The handle [http://hdl.handle.net/1887/25760](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
General Introduction
What is coronary heart disease?

Coronary heart disease (CHD) arises from atherosclerosis in the coronary arteries. Atherosclerosis is a slowly progressing disease resulting in the formation of atherosclerotic plaques and narrowing of the arteries. Rupture of an atherosclerotic plaque leads to thrombus formation and temporal or permanent obstruction of the coronary artery. As a consequence, the underlying tissue is no longer supplied with oxygen and nutrients, resulting in ischemia and possible necrosis of myocardial tissue. The patient experiences this as chest pain that may radiate to the neck, back, or arms, which is usually accompanied with nausea, excessive sweating, shortness of breath and fear of death. Although 20% of the patients (mostly women, diabetics, postoperative patients and elderly people) with a temporal or permanent obstruction of one of the coronary arteries lack chest pain and experience atypical symptoms.

Acute CHD can appear as an acute myocardial infarction (MI) or unstable angina pectoris (UAP). If a patient experiences chest pain and/or other complaints concerning an obstruction of the coronary arteries an electrocardiogram (ECG) should be performed. As shown in figure 1, an abnormal ECG is indicative for CHD, but is not sufficient by itself to distinguish between acute MI and UAP. The most important difference between acute MI and UAP is the presence of necrosis of myocardial tissue. Therefore, as shown in figure 1, the final diagnosis of acute MI or UAP depends on the detection of proteins that are released in the blood due to myocardial tissue necrosis.

![Figure 1](image_url_or_description)
**Mortality of coronary heart disease**

CHD is the main cause of death worldwide. In the Netherlands, 10,400 individuals died because of CHD (6000 males and 4400 females), accounting for 7.6% of all deaths in 2010, making CHD the second most important cause of death in the Netherlands (see table 1).

<table>
<thead>
<tr>
<th>Disease/illness</th>
<th>Males</th>
<th>Females</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung cancer</td>
<td>6.586</td>
<td>3.958</td>
<td>10.544</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>5.724</td>
<td>4.152</td>
<td>9.876</td>
</tr>
<tr>
<td>Dementia (including Alzheimer)</td>
<td>2.544</td>
<td>6.617</td>
<td>9.161</td>
</tr>
<tr>
<td>Stroke</td>
<td>3.315</td>
<td>5.165</td>
<td>8.480</td>
</tr>
<tr>
<td>Heart failure</td>
<td>2.499</td>
<td>3.999</td>
<td>6.498</td>
</tr>
<tr>
<td>COPD</td>
<td>3.466</td>
<td>2.887</td>
<td>6.353</td>
</tr>
<tr>
<td>Infections of the lower respiratory tract</td>
<td>2.613</td>
<td>3.068</td>
<td>5.681</td>
</tr>
<tr>
<td>Colon cancer</td>
<td>2.619</td>
<td>2.508</td>
<td>5.127</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>22</td>
<td>3.261</td>
<td>3.283</td>
</tr>
<tr>
<td>Private accidents</td>
<td>1.246</td>
<td>1.575</td>
<td>2.821</td>
</tr>
<tr>
<td>All causes of death combined</td>
<td><strong>65.259</strong></td>
<td><strong>70.482</strong></td>
<td><strong>135.741</strong></td>
</tr>
</tbody>
</table>

*COPD, Chronic Obstructive Pulmonary Disease*

Since 1970 CHD mortality is declining in western countries, including the Netherlands. This can be explained by technological advances such as bypass surgery and percutaneous coronary interventions, which prevent death during an acute manifestation of CHD. Second, there are interventions that prolong survival once CHD is manifest (secondary prevention). This includes stimulating a healthy lifestyle and medical treatment to reduce high cholesterol levels and high blood pressure in individuals diagnosed with CHD. And finally by primary prevention aimed at preventing CHD in individuals that never had a CHD event. This can be done by stimulating a healthy lifestyle in the general population but also by treating individuals with a high risk for CHD with the same interventions that are used for secondary prevention. For the latter approach to be successful it is important to select only the individuals at high risk for CHD. To do so risk scores have been developed to select these high risk individuals.

