Interaction of Genetic Susceptibility and Traffic-Related Air Pollution in Cardiovascular Disease

Anna Levinsson

Occupational and Environmental Medicine
Department of Public Health and Community Medicine
Institute of Medicine
Sahlgrenska Academy at University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2015

Cover illustration: ‘1.2’ by Anna Levinsson/Wordle

Interaction of Genetic Susceptibility and Traffic-Related Air Pollution in Cardiovascular Disease

anna.levinsson@amm.gu.se

ISBN 978-91-628-9279-1

Printed in Gothenburg, Sweden 2015
Aidla Trading AB/Kompendiet

*For **Whizzski Lad**, my life companion & partner in crime.*
Who moved with me to this rainy place.
We have weathered the tough times together,
I hope that the future holds more light.

Interaction of Genetic Susceptibility and Traffic-Related Air Pollution in Cardiovascular Disease

Anna Levinsson

Occupational and Environmental Medicine
Department of Public Health and Community Medicine, Institute of Medicine
Sahlgrenska Academy at University of Gothenburg, Göteborg, Sweden

ABSTRACT

This thesis aimed at investigating gene-environment interaction in cardiovascular disease (CVD). A study population of 618 coronary heart disease (CHD) cases (of which 192 first-time acute myocardial infarction (AMI) patients) and 3614 randomly selected population controls was genotyped for genetic variants in genes coding for nitric oxide synthase (NOS) and glutathione s-transferase (GST). Exposure to traffic-related air pollution was assessed using modeled mean annual concentrations of nitric dioxide (NO_2) as a marker for long-term exposure.

Among 58 single nucleotide polymorphisms (SNPs) in the *NOS1*, *NOS2* and *NOS3* genes investigated for risk of CHD and hypertension, several strong associations were found, some of which remained statistically significant after Bonferroni correction for multiple testing. The T-allele of *NOS1* SNP rs3782218 was significantly associated with a protective effect for both CHD (odds ratio (OR) 0.6, 95% confidence interval (CI) 0.44-0.80) and hypertension (OR 0.8, 95% CI 0.68-0.97). A second study investigated SNPs in the genes *GSTP1*, *GSTT1* and *GSTCD* for interaction with traffic-related air pollution on risk of AMI and hypertension. The risk of AMI from air pollution exposure seemed to vary by genotype strata (for example *GSTP1* SNP rs596603 with OR 2.1, 95% CI 1.09-4.10 in the genotype TT+GT stratum; OR 1.4, 95% CI 0.73-2.68 in the genotype GG stratum, although the multiplicative interaction was not significant (p-value =0.27)). Finally, the methodology of estimating additive interaction between a dichotomous (e.g. genetic) variable and a continuous (e.g. air pollution) variable using output from a logistic regression model was investigated in detail. The measure of additive interaction in this setting was shown to be highly sensitive to variation in the parameters defining it, and a pragmatic proposal for controlling this variability when extending estimation of additive interaction to new settings was developed. The proposed method was applied to the *GST* genotype and air pollution exposure data to estimate the additive interaction of these exposures on risk of AMI, finding a sub-additive interaction effect for the *GSTCD* AG+GG genotype.

To conclude, the results of this thesis indicate that *NOS* gene variants are associated with both CHD and hypertension, and that variants in the *GST* genes are of importance regarding the risk of hypertension and the risk of AMI due to air pollution exposure.

Keywords: Cardiovascular disease, genetic variants, air pollution, gene-environment interaction

ISBN (printed): 978-91-628-9279-1 **ISBN (e-publ):** 978-91-628-9280-7

SAMMANFATTNING PÅ SVENSKA

Hjärt-kärlsjukdom i dess olika former är den vanligaste dödsorsaken världen över enligt Världshälsoorganisationen (WHO). Även om antalet dödsfall i västvärlden har minskat, tack vare förbättrade riskfaktorer i befolkningen och effektivare behandlingsmetoder, är hjärt-kärlsjukdom den vanligaste orsaken till sjukdom och död. Detta innebär att det är av fortsatt värde att bedriva forskning om hjärt-kärlsjukdomarnas etiologi, dvs. vad som orsakar dem. Ett antal olika riskfaktorer, såsom rökning, kolesterol, hypertoni (högt blodtryck), diabetes, övervikt och stillasittande livsstil, anses idag vedertagna, men de förklarar inte hela risken.

Den här avhandlingen syftar till att undersöka huruvida olika riskfaktorer, närmare bestämt genetiska variationer och exponering för luftföroreningar från trafik, tycks interagera när det gäller risk för hjärtkärlsjukdom.

De diagnoser som använts är kranskärlssjukdom, akut hjärtinfarkt och hypertoni. Genetiska varianter i två grupper av gener, *NOS* respektive *GST*, har studerats. NOS (nitric oxide synthase = kväveoxidsyntas) fungerar bland annat som signalsubstans i hjärnan, blodkoncentrationen av den ökar vid inflammation och den är en del av kemin när blodkärl vidgas och drar ihop sig. GST (glutathione s-transferase) är en antioxidant som hjälper till att motverka de skadliga effekterna av syreradikaler, så kallade oxidanter, i kroppen. Luftföroreningar, t.ex. från trafik, har visats vara kopplade till ökad risk för hjärt-kärlsjukdom. I den här avhandlingen har exponeringen för luftföroreningar från trafik beräknats på så sätt att varje studiedeltagares adress har omvandlats till en koordinat och med hjälp av ett geografiskt informationssystem kopplats till ett värde på årsmedelvärdeshalten NO_2 (kvävedioxid) och NO_x (kväveoxider = kvävedioxid + kvävemonoxid). Studiedeltagarna består av 618 patienter med kranskärlssjukdom och 3614 slumpvis utvalda individer från Västra Götalandsregionen. Alla deltagare genomgick en medicinsk undersökning där bland annat blodtryck, längd och vikt mättes. De fick också fylla i frågeformulär med medicinska såväl som livsstilsfrågor (vilka mediciner personen äter, utbildningsgrad, rökvanor osv).

Resultaten av avhandlingens tre delprojekt kan sammanfattas som att en genetisk variant i *NOS1* genen sågs vara signifikant associerad med både kranskärlssjukdom och hypertoni, med en skyddande effekt för den mindre vanliga varianten. Som markör för luftföroreningar från trafik var NO_2 starkt kopplat till ökad risk för hjärtinfarkt. Effekten av trafikrelaterade luftföroreningar tycktes variera beroende på vilken genetisk variant av *GST*-

generna en individ har. Under arbetet med att undersöka interaktionen mellan genvariationer och luftföroreningsmarkörer identifierades en metodologisk svårighet med att undersöka den additiva interaktionen, dvs om den totala effekten av två exponeringar avviker från summan av deras respektive effekter, t.ex. att den totala effekten är större än summan, när en av exponeringarna mäts som en kontinuerlig variabel. Ett förslag på ett praktiskt inriktat tillvägagångssätt för beräkning av storlek och riktning för en eventuell avvikelse presenterades, som en vidareutveckling av en tidigare känd metod. Tillvägagångssättet tillämpades också på observerade data i fallet med en kategorisk och en kontinuerlig exponeringsvariabel.

Slutsatsen är att resultaten från denna avhandling visar på att varianter i NOS-gener är associerade med både CHD och hypertoni samt att GST-gener är betydelsefulla när det gäller risken för hjärtkärlsjukdom som följd av exponering för luftföroreningar.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Levinsson A, Olin AC, Björck L, Rosengren A, Nyberg F (2014) *Nitric oxide synthase (NOS) single nucleotide polymorphisms are associated with coronary heart disease and hypertension in the INTERGENE study*. Nitric Oxide 39:1-7.

II. Levinsson A, Olin AC, Modig L, Dahgam S, Björck L, Rosengren A, Nyberg F (2014) *Interaction effects of long-term air pollution exposure and variants in the GSTP1, GSTT1 and GSTCD genes on risk of acute myocardial infarction and hypertension: a case-control study*. PLoS One 9(6): e99043.

III. Levinsson A, Olin AC, Ding B, Björck L, Rosengren A, Nyberg F. *Additive interaction involving a continuous variable: a pragmatic approach*. Manuscript.

CONTENT

ABBREVIATIONS IV
DEFINITIONS IN SHORT V
1 INTRODUCTION 1
1.1 Coronary heart disease, hypertension and their known risk factors 1
1.2 Traffic-related air pollution and cardiovascular disease 3
1.2.1 Traffic-related air pollution 3
1.2.2 Inflammatory pathway or direct pathway? 4
1.3 Genetic variation in cardiovascular disease 4
1.4 Gene-environment interactions in cardiovascular disease 5
1.5 Methods for measuring interaction in case-control data 6
1.5.1 Multiplicative and additive interaction 7
1.5.2 Measures of additive interaction 7
1.5.3 Effect measure modification 8
1.5.4 Estimating additive interaction involving a continuous variable 9
2 AIM 10
2.1 Specific aims for each paper 10
3 PATIENTS AND METHODS 11
3.1 Study population and data collection 11
3.2 Air pollution exposure assessment 13
3.3 Genetic analysis 14
3.3.1 SNP genotyping 14
3.3.2 Genetic models and genotype coding 17
3.3.3 Statistical methods in genetic data analysis 18
3.4 Estimating RERI for GST and air pollution data from Paper II, using methodology from Paper III 21
4 RESULTS 23
4.1 Paper I 23
4.2 Paper II 24

4.3 Paper III 26
4.4 Estimating RERI for GST and air pollution data from Paper II, using methodology from Paper III 28
5 DISCUSSION 30
6 CONCLUSION 37
6.1 Paper-specific conclusions 37
7 FUTURE PERSPECTIVES 39
ACKNOWLEDGEMENT 41
REFERENCES 43
APPENDIX 52

