</td>
<td>44.6% (252)</td>
<td>4.95 (2.72-9.02)</td>
<td>3.30 (1.65-6.57)</td>
</tr>
<tr>
<td>Current smoking</td>
<td>51.9% (41)</td>
<td>38.8% (219)</td>
<td>2.16 (1.29-3.62)</td>
<td>1.84 (1.02-3.32)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>3.8% (3)</td>
<td>0.5% (3)</td>
<td>3.99 (0.70-22.84)</td>
<td>3.34 (0.41-27.43)</td>
</tr>
<tr>
<td>Parental history of MI</td>
<td>50.6% (40)</td>
<td>40.2% (227)</td>
<td>1.32 (0.88-1.98)</td>
<td>1.24 (0.78-1.98)</td>
</tr>
</tbody>
</table>

Data are expressed as mean±SD or % (n). HR, hazard ratio; HDL, high-density lipoprotein; MI, myocardial infarction; SBP, systolic blood pressure; BMI, body mass index

$^*$Including 10 cases

$^+$Univariate HR were calculated per unit increase for age, total cholesterol, HDL cholesterol, SBP, BMI, and for the categorical traits. Age in years was used as the time-scale variable.

We determined 16 $^1$H-NMR signals as the best prediction subset using LASSO regression, of which, 13 were available in the secondary study (see figure 1). These 16 $^1$H-NMR signals represent creatinine, serine, glucose, 1,5-anhydroisorbitol, trimethylamine N-oxide (TMAO), ornithine, citrate, glutamate, glycoproteins, an unsaturated lipid structure, valine and five non-annotated signals located at 3.924, 3.145, 2.412, 1.391 and 0.988 ppm. From the 13 signals present in both the primary and secondary study a weighted metabolite score was constructed using the corresponding coefficients (figure 1). This metabolite score was normally distributed in cases and sub-cohort members (supplement I, figure S1), and associated with incident CHD (hazard ratio per standard deviation [HR/SD]=1.91, 95% confidence interval [95%CI]=1.50-2.44). After adjusting for TRFs, this metabolite score remained associated with incident CHD (HR/SD=1.50, 95%CI=1.12-2.01). For the results of the metabolite score based on the 16 signals see supplement I, table S3.

The metabolite score had a C-index of 0.75, (95%CI=0.70-0.80). Adding age and sex to the metabolite score, resulted in a C-index of 0.81 (95%CI=0.77-0.85), which is similar to a C-index when only TRFs are included (C-index=0.82, 95%CI=0.78-0.87, $p=0.327$). When the metabolite score was added to a model containing all TRFs, the C-index increased.
from 0.82 to 0.84, which was non-significant ($p=0.107$). Both the improvement in the NRI (NRI$_{total}$=0.038; $p=0.209$) and the IDI (0.012; $p=0.091$) were non-significant (supplement I, table S4 and S5). Inspecting C-indices for individual TRFs and evaluating improvement of adding the metabolite score indicated that the metabolite score improved the C-indices of all individual TRFs (supplement I, table S6a/b).

**Figure 1** | The subset of signals selected using 10-fold cross-validated LASSO regression and their coefficients in the primary study (left panel) and the secondary study (right panel).

We tested to what extent TRFs explain the variance in the metabolite score in the subcohort of the primary study. HDL-C, sex, BMI, TC, SBP, age and diabetes explained respective 16.4%, 11.7%, 10.0%, 7.9%, 5.6%, 4.1% and 2.3% of the variance in the metabolite score. Current smoking and parental history of MI all explained less than 1% of the variance in the metabolite score (figure 2). When all TRFs were combined, 32.6% of the variance in the metabolite score was explained.

**Secondary study:** To test if the 13 metabolite signals selected in the primary study were relevant for the identification of prevalent CHD cases, we investigated 170 CHD cases and 2157 controls for which equivalent metabolomics profiles were obtained. In this non-prospective cohort, combining all TRFs in one logistic regression model resulted in only age, sex, and parental history of MI to be independently associated with prevalent CHD (table 2). See supplement I, table S7 for the baseline characteristics before excluding individuals with missing data.
A metabolomics profile is associated with the risk of incident coronary heart disease

Table 2 | TRFs and their association with prevalent CHD in the secondary study.