**Risk scores for coronary heart disease**

In 1948 the Framingham Heart Study was initiated, the goal of this study was to identify risk factors for CHD. At that point it was already assumed that CHD was a multifactor disease.
that developed over a long time. The investigators included 5209 individuals, free from CHD aged between 30 and 59 years living in Framingham, Massachusetts and hoped that they could follow the included individuals for as long as 20 years. It was expected that by comparing individuals that developed CHD with individuals that remained healthy during the course of time, risk factors for CHD could be found.

In 1957, the first report on the Framingham study was published. It was found that older age, male sex, high blood pressure, obesity and high cholesterol levels were associated with atherosclerotic heart disease. Twenty years later a general cardiovascular risk profile was proposed. This risk profile included the joint effects of gender, age, systolic blood pressure, serum cholesterol levels, cigarette smoking, electrocardiographic evidence of left ventricular hypertrophy and glucose intolerance. The Framingham score currently used is simplified and based on a larger and more recent population (i.e. the original Framingham cohort plus their offspring). This simplified version includes the traditional risk factors (TRFs) sex, age, total cholesterol or LDL cholesterol, systolic blood pressure, smoking and self-reported diabetes. Left ventricular hypertrophy was no longer included, since it was highly correlated with hypertension. To date several other risk scores have been developed to predict the risk for cardiovascular disease and CHD in individuals free from cardiovascular disease. Of these, the SCORE function is used to assess CHD risk in individuals free from cardiovascular disease in the Netherlands.

The performance of these risk scores can be estimated by the C statistic, which is also known as the area under the receiver operator characteristic (ROC) curve or the C-index in case of survival data. The C-index estimates the probability that of two randomly chosen individuals, the one with a higher score will develop CHD earlier than the one with a lower score. The C-index can vary between 0.5 (50% of the cases are correctly ranked, meaning that the test is not better than flipping a coin) and 1 (all the cases are correctly ranked). For the Framingham risk score and comparable scores the C-index is around 0.75, meaning that in 75% of the cases, two randomly chosen individuals are correctly ranked. This indicates that there is room for improvement of the currently used CHD risks scores, which can be done by adding novel risk markers.

**Metabolomics and coronary heart disease**

Since atherosclerosis is a disease of the arteries it is reasonable to assume that most CHD risk factors can be found in blood. However of the risk factors to be found in blood the current risk scores only use total cholesterol, HDL-cholesterol and triglyceride levels as risk markers. There are a many other compounds to be found in blood, for example free fatty acids, glucose, amino acids, lipid particles that can all play a role in the development of
atherosclerosis and thus CHD. By applying metabolomics technology it is possible to acquire information about the blood metabolome by identifying and quantifying metabolites in a non-biased way.\textsuperscript{34,35} To date more than 4000 different metabolites are already identified and quantified in human blood, and this number is still increasing.\textsuperscript{36} These metabolites reveal information about the metabolic state of an individual and this information might be useful for understanding, diagnosing or predicting CHD.

There are different techniques available for identifying and quantifying these metabolites.\textsuperscript{36} Proton nuclear magnetic resonance (\textsuperscript{1}H-NMR) spectroscopy and chromatography coupled to mass spectrometry (MS) based platforms are mostly used.\textsuperscript{37,38} See box 1 for a description of both techniques and their advantages and disadvantages.\textsuperscript{37,38} \textsuperscript{1}H-NMR spectroscopy is a

Box 1 | Techniques used for measuring metabolites at a large scale

The two most commonly used methods to obtain information about a person’s metabolome in one single measurement are chromatography coupled to mass spectrometry (MS) and proton nuclear magnetic resonance (\textsuperscript{1}H-NMR) spectroscopy.\textsuperscript{37,38}

With chromatography coupled to MS, the compounds within the samples are first separated by separation techniques based on liquid or gas chromatography.\textsuperscript{41} Next, to identify these compounds the samples are placed in the mass spectrometer and ionized, the resulting charged molecules (ions) are then separated by their mass to charge ratio.\textsuperscript{41} Next, these ions fly into the detector, which records the current induced when this happens.\textsuperscript{41} This information is visualized in a mass spectrum.\textsuperscript{41} Depending on the location and the heights of the peaks in the mass spectrum, information about the identity and the abundance of the molecules in the biological sample can be deducted.\textsuperscript{41}