ABBREVIATIONS

AMI	Acute myocardial infarction
AP	Attributable proportion (due to interaction)
BMI	Body mass index
CHD	Coronary heart disease
CI	Confidence interval
CV	Cardiovascular
CVD	Cardiovascular disease
DBP	Diastolic blood pressure
GST	Glutathione S-Transferase
HWE	Hardy Weinberg equilibrium
NO_2	Nitrogen dioxide
NO_x	Nitrogen oxides (nitrogen oxide and nitrogen dioxide)
NOS	Nitric oxide synthase
OR	Odds ratio
RERI	Relative excess risk due to interaction
SBP	Systolic blood pressure
SNP	Single nucleotide polymorphism

DEFINITIONS IN SHORT

Call rate	The percentage of genotyped individuals that had a successful genotyping (ending up with a result) for a particular assay. If the call rate is low, e.g. below 90%, it is suspected that the assay used is incorrect.
Coronary vessel or coronary artery	Blood vessel supplying the myocardium [Persson 1986]
Diplotype	The set of haplotypes in an individual's DNA. [Marchenko et al. 2008]
Ever-smoker	In this dissertation and included publications: A person who has either been a smoker and quit smoking, or is still a smoker.
Former smoker	In this dissertation and included publications: A person who used to smoke daily but quit at least 12 months ago.
Genotype	An individual's genetic constitution at a given locus [Jorde et al. 2005], i.e. the combination of 2 alleles (one on each chromosome copy) at a single locus.
Haplotype	Sequence of genetic markers on the same chromosome within a genomic region of interest. [Marchenko et al 2008]
Hypertensive	In this dissertation and included publications, a person referred to as 'hypertensive' will have at least one of the 3 following characteristics: a) SBP ≥140 mmHg, b) DBP ≥90 mmHg, c) using antihypertensive medication daily.

Hardy Weinberg equilibrium	The population frequencies of genotypes AA, Aa and aa are predicted based on allele frequencies p and q, where A = p, a =q, AA = p^2, Aa = aA = pq and aa = q^2. Since the probability of having a genotype at all is 1, $p^2 + 2pq + q^2 = 1 \Leftrightarrow (p + q) = 1 \Leftrightarrow p = 1 – q$ [Jorde et al. 2005] The assumption is that a genotype distribution (in a population fulfilling the underlying assumptions of large population, random mating, no migration, no mutations and no natural selection) fulfills the conditions above, and if genotype distribution deviates significantly from these conditions, a practical interpretation is that the genotyping process and/or the material used for genotyping may be flawed or contaminated.
Locus	A location on a chromosome (from Latin meaning “place”), for e.g. a gene, SNP or other genetic characteristic. [Jorde et al 2005]
Phenotype	The trait which is observed physically or clinically. In epidemiology often: affected / not affected. [Jorde et al. 2005)
Resistance	The insensitivity of the results of a procedure to small changes in the data. [Andrews 1998]
Robustness	The ability of a model to be insensitive to small changes in the assumptions which specify it. [Andrews 1998]
Single nucleotide polymorphism	Single nucleic base difference in the DNA sequence [Jorde et al. 2005], with minor allele frequency ≥ 1%.

1 INTRODUCTION

While the expression ‘nature or nurture?’ used to drive the research for several diseases, it seems that modern research has largely moved on to ask how nature and nurture interact, i.e. to studies of gene-environment interaction. [Pigliucci 2001, Steele 2014, LaBaer 2002] The current disease pathology paradigm is that risk factors do not act alone, but rather in different formations to cause disease. Of the known risk factors for cardiovascular disease (CVD), the modifiable risk factors smoking, cholesterol levels and hypertension are the three most important [Yusuf et al. 2004], but non-modifiable risk factors such as age and male sex are also of importance. [WHO 2013] Nonetheless, the pattern of how the risk factors connect to form a web of disease risk probabilities still needs to be investigated further.

1.1 Coronary heart disease, hypertension and their known risk factors

Coronary heart disease (CHD) is an umbrella term for several cardiologic diagnoses affecting the coronary heart vessels, which provide the blood supply to the heart. [WHO 2013] CHD includes for example angina pectoris (chest pain due to restricted blood flow) and acute myocardial infarction (AMI). CVD is an even wider umbrella term, including CHD but also other diseases of the heart and diseases of vessels other than the coronaries.

According to the World Health Organization, CVD continues to be the number one cause of death globally. [WHO 2013] Most of these deaths are caused by AMI. The dominant cause of acute CHD, including AMI, is atherosclerosis. [Nilsson 2010] Often atherosclerosis is one of the first recognizable signs of CVD. The pathological mechanisms which initiate and drive atherosclerosis are not fully elucidated, but inflammation is considered one of the major processes that contribute to atherogenesis. [Ikonomidis et al. 2012] At an early stage of disease, inflammation acts in a protective manner against atherosclerosis by absorbing oxidized LDL before it damages the vessel wall. [Nilsson 2010] If, despite the countermeasures, oxidized LDL ingested by macrophages to form foam cells, gathers in the vessel wall in formations called plaques or fatty streaks [Ahlner & Johansson 1994] (Figure 1), the inflammatory response is increased, which reduces the ability to

sustain immunological tolerance towards the oxidized LDL. At this point, inflammation becomes the driving mechanism of atherosclerosis. [Nilsson 2010] During atherosclerosis formation, lipids and inflammatory cells are accumulated in the vessel wall in formation called plaques. While the plaque formation is mostly located in the intima, changes also occur in other parts of the cell wall. The underlying media is often atrophic and containing a decreased number of muscle cells. Inflammation plays an important role not only in the initiation and progression of atherosclerosis but also in plaque rupture, an event that leads to acute vascular events. [Ikonomidis et al. 2012]

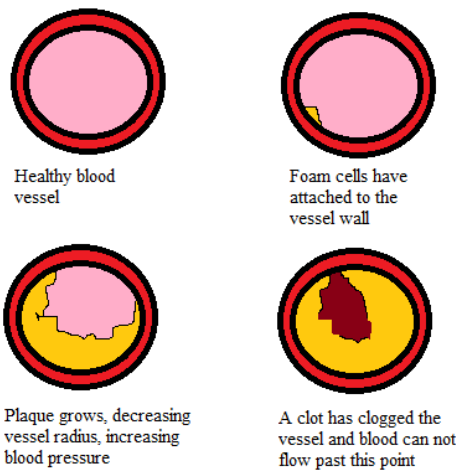

Figure 1. Illustration of gradual plaque build-up. Drawing by Anna Levinsson

The formation of plaques often decreases the radius of the vessel lumen and causes the vessel walls to become more rigid, both of which increase the blood pressure. In turn, the hypertension causes an increase of the inflammatory effects in the vessel by putting more stress on the vessel walls, and also increases the risk of an unstable plaque rupturing. [Ahlner & Johansson 1994] Generally, an AMI occurs by obstruction of a blood vessel because of a local obstruction and a local final clot, or sometimes by embolic obstruction due to a clot originating from a ruptured coronary plaque. [Ahlner & Johansson 1994] As a result, a vessel becomes completely obstructed, thus cutting off the blood flow past the point of obstruction, i.e. to a portion of the heart.

Some of the classic known risk factors for hypertension and CHD are modifiable lifestyle risk factors including smoking, high blood lipid levels, hypertension, diabetes, obesity and physical inactivity. [Yusuf et al. 2004, WHO 2013] Despite this knowledge, people continue to smoke and engage in hazardous lifestyle behavior. In the western world, CVD mortality and morbidity decreases, due to better and more swiftly applied health care. In developing countries, health care is less available and mortality rates rise as CVD morbidity increases with current trends in lifestyle changes. [WHO 2013]

1.2 Traffic-related air pollution and cardiovascular disease

Air pollution is a significant risk factor for human morbidity: the World Health Organization estimates that in 2012, 7 million unnatural deaths were caused by ambient air pollution. [WHO 2013] Of these air pollution-related deaths, 40% were from ischemic heart disease.

1.2.1 Traffic-related air pollution

One of the main sources of human everyday air pollution exposure today is traffic. Traffic-related air pollution consists of a mixture of particles of varying size and gases, including large amounts of NO_x. [HEI 2010] Thus, NO_x or NO_2 is often used as a marker for traffic-related air pollution exposure. [Coogan et al. 2012, Vermylen et al. 2005] While the specific mechanisms of traffic-related air pollution effects on human health are not known, the (mainly pulmonary) exposure, both long-term and acute, to particles and gases has been found to be associated with human disease, including CVD. [Brook et al. 2010, Brook & Rajagopalan 2009, Brunekreef 2007, Peters 2005]

Several studies link ambient air pollution and AMI. A review of epidemiological studies [Vermylen et al. 2005] reported adverse associations between chronic exposure to ambient air pollution and the outcomes cardiovascular mortality, cardiopulmonary mortality and increased intima-media thickness, an indicator of atherosclerosis. The strongest association was a nearly doubled risk of cardiopulmonary mortality when living near a major road.