<table>
<thead>
<tr>
<th></th>
<th>Cases (n=170)</th>
<th>Controls (n=2157)</th>
<th>OR (95%CI)*</th>
<th>OR (95%CI)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>60.9±11.7</td>
<td>47.7±14.0</td>
<td>1.08 (1.07-1.10)</td>
<td>1.08 (1.05-1.10)</td>
</tr>
<tr>
<td>Total cholesterol (mmol/L)</td>
<td>5.0±1.1</td>
<td>5.6±1.1</td>
<td>0.57 (0.49-0.67)</td>
<td>0.47 (0.37-0.59)</td>
</tr>
<tr>
<td>HDL cholesterol (mmol/L)</td>
<td>1.2±0.3</td>
<td>1.3±0.4</td>
<td>0.30 (0.18-0.49)</td>
<td>1.25 (0.60-2.61)</td>
</tr>
<tr>
<td>SBP (mmHg)</td>
<td>148.2±22.9</td>
<td>139.1±19.7</td>
<td>1.02 (1.02-1.03)</td>
<td>1.00 (0.99-1.01)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>28.1±4.4</td>
<td>26.8±4.6</td>
<td>1.06 (1.03-1.09)</td>
<td>1.01 (0.96-1.07)</td>
</tr>
<tr>
<td>Men</td>
<td>61.2% (104)</td>
<td>42.5% (917)</td>
<td>2.40 (1.79-3.22)</td>
<td>1.93 (1.18-3.14)</td>
</tr>
<tr>
<td>Current smoking</td>
<td>36.5% (62)</td>
<td>39.1% (844)</td>
<td>0.90 (0.66-1.24)</td>
<td>1.18 (0.75-1.88)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>12.9% (22)</td>
<td>4.3% (93)</td>
<td>3.44 (2.15-5.49)</td>
<td>1.25 (0.56-2.79)</td>
</tr>
<tr>
<td>Parental history of MI</td>
<td>31.2% (53)</td>
<td>19.9% (430)</td>
<td>2.03 (1.53-2.71)</td>
<td>1.60 (1.16-2.21)</td>
</tr>
</tbody>
</table>

Data are expressed as mean±SD or % (n). OR, odds ratio; HDL, high-density lipoprotein; MI, myocardial infarction; SBP, systolic blood pressure

*Univariate OR were calculated per unit increase for age, total cholesterol, HDL cholesterol, SBP, BMI, and for the categorical traits.

†All the variables were added into one multivariable logistic regression model.

The metabolite score, based on the 13 best predicting signals in the case-cohort study, but with weights based on the secondary study (figure 2), was associated with prevalent CHD before (OR=2.72, p<0.001) and after adjusting for TRFs (odds ratio [OR]=1.59, p=0.002). After excluding statin users (n=299), similar results were obtained (supplement I, table S8).

The proportion of variance in the metabolite score explained by age, SBP and diabetes was higher for the secondary study than for primary study, whereas the variables TC, HDL-C and BMI explained a lower proportion of the variance (figure 2). The proportion of variance explained by sex, current smoking and parental history of MI was comparable for both studies. With all TRFs combined, 32.1% of the variance of the metabolite score could be explained, which is comparable to that of the primary study.
**Discussion**

A metabolite score, based on $^1$H-NMR spectroscopy, is significantly associated with incident CHD independent of TRFs. When combined with age and sex, this score was as predictive for incident CHD as all TRFs combined. A score based on the same $^1$H-NMR signals was also associated with prevalent CHD, independent of TRFs.
The observation that the metabolite score could not improve CHD risk prediction beyond TRFs in individuals free from CHD is in line with a previous study published by El Harchaoui et al. However, Würtz et al. found that metabolites measured by \(^1\)H-NMR spectroscopy improved risk stratification for subclinical atherosclerosis in comparison to conventional lipids. In this study both \(^1\)H-NMR determined lipoproteins and low-molecular weight metabolites were included. We only included information on low-molecular weight metabolites, whereas in the study by El Harchaoui et al. only information on lipoproteins was included. Perhaps the combination of both lipoproteins and low-molecular-weight metabolites results in the optimal prediction of CHD.