\textsuperscript{1}H-NMR spectroscopy is used to find clues about the structure of unknown molecules.\textsuperscript{37,42} \textsuperscript{1}H stands for hydrogen-1, which is the most common hydrogen isotope.\textsuperscript{42} This hydrogen isotope contains 1 proton and 1 electron and absorbs electromagnetic radiation at a specific frequency or resonance, depending on the force of the magnetic field used.\textsuperscript{42} In a \textsuperscript{1}H-NMR spectrometer, the electromagnetic radiation is fixed at a specific frequency but the magnetic field can be varied from high to low.\textsuperscript{42} When hydrogen atoms absorb electromagnetic radiation, this is visualized by peaks in the \textsuperscript{1}H-NMR spectrum.\textsuperscript{42} The location of the peaks depends on the chemical environment (i.e. the other atoms located nearby the hydrogen atom) of the hydrogen atoms.\textsuperscript{42} As a result a high or a low magnetic field is needed for the hydrogen atoms to absorb electromagnetic radiation.\textsuperscript{42} The heights and areas of the peaks reveal information about the number of hydrogen atoms in these specific chemical environments.\textsuperscript{42} Based on this information it is possible to identify molecules in a biological sample.\textsuperscript{42}

| Table | The advantages and disadvantages of \textsuperscript{1}H-NMR and chromatography coupled to MS based techniques.\textsuperscript{43} |
| Method | Advantages/Disadvantages |
| \textsuperscript{1}H-NMR spectroscopy | + highly reproducible <br> + fast <br> - relatively few metabolites can be identified and quantified <br> - low sensitivity, only high abundant metabolites can be detected |
| Chromatography coupled to MS | + many different metabolites can be identified and quantified <br> + high sensitivity, also low abundant metabolites can be detected <br> - slow <br> - sample preparation is difficult |
robust and reliable technique that is highly reproducible,\textsuperscript{39} which is a prerequisite for the identification of novel biomarkers that may be used for CHD risk prediction improvement.\textsuperscript{40} Moreover, \textsuperscript{1}H-NMR spectroscopy is fast, especially when compared to chromatography coupled to MS based platforms. The downside of \textsuperscript{1}H-NMR spectroscopy is it low sensitivity compared to MS based platforms. For example with \textsuperscript{1}H-NMR spectroscopy it is possible to identify and quantify up to 44 different metabolites in blood, whereas with chromatography-MS based methods up to hundreds of different metabolites can be identified and quantified.\textsuperscript{36} Thus with \textsuperscript{1}H-NMR spectroscopy it is possible to generate a robust and reliable profile of the metabolome in large study populations in relatively little time and it is hoped that this profile can be used to improve CHD risk prediction beyond TRFs.

**Genetics and coronary heart disease**

Twin, family and adoption studies showed that CHD has a heritable component.\textsuperscript{44-47} In a Swedish twin-registry study it was found that in males 57% of the variance in death from CHD can be explained by heritable factors and for females this was 38%.\textsuperscript{46} The proportion in CHD death explained by a non-shared environment was 43% for males and 62% for females, the shared environment did not explain any of the variance found.\textsuperscript{46} In a Danish twin study comparable results were found.\textsuperscript{47} Thus CHD is caused by a combination of genetic and environmental risk factors. Not only CHD, but also CHD risk factors have a large heritable component, including smoking,\textsuperscript{48} blood pressure,\textsuperscript{49} and blood lipid levels.\textsuperscript{50}

In 1996, Risch and Merikangas proposed that finding genetic variants associated with common complex disease in the common population required large scale testing by association analysis.\textsuperscript{51} With the completion of a physical map of the human genome in 2001,\textsuperscript{52} and the identification of more than 1.4 million single nucleotide polymorphisms (SNPs)\textsuperscript{53} it became possible to associate diseases and/or traits with common genetic variants on a genome-wide scale. These studies are commonly referred to as genome-wide association studies (GWAS).