1.2.2 Inflammatory pathway or direct pathway?

The particulars of the biological mechanisms by which pulmonary exposure to air pollution leads to CVD outcomes are not fully understood. [Zanobetti, Baccarelli & Schwartz 2011] One potential pathophysiological pathway is that pulmonary exposure to air pollution induces local pulmonary oxidative stress, which leads to release of pro-thrombotic and inflammatory cytokines into the blood stream, as well as an increased level of reactive oxygen species (ROS) in the heart. [Zanobetti, Baccarelli & Schwartz 2011, Shrey et al. 2011, Bessa et al. 2009] When the pulmonary stress responses are insufficient to handle the levels of ROS, a range of pulmonary inflammatory processes are activated, which enhances expression of inflammatory cytokine genes, in turn inducing systemic inflammation and systemic oxidative stress. Inflammation furthers progress of atherosclerosis and can potentially trigger acute plaque rupture. [Campen et al. 2012] The release of pro-thrombotic agents into the blood stream can also trigger clot formation and put the individual at increased risk of ischemic heart disease, especially if vessels are atherosclerotic, i.e. already inflamed and more vulnerable. [Siegbahn 2010]

Besides the inflammatory pathway, other mechanisms have been suggested, for example direct translocation of particles across the pulmonary epithelium and lung-blood barrier into the cardiovascular system, i.e. penetrating cellular membranes, which has been shown experimentally in both animals and humans. [Peters et al. 2006, Vermylen et al. 2005] Once the particles have reached the blood, they may reach specific organelles in the blood cells, or induce the release of cytokines and inflammatory mediators throughout the body by way of the cells. This is sometimes referred to as the direct pathway. [Peters et al. 2006]

1.3 Genetic variation in cardiovascular disease

Previous research has identified some associations between the genes investigated in this thesis (*NOS1*, *NOS2*, *NOS3*, *GSTP1*, *GSTT1* and *GSTCD*), and the CVD outcomes studied here or related outcomes. SNPs in *NOS1* have been associated with blood pressure [Iwai et al. 2004, Padmanabhan et al. 2010], and *NOS1* has been identified as a candidate gene for stroke [Meschia et al. 2011]. A copy-number variation in *NOS2* has been linked to CHD and CV events. [Tepliakov et al. 2010, Gonzales-Gay et al. 2009] *NOS3* is the most studied of the three genes and SNPs in this gene

have been associated with different CHD manifestations including myocardial infarction, as well as treatment-resistant hypertension and ischemic stroke. [Johnson et al. 2011, Casas et al. 2006, Hingorani et al. 1999, Jàchymovà et al. 2001, Berger et al. 2007, Niu & Qi 2011] In addition, variants in the NOS genes have been investigated regarding pulmonary outcomes, including lung function and chronic obstructive pulmonary disease [Aminuddin et al. 2013], and inhibition of *NOS2* function has been associated with reduced pulmonary fibrosis [Janssen et al. 2013]. Inducible NOS (expressed by the *NOS2* gene) has also been implicated in many inflammatory diseases, and expression of inducible NOS can be induced by inflammatory stimulants and mediators. [Förstermann & Sessa 2012] Variants in *NOS2* and *NOS3* genes have been associated with airway inflammation. [Dahgam et al. 2012]

The gene deletion causing the *GSTT1* polymorphism results in almost no enzymatic activity in individuals with the null genotype, potentially putting them at increased risk of oxidative stress and inflammation. [Stephens, Bain and Humphries 2008, Pemble et al. 1994] The GSTT1 null polymorphism has previously been studied regarding association with CHD with inconclusive results. [Nørskov 2013, Du et al. 2012] For variants in *GSTP1*, no significant interactions for CVD have been reported, as far as we know. However, SNPs in *GSTP1* have also been investigated regarding associations with lung function, and have been shown to modify the effect of air pollution on lung function. [Mordukhovich et al. 2009, Probst-Hensch et al. 2008] Thus, considering the inflammatory pathway, an association with CVD outcomes is possible. Associations between SNPs in the *GSTCD* gene and pulmonary function have also been reported, and are supported by a meta-analysis. [Repapi et al. 2010, Hancock et al. 2010] In addition, variants in the GST genes have been tentatively associated with other health outcomes, e.g. asthma and several types of cancer. [Minelli et al. 2009, White et al. 2008, Dunning et al. 1999]

1.4 Gene-environment interactions in cardiovascular disease

Without making assumptions about which, if any, of the inflammatory and the direct pathway is correct, it still seems plausible that genes with antioxidative and inflammatory effects may be involved in the mechanism underlying the association of air pollution exposure with CVD. Consider one

amino acid sequence of DNA which may be associated with production of an antioxidant defense adequate for holding back an exposure-induced inflammatory onslaught. If a mutation occurs in this sequence, one genotype may be synonymous with the original nucleotides and the protein synthesis will function normally, while another genotype may change the DNA sequence. The result of a change may be a different protein sequence or a truncated protein sequence, which affects regulation, or a change in splicing. All of these may result in changed protein function which may cause less or no production of a protein, which may upset the redox homeostasis. [Young et al. 2006, Wang et al. 2001]

A review of studies investigating gene-environment interaction in relation to cardiovascular health effects showed that genes in the oxidative stress pathway modify the risk of CVD due to air pollution exposure. [Zanobetti, Baccarelli & Schwartz 2011] Several studies have also investigated interaction between APOE genotype and an environmental exposure variable in CHD risk. [Gustavsson et al. 2012, Talmud 2007] One such study investigated multiplicative interaction effect between smoking habits and APOE genotype on risk of CHD and found a statistically significant interaction. [Talmud 2007]

1.5 Methods for measuring interaction in case-control data

Logistic regression is the work horse of epidemiology for estimating odds ratios as effect estimates of relative risk when the outcome is dichotomous, e.g. diseased / not diseased. [Skrondal 2003] Since it is inherently multiplicative, all analyses of statistical interaction using results from logistic regression are on the multiplicative scale. [Ahlbom & Alfredsson 2005] In case-control data, absolute risks cannot be estimated directly because the underlying sampling fractions are unknown. [Rothman, Greenland & Lash 2008] However, under appropriate control sampling conditions, the odds ratio from logistic regression can be equivalent with the risk ratio and can also give estimates of the rate ratio and the incidence odds ratio. Under the ‘rare disease assumption’, each of the measures is also an approximate estimate of the others. [Pearce 1993, Greenland & Thomas 1982] The purpose of the epidemiologic studies within this thesis is to understand disparities in disease risk between groups, and considering the reasoning

above, the odds ratio obtained from logistic regression is an appropriate measure for such studies.

When investigating the joint effects of genetics and environmental exposure on the risk of an outcome, there is a need to define the characteristics of this interaction and to find a suitable measure for it. In current epidemiology, two kinds of interaction are mainly discussed, namely additive and multiplicative, sometimes referred to as biological and statistical. [Kaufman 2009]

1.5.1 Multiplicative and additive interaction

In the estimation of relative risk by regression methods, e.g. analysis of case-control data using logistic regression, the insertion of a product term of two exposure variables of interest gives an estimate of the multiplicative interaction between the two exposures, per variable unit. Additive interaction cannot be directly estimated in a logistic regression model, but in the case of two dichotomous variables, methods for using the output from the logistic regression to calculate an estimate of additive interaction, for example RERI, are fairly well characterized. [Knol et al. 2007]

1.5.2 Measures of additive interaction

RERI (Relative Excess Risk due to Interaction) is one measure of additive interaction developed by Rothman [Rothman 1986], originally expressed as

$$\mathbf{RERI} = R_{11} - R_{10} - R_{01} + R_{00} \qquad \mathbf{[1]}$$

where $R_{jk} \equiv P(Y = 1|x_1=j, x_2=k)$ is the conditional risk or probability that the outcome variable Y takes the value 1 given the values j, k of the exposures x_1, x_2. The equation can also be expressed using risk ratios (RR) by dividing all factors by the baseline risk R_{00}:

$$\mathbf{RERI}_{RR} = RR_{11} - RR_{10} - RR_{01} + 1 \qquad \mathbf{[2]}$$

When estimating the risk ratios using logistic regression, odds ratios replace the risk ratios in the formula and the formula can be rewritten with the beta coefficients obtained from a logistic regression for a dichotomous outcome.

$$\mathbf{RERI} = e^{\beta_1+\beta_2+\beta_3} - e^{\beta_1} - e^{\beta_2} + 1 \qquad \mathbf{[3]}$$

In the simple form expressed in equations [1], [2] and [3] above, both exposures are assumed to be dichotomous. The regression model consists of

the two exposures (coefficients β_1 and β_2), their product term (coefficient β_3) and any relevant covariates. If RERI = 0, the interpretation is that the joint effect of the two exposures is equal to the sum of their main effects, meaning there is no additive interaction. If RERI ≠ 0, there is deviation from additivity of risks and the precision of the estimate can be evaluated using confidence intervals. Confidence intervals can be calculated using different techniques, for example bootstrapping or the Wald-type method originally presented by Hosmer & Lemeshow. [Richardson & Kaufman 2009, Hosmer & Lemeshow 1992]

Other measures of additive interaction are available, such as the synergy index (S) and attributable proportion (AP). However, as AP is simply a function of RERI $\left(\text{expressed with risk ratios, } \mathbf{AP} = \frac{RERI}{RR_{11}}\right)$ it can easily be calculated along with RERI if one prefers a measure interpreted as the attributable proportion of disease which is due to interaction among persons with both exposures. However, AP is not defined for negative interaction ($RR_{11} < 0$) as the proportion would then be negative. [Skrondal 2003]

S, expressed with risk ratios, is defined $\mathbf{S} = \frac{R_{11}-1}{(R_{10}-1)+(R_{01}-1)}$. The measure in focus for this thesis was RERI, which has been recommended as the preferred measure by some authors [Knol & VanderWeele 2012, VanderWeele 2011].