The metabolite score represents the metabolites valine, ornithine, glucose, 1,5-anhydro-sorbitol, creatinine, an unsaturated lipid structure, glutamate and, glycoproteins, citrate, and TMAO, of which TMAO was not available in the secondary study. Most of these metabolites have been associated with CHD or CHD risk factors before. Valine has been associated with metabolic risk factors, insulin resistance, incident type 2 diabetes, and future cardiovascular events. Ornithine, is produced by splitting of urea from arginine, resulting in a lower bioavailability of arginine. Arginine is necessary to produce nitric oxide, which is essential for a normal endothelial function. This pathway has been linked to CHD and CHD mortality. The presence of glucose and 1,5-anhydro-sorbitol, a short-term marker for glycemic control, could indicate that our metabolite score marks individuals at higher risk of developing diabetes or insulin resistance and thereby CHD. Low creatinine levels are a marker for high HDL-C and low LDL cholesterol levels. Thus the presence of creatinine and an unsaturated lipid structure could indicate that our metabolite risk score is a marker for an unfavorable lipid profile. This is confirmed by the proportion of variance explained by HDL-C and TC levels, 7.9% and 16.4%. In the secondary study these explained variances are only 0.9% and 4.6%, but this discrepancy might be caused by statin treatment, resulting in lower cholesterol levels for the cases compared to controls. A secondary explanation for the incorporation of creatinine in the metabolite score is that elevated creatinine levels may indicate kidney dysfunction, which is associated with cardiovascular disease. Increased TMAO levels have been associated with cardiovascular risk before. Thus it seems that the LASSO procedure selected relevant metabolites that have been associated with CHD and CHD risk factors before.

Several issues have to be resolved before it can be concluded if a metabolomics approach is useful for CHD risk prediction. First, the known, quantifiable serum metabolome consists of 4229 metabolites, of which only 36 (0.9%) were included in this study. Other studies that use \(^1\)H-NMR spectroscopy also incorporated lipoproteins. Therefore we hope that we can achieve better in follow-up studies when incorporating H-NMR determined lipoproteins.
in addition to low-molecular-weight metabolites in our analysis. Moreover, other metabolomics platforms should be also be included.37 Second, the 16 signals provided by our study should be measured in large prospective cohorts for replication and to determine universally applicable weights. The current study is too small for that purpose. Third, we had non-fasting samples in the primary study and fasted samples in the secondary study. However, we still found that the $^1$H-NMR signals selected in the primary non-fasted study were also associated with prevalent CHD in the fasted secondary study. This indicates that we have selected $^1$H-NMR signals that are robust whether fasted or non-fasted samples are used. Fourth, constructing robust prediction models constitutes a statistical challenge, especially in a high-dimensional setting. In our case, model selection by LASSO regression resulted in predictor selection that eliminated high correlations among predictors. This can lead to reduced transferability of prediction models as correlation structures of predictors can vary between studies. A wide variety of penalized regression models are available (e.g. elastic net, ridge regression) and further research is needed to select the appropriate methods for the application at hand.

**Conclusion**

A metabolite score derived from a single point metabolome measurement is associated with the risk of CHD independent of TRFs, but thus not improve risk prediction beyond TRFs On the other hand, LASSO regression resulted in the selection of relevant metabolites, suggesting that more comprehensive metabolomics methods may be promising tools to further improve upon CHD disease understanding and risk stratification.

**Acknowledgements**

We thank the MORGEN-project steering committee consisting of (not mentioning co-authors of this article) HB Bueno de Mesquita, HA Smit and JC Seidell. Furthermore we thank the epidemiologists and field-workers of the Municipal Health Services in Maastricht for their contribution to baseline data collection and those involved in logistic management (A Jansen and J Steenbrink) and data management (A Blokstra, A van Kessel, P Steinberger, E den Hoedt, I Toxopeus, J van der Laan). We thank A Hilton, V Visser, P Erkens, S Philippens, J Bremen and B Bleijlevens for assistance in clinical data collection. Statistics Netherlands is acknowledged for providing data on causes of death. We thank Joris Deelen, Dennis Kremer, Ruud van der Breggen, Yingchang Lu, Mariken Tijhuis and Cecile Povel for assistance with practical work.
A metabolomics profile is associated with the risk of incident coronary heart disease

We thank the participants from the Genetic Research in Isolated Populations, Erasmus Rucphen Family studies, as well as the general practitioner and other clinicians who made this work possible.

**Funding sources**

This study was funded by a grant from the Netherlands Heart Foundation (2006B195), the Netherlands Consortium for Healthy Ageing (05060810) and by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007).

The MORGEN-Project was financially supported by the Ministry of Public Health, Welfare and Sports of the Netherlands.

The ERF study was funded by the Centre for Medical Systems Biology in the framework of the Netherlands Genomics Initiatives and by the European Network of Genomic And Genetic Epidemiology consortium.

The funding sources had no involvement in the design and conduct of the study; or with the collection, management, analysis or interpretation of the data and reporting or the decision to submit the manuscript for publication.
References


A metabolomics profile is associated with the risk of incident coronary heart disease