In 2009 a catalog with published GWAS came online.\textsuperscript{54} On 13/12/12 this catalog (http://www.genome.gov/gwastudies/) contained 1459 publications and 8084 SNPs. Of these studies, 153 publications are related to CHD or CHD related risk factors (i.e. blood pressure, blood lipid levels, diabetes type and anthropomorphic related traits). Together these publications contain 914 SNPs associated with CHD or CHD risk factors.

The GWAS approach has delivered many SNPs for complex traits and with the progress made through meta-analysis leading to stable risk estimates for each SNP, the usefulness of these SNPs for CHD risk prediction should be tested. A limitation of GWAS is that every identified SNP has a very small effect size.\textsuperscript{55} Because of the small effect sizes of SNPs
associated with CHD or its risk factors, SNPs were combined into a genetic risk score (GRS).\textsuperscript{56-58} It was expected that such a GRS, based on SNPs associated with CHD or CHD risk factors is associated with incident CHD and perhaps can improve CHD risk prediction beyond the TRFs on which current CHD risk scores are based.\textsuperscript{56-58}

Metabolic health

To find risk markers that may improve CHD risk prediction, most studies focus on finding risk factors for CHD. One may question if a low CHD risk is simply caused by the absence of risk factors or that there are protective factors involved, protecting individuals against CHD even in the presence of established CHD risk factors. To answer this question it is relevant to study which factors mark metabolic health.

This can be done in the offspring of long-lived individuals that are characterized by a lower prevalence of age-related diseases, including diabetes, hypertension CHD and stroke.\textsuperscript{59-62} This indicates that these long-lived individuals have a genetic or familial factor that somehow protects them against the development of metabolic diseases, which make them interesting for studying which factors are involved in healthy ageing and longevity and thus possible markers that are protective against CHD.

Box 2 | Studies used in this thesis

Cardiovascular Registry Maastricht cohort study - The Cardiovascular Registry Maastricht (CAREMA) cohort study was set up to study coronary heart disease (CHD) and associated risk factors.\textsuperscript{63} The cohort consists of 21,148 participants born between 1927 and 1977 who were randomly sampled from Maastricht and surrounding communities between 1987 and 1997.\textsuperscript{63} This cohort was followed until December 2003. During this follow-up period, 742 incident CHD cases (i.e. acute myocardial infarction, unstable angina pectoris or death due to CHD) occurred.\textsuperscript{63,64} From the CAREMA cohort a sub-cohort was drawn, consisting of 2221 participants, including 116 cases. Only the cases experiencing a CHD event and the members of the sub-cohort are included in the data-analysis.\textsuperscript{65} This type of study design is called a case-cohort study.\textsuperscript{65}

Erasmus Rucphen Family study - The Erasmus Rucphen Family (ERF) study is a large, family-based study with participants from a genetically isolated community located in the Southwest of the Netherlands.\textsuperscript{66} The ERF study includes individuals that are the living descendants of 22 couples that had at least six children baptized in the community church between 1850 and 1900. Participants were thus not selected based on disease of interest.\textsuperscript{66} Participants were classified as a CHD case when they indicated that they had experienced a myocardial infarction, underwent a coronary revascularization procedure or if they reported angina symptoms during an interview with the study physician. Participants were also classified as a CHD case if they showed signs of myocardial infarction on the electrocardiogram. In this study of 2415 participants, 170 participants were classified as a CHD case and 2157 participants were categorized as controls.

Leiden Longevity Study - This study was set up to investigate if there was evidence of genetic enrichment for exceptional survival using a family approach.\textsuperscript{67} For this study large series of sib pairs aged 90 years and over were collected.\textsuperscript{67} In addition the offspring of the long-lived siblings and the partners of the offspring were also collected for further study.\textsuperscript{67} The recruitment for this study started in July 2002.\textsuperscript{67} In total 421 families were included, representing 982 long-lived siblings, their offspring (n=1671) and the partners of the offspring (n=745).
Chapter 1

Aims of this thesis
The main goal of this thesis was to assess if CHD risk prediction based on the TRFs could be improved by adding novel risk markers. First it was tested if metabolomics data could improve CHD risk prediction beyond TRFs. Second it was investigated if genetic markers could improve CHD risk prediction. A secondary goal was to find out if a comprehensive lipid profile could mark metabolic health as opposite to age-related diseases like CHD in middle-aged individuals. Finally, as a first step in understanding more of CHD pathology we combined metabolomics data (a comprehensive fatty acid profile) with genetic information on the FADS1 gene. See box 2 for the studies used to investigate these questions.