1.5.3 Effect measure modification

A method for evaluating the presence of interaction that works well for continuous variables is effect measure modification, or heterogeneity of effects as it is also called. [Rothman, Greenland & Lash 2008, Greenland & Morgenstern 1989] In practice, the method amounts to stratifying for one variable and estimating the exposure effect for an outcome in each stratum, then comparing effect estimates across strata. If the stratum-specific effect estimates are equal, the measure is said to be homogenous, constant or uniform across strata, while if it is not, it is said to be heterogeneous, modified or varying across strata. [Rothman, Greenland & Lash 2008] When investigating effect measure modification using linear regression analysis models (i.e. for a continuous phenotype), effect measure modification is equivalent with additive interaction, and when using logistic regression analysis models, i.e. with the effect estimates expressed as odds ratios, effect measure modification is equivalent with multiplicative interaction. [Greenland 2009, Rothman, Greenland & Lash 2008] The latter form of the method is used in Paper II to study air pollution effect measure modification

by genetic variants in *GST*-genes for outcomes AMI and hypertension. [Paper II]

1.5.4 Estimating additive interaction involving a continuous variable

A recently published article suggested that estimating RERI using continuous variables was possible, if the baseline and interval size (increment) for each variable was explicitly defined. [Katsoulis and Bamia 2014] However, a major problem with estimating additive interaction involving a continuous variable, using effect estimates from logistic regression, is that for the continuous variable, there is not one unequivocal estimate, but rather an infinite set of estimates, with the estimate of RERI depending on the interval where the additive interaction is estimated and the variable units. [Paper III, Knol et al. 2007] This is not consistent with the original definition of RERI, where for a given dataset, the interaction parameter estimate was seen to be constant. [Rothman 1986] The RERI measure is sensitive to variations in the parameters defining it, which was a focus of study in this thesis and will be further discussed in the Results sections for Paper III and Chapter 5 (Discussion).

2 AIM

The overall aim of this thesis was to study the main effects of genetic variants in genes associated with oxidative stress and inflammation on the outcomes CHD, hypertension and AMI, as well as to study cardiovascular effects of traffic-related air pollution in interaction with genetic variants in the *GST* gene family.

2.1 Specific aims for each paper

I. The overall aim was to comprehensively investigate main effects of polymorphisms in the *NOS* genes on risks of both CHD and hypertension in the same source population. The first aim was to determine which of the *NOS* genes and SNPs were most strongly associated with the two CV phenotypes. Then, recognizing that multiple SNPs in the same gene can be markers for the same effect, a second aim was to explore this aspect with haplotype analyses.

II. The first aim was to investigate main effects of long-term traffic-related air pollution exposure, as well as variants in *GSTP1*, *GSTT1* and *GSTCD*, on risk of acute myocardial infarction (AMI) and hypertension. The second, major, aim was to study whether air pollution effects were modified by the investigated genetic variants.

III. This was a methodological exploration, aiming to identify the various problems with estimating additive interaction for a dichotomous outcome and involving a continuous variable, and to propose a pragmatic approach for generating more interpretable and consistent results based on logistic regression coefficient estimates.

3 PATIENTS AND METHODS

3.1 Study population and data collection

The INTERGENE/ADONIX (INTERplay between GENEtic susceptibility and environmental factors for the risk of chronic diseases in West Sweden / ADult-Onset asthma and exhaled NItric oXide) study was the source of the data used for this thesis. From April 2001 until December 2004, INTERGENE/ADONIX recruited CHD cases and a population control cohort from the greater Gothenburg area in Sweden. All participants were aged 25-75 years at the time of selection. [Berg et al. 2008, Berg et al. 2005] For the population control cohort, 8820 randomly selected individuals were invited to participate in the study. 194 of these had either left the country, moved to a different part of Sweden, were deceased or had an unknown address. [Strandhagen et al. 2010] Of the remaining 8626 eligible individuals, 3614 participated, which yields a participation rate of 41.9%. As CHD cases, the study included consecutive inpatients admitted to wards at 3 locations (Östra, Mölndal and Sahlgrenska) of the Sahlgrenska University Hospital, Gothenburg, Sweden or outpatients with significant coronary lesions identified from coronary angiograms. Altogether, the INTERGENE/ADONIX study included 618 CHD patients (73.4% men and 26.6% women), 295 with a first episode of acute myocardial infarction (AMI) or unstable angina pectoris, and the remainder with chronic CHD, defined as either prior AMI or positive angiogram. 192 patients were individuals presenting with first-time AMI. Focusing on data used for this thesis, characteristics and demographics of participants are presented in Table 1.

Study participants received questionnaires and were invited to a medical examination, during which body height and weight was measured to the nearest 1 cm and 0.1 kg with the participants lightly dressed and without shoes. BMI was calculated from weight (kg) and height (m) using the formula BMI = weight/height^2. Blood pressure was measured in a sitting position after 5 minutes rest, using a validated sphygmomanometer (Omron 711 Automatic IS; Omron Healthcare Inc., Vernon Hills, IL). The pressure was measured twice and the mean of the two measurements was recorded. Blood samples were collected, after ≥4 hours of fasting, for immediate serum lipid (total cholesterol, HDL cholesterol and triglycerides) analysis and storage for DNA extraction.

Table 1. Demographic characteristics of the INTERGENE/ADONIX study participants, subdivided into CHD patients and population controls, by sex.

	CHD cases		Controls	
	Women	**Men**	**Women**	**Men**
Characteristic	**N (%)**	**N (%)**	**N (%)**	**N (%)**
Total	165 (26.7%)	453 (73.3)	1910 (52.9%)	1704 (47.1%)
Age				
≤34 years	1 (0.6%)	1 (0.2%)	247 (12.9%)	198 (11.6%)
35-44 years	2 (1.21%)	16 (3.5%)	415 (21.7%)	351 (20.6)
45-54 years	25 (15.2%)	78 (17.2%)	420 (22.0%)	378 (22.2%)
55-64 years	59 (35.8%)	172 (38.0%)	468 (24.5%)	458 (26.9%)
≥65 years	78 (47.3%)	186 (41.1%)	360 (18.9%)	319 (18.7%)
Hypertension[a]	130 (78.8%)	326 (72.0%)	732 (38.3%)	756 (44.4%)
Diabetes	30 (18.2%)	78 (17.2%)	47 (2.5%)	81 (4.8%)
Ever smokers	100 (60.6%)	352 (77.7%)	953 (49.9%)	895 (52.5%)
BP-lowering treatment	269 (59.4%)	111 (67.3%)	249 (13.0%)	211 (12.4%)
Lipid-lowering treatment	116 (70.3%)	352 (77.7%)	104 (5.5%)	117 (6.9%)
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Age, years	62.7 (8.15)	61.4 (8.44)	51.2 (13.3)	51.6 (12.9)
BMI, kg/m2	28.2 (4.79)	27.7 (3.98)	25.6 (4.35)	26.7 (3.53
LDL cholesterol, mM	2.6 (0.92)	2.5 (0.81)	3.2 (0.98)	3.4 (0.95)
HDL cholesterol, mM	1.6 (0.43)	1.3 (0.34)	1.8 (0.45)	1.5 (0.38)
Total cholesterol, mM	4.9 (1.14)	4.5 (1.01)	5.5 (1.12)	5.5 (1.07)
SBP, mmHg	134 (22.9)	134 (20.3)	128 (22.3)	135 (20.0)
DBP, mmHg	79 (11.0)	83 (11.2)	81 (10.4)	83 (10.4)

CHD: coronary heart disease; BP: blood pressure; BMI: Body Mass Index; LDL: low-density lipoprotein; HDL: high-density lipoprotein; SBP: systolic blood pressure; DBP: diastolic blood pressure

[a] Defined as SBP ≥140 mmHg, DBP ≥90 mmHg or taking anti-hypertensive drugs daily

The questionnaires addressed medical history, socio-economic factors and dietary behavior. For this thesis, mainly medical history and some socio-economic variables were used, along with data collected during the medical examination. One of the assessed socio-economic variables was education. The questionnaire asked participants to mark their highest obtained educational level out of six alternatives: a: elementary school, b: lower secondary school, c: training/girl school, d: upper secondary/grammar school, e: university/college and f: other. These categories were then combined into three educational levels and coded as 1: primary (a, b, c and f), 2: secondary (d) and 3: tertiary (e). For smoking, two variables were constructed from the questionnaire responses. A 2-level never/ever variable, where a person was categorized as a never-smoker if s/he had never smoked and an ever-smoker if s/he indicated that s/he either smoked currently or had stopped smoking. For the 3-level variable, the levels were never/former/current, where never was equal to 'never' in the 2-level variable, 'former' if the individual indicated having stopped smoking at least 12 months previously and 'current' if the individual was currently smoking or had stopped less than 12 months ago.

The study was approved by the local ethical committee and all participants provided written informed consent.

3.2 Air pollution exposure assessment

Modeled annual average levels of NO_2 outside each participant's baseline home address were used for exposure assessment. Each participant's home address was translated into geographical coordinates and combined with modeled levels of NO_2 in a geographical information system (GIS). The dispersion model, which is hosted by the local authorities, contains both emission data and meteorological information and has been previously validated against actual measurements, showing good agreement. (Johansson et al. 2006) The main output from the model is NO_x values with high spatial resolution (20*20 meters), which were then converted to estimated NO_2 using local empirical relationships. Due to the availability of concentration grids, the calculated exposure levels represented the years 2006 and 2007 and not the exact years of inclusion (2001-2004). For individuals with air pollution data for both years, we used the 2007 value because the geographical area covered was increased from 2006. (Figure 2) For individuals with exposure data from only one year, this value was used. Correlation between values for individuals with values from both years was 0.98 for NO_2. This high degree

of stability over years indicates that 2007 is a good indicator also for the long-term spatial distribution of exposure levels during 2001-2004.

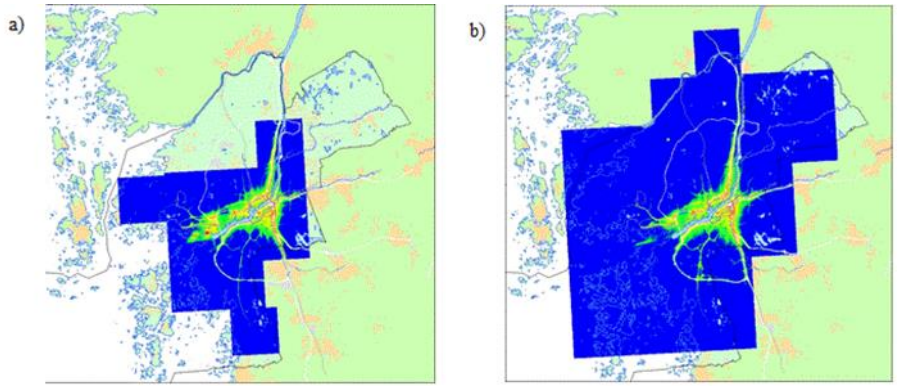

Figure 2. a) Geographical area covered by the dispersion model used to calculate annual average NO^2 exposure in 2006, b) geographical area covered by the dispersion model used to calculate annual average NO^2 exposure in 2007.Figures reproduced from PLoSONE9(6).

Since long-term air pollution exposure assessment of this type essentially is a spatial contrast, a spatially biased recruitment of cases and controls could constitute a problem. Such potential spatial bias by geographical clustering of cases' home addresses in areas closer to the three source hospitals was handled by adjusting the regression model for residential area, based on the postal code for the participants' indicated home addresses. By thus first estimating the effect for each residential area and then pooling the effect (which is the mechanism of adjusting for a variable in a regression model), random selection of both cases and controls from the source population within each area could be more reasonably assumed, although the case-control ratio could vary across residential areas.

3.3 Genetic analysis

3.3.1 SNP genotyping

The three *NOS* genes each code for a certain type of NOS protein. *NOS1* codes for neuronal NOS (nNOS), which among other functions acts as a neurotransmitter in the brain, *NOS2* for inducible NOS (iNOS) which is expressed e.g. in inflammation, and *NOS3* for endothelial NOS (eNOS) which for example is involved in processes regulating blood pressure.

For the three nitric oxide synthase genes, 58 tagging SNPs were selected to capture genetic variation across each gene (Table 2). Tag SNP selection was done using the European ancestry genotype information from the HapMap phase III database (http://www.hapmap.org) with a pairwise approach, SNP minor allele frequency $\geqslant 0.05$ and r^2 between SNPs $\geqslant 0.8$, and including 100 kb upstream and 50 kb downstream of the genes.

Table 2. Descriptive data for all 58 SNPs in the three NOS genes genotyped for INTERGENE/ADONIX.

Gene	dbSNP ID	Location	Alleles (Major/Minor)	Minor allele frequency	HWE* p-value	Call rate (%)
NOS1	rs10774907	Chr12:116131786	G/A	0.28	0.54	98.6
	rs2682826	Chr12:116137221	G/A	0.27	0.44	96.1
	rs816363	Chr12:116144850	C/G	0.40	0.91	98.0
	rs816347	Chr12:116174306	G/A	0.08	0.10	97.4
	rs2293054	Chr12:116186097	G/A	0.28	0.97	97.1
	rs2293055	Chr12:116186267	G/A	0.10	1.00	98.2
	rs6490121	Chr12:116192578	A/G	0.32	0.51	97.2
	rs2293050	Chr12:116203205	C/T	0.41	0.33	98.0
	rs7314935	Chr12:116203220	G/A	0.13	0.79	97.6
	rs9658354	Chr12:116208608	A/T	0.41	0.47	98.7
	rs9658350	Chr12:116208811	A/G	0.19	0.83	92.7
	rs7977109	Chr12:116214723	A/G	0.49	0.09	93.4
	rs532967	Chr12:116216722	G/A	0.18	0.34	98.0
	rs11611788	Chr12:116222759	T/C	0.11	0.23	98.7
	rs7310618	Chr12:116231689	C/G	0.11	0.10	98.0
	rs553715	Chr12:116238239	G/T	0.40	0.06	98.3
	rs2077171	Chr12:116240885	C/T	0.31	0.19	97.1
	rs12578547	Chr12:116247730	T/C	0.25	0.46	95.1
	rs499262	Chr12:116250777	C/T	0.18	0.33	90.9
	rs3782218	Chr12:116255894	C/T	0.16	0.25	92.2
	rs12424669	Chr12:116263339	C/T	0.13	0.65	98.5
	rs1552227	Chr12:116263418	C/T	0.29	0.72	98.4
	rs693534	Chr12:116269101	G/A	0.39	0.13	97.6
	rs1123425	Chr12:116270480	A/G	0.43	0.49	98.0
	rs17509231	Chr12:116278706	C/T	0.14	0.75	97.4
	rs9658253	Chr12:116285009	C/T	0.20	0.05	98.2
	rs41279104	Chr12:117877484	C/T	0.12	0.15	96.9
NOS2	rs4796024	Chr17:23103071	C/T	0.09	0.36	98.0
	rs4795051	Chr17:23103624	C/G	0.43	0.71	98.8
	rs9901734	Chr17:23105156	C/G	0.23	0.90	98.6
	rs2255929	Chr17:23112094	T/A	0.43	0.22	98.2
	rs2297514	Chr17:23117442	T/C	0.39	0.29	97.9
	rs2297515	Chr17:23117460	A/C	0.13	0.66	97.4
	rs2248814	Chr17:23124448	G/A	0.41	0.60	98.0
	rs2314810	Chr17:23128237	G/C	0.05	0.95	98.5
	rs12944039	Chr17:23128891	G/A	0.20	0.36	98.0
	rs4795067	Chr17:23130802	A/G	0.38	0.56	98.2
	rs3729508	Chr17:23133157	C/T	0.40	0.10	98.3
	rs944725	Chr17:23133698	C/T	0.41	0.99	96.4
	rs8072199	Chr17:23140975	C/T	0.49	0.96	96.2
	rs2072324	Chr17:23141023	C/A	0.18	0.29	96.1
	rs3730013	Chr17:23150045	G/A	0.31	0.73	98.0
	rs10459953	Chr17:23151645	G/C	0.36	0.16	97.8
	rs2779248	Chr17:23151959	T/C	0.38	0.56	97.7

	rs2301369	Chr17:23154123	C/G	0.38	0.54	96.5
NOS2A	rs2779252	Chr17:23155497	G/T	0.05	0.77	98.5
NOS3	rs10277237	Chr7:150314277	G/A	0.21	0.07	98.0
	rs1800779	Chr7:150320876	A/G	0.35	0.33	97.6
	rs2070744	Chr7:150321012	T/C	0.35	0.45	98.2
	rs3918226	Chr7:150321109	C/T	0.08	0.25	98.4
	rs3918169	Chr7:150325539	A/G	0.16	0.93	97.3
	rs3793342	Chr7:150326128	G/A	0.15	0.60	98.0
	rs1549758	Chr7:150326659	C/T	0.29	0.28	98.2
	rs1799983	Chr7:150327044	G/T	0.30	0.27	98.2
	rs3918227	Chr7:150331879	C/A	0.10	0.43	98.0
	rs3918188	Chr7:150333714	C/A	0.37	0.60	97.4
	rs1808593	Chr7:150339235	T/G	0.20	0.96	96.1
	rs7830	Chr7:150340504	G/T	0.38	0.46	98.0

* HWE: Hardy-Weinberg equilibrium

GST genes code for metabolizing enzymes which, for example, are involved in counteracting the effects of oxidative stress. [Raza 2011] In total, 9 SNPs were chosen based on literature findings; 7 in the *GSTP1* gene, one to capture the null variant of *GSTT1* and one in *GSTCD*. (Table 3)

Table 3. Descriptive data for the GST-SNPs genotyped for INTERGENE/ADONIX.

Gene	dbSNP ID	Location	Alleles (Major/Minor)	Minor allele frequency	HWE* p-value	Call rate (%)
GSTP1	rs1138272	Chr11:67110155	C/T	0.08	0.70	98.5
GSTP1	rs1695	Chr11:67109265	A/G	0.33	0.28	98.0
GSTP1	rs1871042	Chr11:67110420	C/T	0.34	0.26	97.8
GSTP1	rs596603	Chr11:67116179	G/T	0.43	0.23	98.2
GSTP1	rs749174	Chr11:67109829	G/A	0.34	0.26	98.1
GSTP1	rs762803	Chr11:67108832	C/A	0.43	0.45	97.5
GSTP1	rs7927381	Chr11:67103319	C/T	0.09	0.72	97.1
GSTCD	rs10516526	Chr4:106908353	A/G	0.06	0.005	98.5
GSTT1	rs2266637	Chr22: 22706845	Non-null/null genotype	Frequency of null genotype: 0.150	-	94.1

* HWE: Hardy-Weinberg equilibrium

SNPs were genotyped using a Sequenom MassARRAY platform (Sequenom San Diego, CA, USA) or a competitive allele-specific PCR system KASPar (KBioscience, Hoddesdon Herts, GB). All SNPs had a call rate $\geqslant$90%

(Tables 2 & 3). SNPs with a Hardy–Weinberg Equilibrium (HWE) p-value ⩽0.001 and individuals with a genotype success rate below 75% were excluded.