Outline of this thesis
In chapter 2 we investigated if a metabolomics approach could be useful for improving and understanding CHD risk prediction beyond TRFs. Therefore we obtained a metabolic profile using $^1$H-NMR spectroscopy in the individuals who entered the Cardiovascular Registry Maastricht (CAREMA) cohort study from 1993 until 1997 and were selected for the sub-cohort or experienced a CHD event during follow-up. Using least absolute shrinkage and selection operator (LASSO) regression, a subset of the most informative metabolites for predicting incident CHD was obtained. Subsequently, it was tested if a score based on these metabolites was associated with incident CHD within the same case-cohort study. Using the Erasmus Rucphen Family (ERF) study it was tested if the selected metabolites were relevant for differentiating between prevalent CHD cases and CHD-free controls.

In chapter 3 we compared two methods for data reduction, our own targeted approach explained in chapter 2 and a more commonly used method which involves partitioning the $^1$H-NMR spectra into discrete sections and calculate the sum of the spectral intensity in each section. Moreover, we also evaluated the impact of two scaling methods (i.e. autoscaling and pareto scaling) which are commonly used as data pre-treatment methods in metabolomics data-analysis.

In chapter 4, as apposite to the development of CHD, factors that promote longevity and metabolic health were studied. Therefore offspring of long-lived siblings were compared to their same-aged spouses representing the general population. Siblings are considered long-lived when they are 90 years or older and have a brother or sister who is also 90 years or older. These long-lived siblings could not be easily compared a control group since most subjects in their birth cohort were already deceased. However at middle-age their offspring also showed less overall mortality, and metabolic disease compared to the general Dutch population. Therefore the offspring of long-lived siblings were compared to their same...
aged partners. With this approach it is possible to study the differences between people that have a tendency to become long-lived and people with an average life expectancy. The focus for this chapter was on lipid metabolism. Thus a comprehensive lipid profile was obtained using $^1$H-NMR spectroscopy.\textsuperscript{73}

In chapter 5 it was investigated if GWAS identified SNPs known to be associated with CHD or CHD risk factors (i.e. blood pressure, diabetes type 2, blood lipid levels and anthropomorphic related traits) could improve CHD risk prediction beyond traditional CHD risk factors. Based on these SNPs GRSs were constructed. First we focused on counted GRSs. The first counted GRS was based on SNPs associated with CHD and CHD risk factors, the second one was based on SNPs associated with CHD risk factors and the third one was based on SNPs associated with CHD only. The fourth GRS was also based on SNPs associated with CHD, but in contrast to the counted GRS, we adjusted for the different effect sizes of the CHD associated SNPs. The effect sizes for these SNPs were based on two large meta-analysis.\textsuperscript{74,75} As an exploratory approach, we selected a subset of the most informative SNPs using LASSO regression.

In chapter 6, the fatty acid metabolism was explored for gaining more understanding in the development of CHD. Lipid metabolism, including fatty acid metabolism plays an important role in CHD etiology. For example, a high intake of polyunsaturated fatty acids reduces CHD risk.\textsuperscript{76-78} The fatty acid intake is often estimated by assessing the fatty acid contents in biological tissues.\textsuperscript{79} This profile, however, does not only reflect dietary intake, but also fatty acid metabolism in the body itself,\textsuperscript{79} in which $\delta$-5 and $\delta$-6 desaturase play an important role.\textsuperscript{80,81} In this chapter we explored the role of $\delta$-5 and $\delta$-6 desaturase in CHD, in addition we investigated the influence of a genetic variation in the fatty acid desaturase 1 gene.
References


46. Sundquist K, Winckley M, Li X, Ji J, Hemminki K, Sundquist J. Familial transmission of coro-
Chapter 1