3.3.2 Genetic models and genotype coding

Consider a genetic single nucleotide locus where there is genetic variability and whose nucleotide is either C or A on one strand of the chromosome (here considered to be the reference strand). Since we have two of each chromosome, the possible combinations are CC (C on this strand on both chromosomes), CA (C on this strand on one chromosome and A on this strand on the other chromosome) and AA (A on this strand on both chromosomes). The nucleotide with the lowest frequency in the population at hand is called the minor allele and the other consequently the major allele. Usually, the major allele is set as reference and thus the minor allele is called the 'risk' allele even though it may have a protective effect for the studied outcome. The minor allele frequency may vary between populations due to selection, and especially for small populations also due to genetic drift. [Rosenberg et al. 2002] Sometimes the minor allele in one population may even be the major allele in another population.

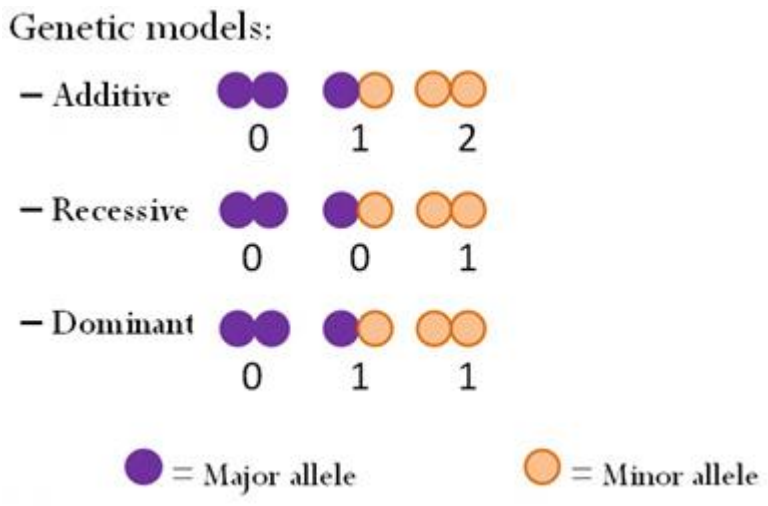

Figure 3. Coding for statistical analysis of the three genetic models: additive, recessive and dominant. Figure by Anna Levinsson

Assume that C is the risk (minor) allele. Under a dominant genetic model, disease risk increases if a person has at least one C allele. Thus we code CC and CA =1 while AA =0. (Figure 3) Under a recessive genetic model, disease risk increases only if two risk alleles are present, coded CC =1 and CA, AA =0. Under an additive genetic model, disease risk increases for each copy of the C-allele present, and is coded so that AA =0, CA =1 and CC =2. [Jorde et al. 2005] Also note that the dominant model for the minor allele is the same as the recessive model for the major allele and vice versa.

In Paper I, all 3 genetic models were used with the intention of identifying the best-fitting genetic model for each SNP. For Paper II and the applied example in Paper III, we used the dominant genetic model only, to improve statistical power and the stability of the regression models. It is notable that the dominant genetic model often detects the same associations as the additive model, with relatively similar power, given that the only difference in coding of the genotype is that 'homozygous for the risk allele' =2 for the additive model and =1 for the dominant model. [Lettre et al. 2007] The similarity in power is due to the fact that the number of individuals coded 2 in the additive model often is small.

Individuals of non-European birth (5%) were excluded from all analyses. Of those reporting European birth origin and included, 90% reported being of Swedish origin.

3.3.3 Statistical methods in genetic data analysis

Paper I

In Paper 1, a stepwise method was used to identify the SNPs most strongly identified with the outcomes CHD and hypertension. For each outcome, the following procedure was carried out.

First, all SNPs were coded according to the additive genetic model, which has the greatest power of the three models to detect an association in many settings, and analyzed in single-SNP logistic regression models, adjusted for age and sex. The SNPs that had a p-value of 0.2 or less were taken to the next step, where a stepwise selection was made using an entry p-value of 0.1 and a limit p-value=0.2 for staying in the model. The SNPs remaining in the model were advanced to the third step. Given that the additive genetic model is not always the best or true fit for a genotype and in order to allow SNPs with the recessive or dominant genetic model as the best fit (which may not have

been captured by the additive genetic coding) to qualify for the final (third step) model, each SNP was also coded to these two genetic models and entered in single-SNP logistic models adjusted for age and sex. The SNPs with a p-value =0.05 or less in these models were also taken to the last step of the procedure. The p-value was set lower at 0.05 since no intermediate selection step was used. Finally, to identify the most strongly associated SNPs and their best-fit genetic models, all qualified SNPs (being selected by one or more of these steps) were coded to all three genetic models and entered into a stepwise logistic model, adjusted for age and sex and potentially containing several SNPs, with entry p-value = 0.1 and stay p-value = 0.05. (Figure 4) A SNP was only allowed to remain in the model coded to one genetic model.

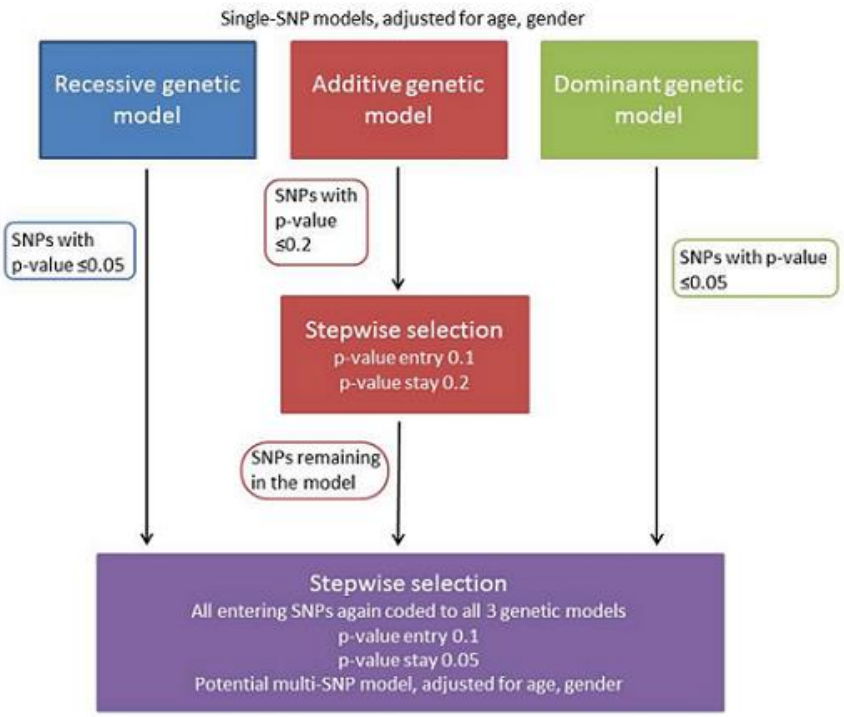

Figure 4. Flow chart describing the steps in the statistical analysis for identifying the SNPs most strongly associated with each CV phenotype. Printed in Nitric Oxide 39 (2014) 1-7.

A possible result of the stepwise analyses was that several SNPs in the same gene (i.e. on the same chromosome) were selected. SNPs on the same chromosome can be analyzed using haplotype analysis to indicate if the SNPs are markers for the same observed effect. This was carried out using the *haplologit* command in STATA. In short, this command first estimates the initial haplotype frequencies. [Marchenko et al. 2008] Then, haplotype-effects logistic regression is used to estimate coefficients for risk haplotypes, environmental covariates (if included) and their interactions simultaneously with the final haplotype frequencies.

Paper II

The same core procedure was used for both AMI and hypertension, but for AMI the full dataset was used (cases and control cohort) while for hypertension only the individuals in the population control cohort were included, divided into hypertension cases and non-hypertension controls. All analyses used logistic regression models adjusted for age, age squared (included due to indicated non-linearity in the age variable) and sex. Because of potential selection bias for cases and controls due to the spatial distribution of cases' home addresses in areas closer to the two source hospitals, meaning that the probability of seeking care in a participating hospital, and thus the possibility of becoming a case, was not the same in all residential areas, all analyses were adjusted for residential area, based on the postal code. In addition, this controls for the control sampling fraction potentially varying across areas, due to non-participation. All analyses involving air pollution exposure were also adjusted for educational level as a proxy for socio-economic and lifestyle variable. Since pre-analyses indicated a confounding of genotype effect by BMI, BMI was also included in all genotype analyses. Covariates as potential confounders were selected from literature and tested one at a time. The covariates whose entry into the model changed the effect estimate for genotype or air pollution by at least 5% compared to the effect estimate for respective exposure and outcome in models with no other variables, were considered confounders and included in main analysis models (main effects and interaction).

First, effects of NO_2 (as a marker for vehicle exhaust pollutants) on risk of AMI and hypertension were analyzed separately. Thereafter, effects of genetic variants on risk of AMI and hypertension were studied. For *GSTP1*, each of the 7 SNPs was analyzed coded to the dominant genetic model (0 for two copies of the major allele, 1 for heterogeneous genotype or two copies of the minor allele). For this gene, only the SNP or SNPs with the strongest

effects on AMI or hypertension were studied for interaction with air pollution exposure on risk of each respective outcome. For *GSTCD*, a single variant (rs10516526) was studied, coded to the dominant genetic model. The *GSTT1* null genotype was studied using the two genotypes captured by the SNP rs2266637.

Finally, interaction between air pollution and genetic variants was investigated by estimating effects of air pollution in analyses stratified by genotype in a common regression model, one SNP at a time. The p-value of the product-term of SNP and air pollution was considered an indicator of the presence of multiplicative interaction between the two exposures (the null hypothesis being no interaction). For these models, the possibility of smoking modifying the interaction between air pollution and genetic variants on AMI was also assessed, by stratifying the analyses of the effect of air pollution on risk of respective outcome by both genotype and 3-level smoking status.

3.4 Estimating RERI for GST and air pollution data from Paper II, using methodology from Paper III

In paper II, interaction between genetic variants and air pollution exposure on risk of AMI and hypertension was investigated using stratified effect methodology. In paper III, an approach for dealing with additive interaction between one dichotomous (e.g. dominant or recessive genetic variable) and one continuous (for example, ambient air pollution measured with NO_2 as a marker) variable was presented, and this approach was subsequently applied to the data in paper II and presented here.

Using the outcome AMI, several dichotomous genetic variables and the continuous air pollution exposure, RERI was estimated using the method from paper III.

Let

X: dichotomous genetic variable {0,1}

Y: continuous air pollution exposure variable with unit $10\mu g/m^3$

d_x: increment in $X = (x_o \rightarrow x_1)$ where x_o represents baseline and x_1 "elevated" exposure level of interest

d_y: increment in Y = ($y_0 \rightarrow y_1$) where y_0 represents baseline and y_1 "elevated" exposure level of interest

$\beta_{X|Y}$: regression coefficient estimate for X when continuous variable defined as Y

β_Y: regression coefficient estimate for Y

β_{XY}: regression coefficient estimate for interaction factor XY

and

$$Z: \frac{Y - mean(Y) - \frac{max(Y) - min(Y)}{2 * 1000}}{\frac{max(Y) - min(Y)}{1000}}$$

Then

$$RERI = e^{\beta_{X|Z} + \beta_Z + \beta_{XZ}} - e^{\beta_{X|Z}} - e^{\beta_Z} + 1 \qquad \text{[Paper III]}$$

and what remains is to calculate the mean and the range of the air pollution exposure variable and to estimate the regression coefficients using logistic regression. Confidence intervals for RERI were calculated using the Wald-type method.

4 RESULTS

4.1 Paper I

Several SNPs where found to be associated with CHD, of which one previously unpublished SNP, *NOS1*: rs3782218 coded according to the additive genetic model, was strongly associated with both CHD (odds ratio (OR) 0.6, 95% confidence interval (CI) 0.44-0.80) and hypertension (OR 0.8 95% CI 0.68-0.97). The statistical significance of these results held even after Bonferroni correction for multiple testing. Several other SNPs in the *NOS2* and *NOS3* gene were associated with an adverse effect for either CHD or hypertension, with ORs ranging from 1.2 – 2.2. (Paper I: Tables 2 & 3)

For each outcome, another significant *NOS1* SNP association in addition to rs3782218 was found. A haplotype analysis for respective outcome including rs3782218 and the other respective *NOS1* SNP indicated that for CHD, the rs3782218 T-allele may be the main marker for the observed effect, while for hypertension it seems that both *NOS1* SNPs investigated may be markers for the same observed effect. (Tables 4 & 5)

Table 4. Haplotype analysis of NOS1 SNPs associated with CHD. SNP order in the haplotype from left to right is rs2682826 and rs3782218. Printed in Nitric Oxide (2014) 39:1-7.

		Modelling of all haplotypes[a]			**Modelling haplotype GT against all others[a]**		
Haplotype[b]	**Sample frequency**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
GT	0.11	0.50	0.34 - 0.71	$1.50*10^{-4}$	0.43	0.30 - 0.61	$1.85*10^{-6}$
AT	0.05	0.61	0.35 - 1.05	0.07	Reference		
GC	0.61	Reference					
AC	0.23	1.14	0.96 - 1.35	0.15			

[a] Model is adjusted for gender, age, diabetes status, smoking, systolic blood pressure, high- and low-density lipoprotein and the other haplotypes

[b] For rs2682826: A is the minor allele, for rs3782218: T is the minor allele

CHD: coronary heart disease, OR: odds ratio, CI: confidence interval

Table 5. Haplotype analysis of NOS1 SNPs associated with hypertension. SNP order in the haplotype from left to right is rs7314935 and rs3782218. Printed in Nitric Oxide (2014) 39: 1-7.

		Modelling of all haplotypes[b]			Modelling haplotype GT against all others[b]		
Haplotype[a]	Sample frequency	OR	95% CI	p-value	OR	95% CI	p-value
GT	0.17	0.86	0.73 - 1.01	0.06	0.84	0.72 - 0.98	0.03
AT	0.0016	1.45	0.11 - 18.57	0.77	Reference		
GC	0.70	Reference					
AC	0.13	1.14	0.96 - 1.34	0.13			

[a] For rs7314935: A is the minor allele, for rs3782218: T is the minor allele

[b] Model is adjusted for gender, age, diabetes status, body mass index, total cholesterol and the other haplotypes

OR: odds ratio, CI: confidence interval

4.2 Paper II

In Paper II, the main effect of long-term estimated NO_2 exposure at the residential address (as a marker of long-term air pollution exposure) on risk of AMI was estimated to be OR 1.8, 95% CI 1.04-3.03. Three *GSTP1* SNPs were associated with hypertension, even after Bonferroni correction. (Table 6)

The interaction analyses indicated that the effect of air pollution exposure on risk of AMI varies between genotypes for all 3 SNPs (one in *GSTP1*, one in *GSTT1* and one in *GSTCD*) tested (Table 7), with a significant effect seen in one genetic stratum, although the interactions were not statistically significant due to the limited sample size.

Table 6. Effects of the most strongly associated SNPs in GSTP1, the GSTCD SNP rs10516526 and the GSTT1 SNP rs2266637 (null genotype) on risk of AMI and hypertension. Adapted from PLoSOne (2014) 9(6):e99043.

			Effect estimates and precision*		
Gene: SNP	**Outcome**	**Genetic model**	**OR**	**95% CI**	**p-value**
GSTP1: rs596603	AMI	(TT + GT) vs. GG	0.77	0.51-1.16	0.21
GSTP1: rs1871042	Hypert.	(TT + TC) vs. CC	0.66	0.50-0.87	0.003
GSTP1: rs749174	Hypert.	(AA + AG) vs. GG	0.66	0.50-0.88	0.004
GSTP1: rs762803	Hypert.	(AA + CA) vs. CC	0.66	0.49-0.89	0.006
GSTCD: rs10516526	AMI	(GG + AG) vs. AA	0.69	0.34-1.38	0.29
	Hypert.	(GG + AG) vs. AA	0.87	0.55-1.36	0.53
GSTT1: rs2266637	AMI	Null vs. Non-null	0.65	0.33-1.27	0.20
	Hypert.	Null vs. Non-null	0.88	0.59-1.33	0.55

Models are adjusted for age, age squared, sex and BMI.

Table 7. Effect of long-term traffic-related air pollution (using annual mean of NO_2 as exposure indicator) on risk of AMI, stratified by genotype. Adapted from PLoSOne (2014) 9(6):e99043.

		Effects per 10µg/m^3 of NO_2			
Gene:SNP	**Genotype**	**OR**	**95% CI**	**p-value**	**Interaction p-value**
GSTP1: rs596603	TT + GT	2.12	1.09 – 4.10	0.03	0.27
	GG	1.40	0.73 – 2.68	0.31	
GSTCD: rs10516526	GG + AG	1.01	0.28 – 3.73	0.98	0.23
	AA	2.25	1.25 – 4.06	0.0007	
GSTT1: rs2266637	Null	1.40	0.33 – 5.96	0.65	0.60
	Non-null	2.02	1.13 – 3.60	0.02	

Models are adjusted for age, age squared, sex, BMI, residential area and educational level.

4.3 Paper III

The paper starts out from the standpoint that if main effects are estimated using logistic regression, then interaction effects, both multiplicative and additive, should also be estimated using the logistic model, for the purpose of interpretation and understanding how they relate to the chosen main effects model. It is pointed out that when the additive interaction of interest is between two dichotomous variables, methods that align with the original definitions of departure from additivity of risks have been defined specifically for this case and work well. For the less investigated but at least as commonly occurring situation of one dichotomous and one continuous variable, the paper proposes a pragmatic approach for estimating the additive interaction, based on the notion that RERI ought to be estimated in an interval which well represents the full set of variable data. This boils down to estimating RERI in a very small interval, approaching zero in length, which surrounds a suitable measure of location, for example the mean. As already mentioned in section 3.4, the proposal involves a simplification by transforming the continuous variable to a variable with the minimum 0 at the mean of the original variable minus half the interval to be used and divided by the range of the original variable/1000 i.e.

Z: $\frac{Y - mean(Y) - \frac{max(Y) - min(Y)}{2*1000}}{\frac{max(Y) - min(Y)}{1000}}$, which can then be used in the dichotomous variable logistic regression output adaption of the original RERI equation presented by Rothman (see section 1.4.1, Equation [2]).

However, this estimate of RERI is dependent on the interval for which it was estimated (which defines the unit of the continuous exposure), something that can be adjusted by standardizing RERI by division with the interval used, i.e. (max(Y) – min(Y)) / 1000 and then adapting the estimate to the scale of the exposure main effect with simple multiplication, based on the fundamental additive property of RERI.

4.4 Estimating RERI for GST and air pollution data from Paper II, using methodology from Paper III

For this example, 120 AMI cases and 1483 randomly selected controls had exposure data. The baseline and increment for the continuous variable were calculated from the data:

$$d_y = \frac{range(Y)}{1000} = \frac{4.359}{1000} = 0.004359$$

$$y_0 = mean(Y) - \frac{dy}{2} = 1.5572 - \frac{0.004359}{2} = 1.5550$$

All regression models were adjusted for age, age squared, sex, BMI, level of education and residential area.

Table 8. Estimated RERI for interaction between respective SNP and long-term traffic-related air pollution exposure (per 10 µg/m^3) for outcome AMI

Gene: SNP	RERI	95 % CI RERI	$\beta_{X\|Y}$ [a]	β_Y [b]	β_{XY} [c]	OR $_{X\|Y}$	OR $_{Y}$	OR $_{XY}$	p-value^
GSTP1: rs596603[d]	-0.31	-1.21 – 0.58	0.91	0.75	-0.41	2.48	2.12	0.66	0.27
GSTCD: rs10516526[e]	-0.80	-1.58 – -0.03	0.57	0.81	-0.80	1.77	2.25	0.45	0.23
GSTT1: rs2266637[f ~]	-0.51	-1.47 – 0.46	0.08	0.70	-0.38	1.08	2.01	0.68	0.60

* The unit of the measure is per 10µg/m^3. ^ P-value for the product term of the two exposures, i.e. the test of no multiplicative interaction. ~ In the stratified effects regression model used for GSTT1 in Paper II, the educational level variable was mistakenly used as a continuous variable instead of as a categorical. Estimates were only marginally changed. The results for GSTT1 corresponding to Table 4 in Paper II from the correct model are presented as an erratum in Appendix Table 2.

[a] $\beta_{X|Y}$: regression coefficient estimate for genetic variable X when continuous air pollution exposure variable defined as Y, [b] β_Y: regression coefficient estimate for continuous air pollution exposure variable Y, [c] β_{XY}: regression coefficient estimate for interaction factor XY, [d] dominant genetic model, the coding of the genetic variable was reversed (compared to Paper II) in order to obtain positive main effect estimates for calculation of RERI and evaluation of deviation from additivity of risks, as recommended by Knol et al. (2011), [e] dominant genetic model, [f] null genotype as risk genotype.

For all three genotypes, the RERI estimate as well as the multiplicative interaction estimate is negative. The 95% confidence intervals show that for the GSTCD combined AG+GG genotype, there is a significant deviation from additivity, in this case inferring that the sum of the genotype and the exposure effects is less than expected, i.e. sub-additivity. For the other two SNPs, the no additive interaction null hypothesis cannot be rejected and the p-values for the product term beta coefficients imply that we cannot reject the null hypothesis of no multiplicative interaction.

5 DISCUSSION

Results from gene-environment interaction studies can provide clues to many currently unanswered questions. By inference from the function of genes associated with disease, it is possible to gain better insight into specific pathways of disease pathology. [Tabor, Risch and Myers 2002] This may lead to the development of new medicines and therapeutic approaches. From a public health perspective, individuals at particularly high risk, i.e. those with high-risk genotype and high-risk environmental exposure, may be identified and potentially given targeted prevention, advice and health care. [Ottman 1995] Highly exposed groups could be identified, and such individuals could potentially be invited to a genotype test to determine increased risk due to genetic susceptibility, which would indicate an incentive for healthy lifestyle management. [Khoury 1997, Khoury & Wagener 1995, Ottman 1995]

In the last few years, genotyping services for individuals have arrived in Britain and the US, where the debate over ethics and validity has been fierce. [Cooper 2014, Annas & Elias 2014a] So far, neither the validity of the tool, nor the health effects and impact (on an individual level or a population level) of the information provided by the service, are known. The American Food and Drug Administration has recognized potential risks and has sent a warning letter to one company marketing such services [Annas & Elias 2014a, FDA 2013], and calls have been made for an international harmonization of standards regarding personal genotyping and handling of resulting data. [Annas & Elias 2014b, Yuji, Tanimoto, Oshima 2014, Annas & Elias 2014a] A prospective study of personal genomic testing has been launched in the US, by researchers in collaboration with two genomic profiling service companies. [Carere et al. 2014)

On a more theoretical note, interaction analysis is complex and both theory and practice have inconsistencies. [Rothman, Greenland, Lash 2008] In particular, the concept of interaction as departure from additivity or multiplicativity of risks is still broadly discussed, and neither concepts nor methods or interpretation are well characterized or agreed upon. [VanderWeele 2011, Kaufman 2009, Ahlbom & Alfredsson 2005, Skrondal 2003] While multiplicative interaction can easily be estimated directly in a logistic regression model, the concept of additive interaction answers somewhat different questions relevant to public health. [Rothman, Greenland,

Lash 2008] Regardless of the pros and cons, characteristics and relative ease of estimation for such different measures of interaction, consensus is needed to allow comparison and replication of results across studies. In order to better understand and ultimately determine how factors interact and the effect on disease risk, clearly defined methods are needed.

Paper II used modeling of stratum-specific risk to investigate whether the effect of air pollution differs between genotype strata (a concept known as effect measure modification, presented briefly in section 1.5), thereby studying whether the risk of disease due to the environmental exposure as measured in the statistical model is *modified* by the genotype. Including the product term of the genotype variable and the air pollution exposure variable (i.e. the two exposure variables) in the logistic regression model estimating relative risks yields a p-value for the significance of the multiplicative interaction between risk factors, and the stratified effect estimates were easily obtainable by a reparametrization of this model.

In the methodological Paper III, additive interaction was the focus and in simulations, a range of RERI estimates based on simulated parameter values for baseline and increment showed that the estimated RERI varies significantly even within the same dataset (the regression coefficients used in RERI calculations are estimated from the dataset and are thus fixed for a given model and dataset). Hence, some consensus on how to select values for the other parameters needed for RERI calculation must be reached in order for estimates to be interpretable and better comparable across studies. Other currently available measures of additive interaction (AP, synergy index) have similar issues and do not provide unequivocal estimates either. [Knol et al. 2007]

In epidemiological studies, and when using either multiplicative or additive interaction, there is always a risk of bias from various sources that needs to be considered. The INTERGENE/ADONIX study is a well characterized, population-based study with high-quality genotyping data. Potential selection and non-participation bias in the population control sample part of INTERGENE/ADONIX study has been investigated. [Strandhagen et al. 2010] The participation rate was 41.9%, and it was concluded that participants were somewhat more likely to be women, be well educated, be married and have a high income as well as being of Nordic origin compared to non-participants. Neither of these findings is likely to have any significant adverse impact on the case-control analyses in this thesis, especially not if the

same selection patterns can be assumed for the cases, which is reasonable. The fact that more women attended is in fact an advantage, insofar as that past cardiovascular research has often been conducted largely on men. Cardiovascular events occur in women too, and women need to be included in both study populations and clinical trials to ensure that research findings are relevant to women and that drugs approved from clinical trials are safe for women as well. [Kim et al. 2008, Stramba-Badiale 2009] That many of the participants are of Nordic origin is also advantageous, as we require a population with the similar overall ancestry for assessment of specific genetic risks in the genetic analyses. [Rosenberg et al. 2002] The small among-populations genetic variance in Europe makes it reasonable to extend inclusion criteria to (self-reported) European ancestry. Being married and well educated does not affect the genetic susceptibility. However, educational level in participants was associated with several CVD risk factors (lower risk factor prevalence for more well educated) including hypertension, cholesterol levels and smoking, and may be associated with air pollution exposure, so that adjusting for educational level in analyses is likely advantageous. [Strandhagen et al. 2010]

A relatively high prevalence of hypertension was observed, using a well-established definition for hypertension (SBP $\geq$140, DBP $\geq$90 or taking anti-hypertensive medication daily): 45.9% in the total study population, 73.8% of CHD cases and 41.2% of population controls. However, this is believed to be a demographic characteristic rather than an indication of biased measurements. [Chow et al. 2013] The selection of CHD case individuals was made with an emphasis on specificity, and all diagnoses have been validated; thus addressing potential misclassification of diagnosis among included cases. Because of the case recruitment strategy being through the major hospital, some true cases are likely not to have been given the possibility of participation in the study, but potential lack of sensitivity in case selection generally does not cause bias, as long as the case subset is non-differential regarding exposure. It is possible that cases missing due to lack of sensitivity are differential with respect to air pollution exposure, since cases may be more likely to participate if they reside close to recruiting hospitals. Therefore, potential spatial bias was addressed by adjusting regression models for residential area, as discussed in section 3.2.

Unfortunately, due to limitations of the area covered by the dispersion model, estimates of long-term air pollution exposure (Paper II and III) were not available for the entire geographical study area, thus excluding many case

and control individuals with otherwise valid data from the present analysis. Modeling air pollution levels for a larger geographical area would enable inclusion of more of the participants and increased power, but such models remain to be developed. The limitations of the dispersion model is the reason why the number of first-time AMI cases used in Paper II and III dropped from 192 potential AMI cases to only 119 once the exposure estimation was finished.

The two main exposures used in this thesis are genetic variants and long-term traffic-related air pollution. Genetic variables are generally very well measured and thus suffer from low misclassification rates. Potentially mismeasured genotypes are also eliminated by excluding SNPs that deviate radically from Hardy-Weinberg equilibrium, which can indicate potential genotyping error. The main point where misclassification may be introduced is when the genetic model for analysis is chosen, which prompted the use of the stepwise procedure in Paper I. Regarding assessment of air pollution exposure, the stated home address of each individual was used as the geographical reference point. Obviously individuals do not spend 24 hours a day every day right outside their house, but rather move between home, work and any spare time activities. Also, the air indoors may be more or less polluted than the estimated outdoor air levels. However, the assumption is that for the type of air pollution exposure assessment used in this thesis, any exposure measurement error is likely to be non-differential with respect to outcome, meaning that it is on average equal among participating cases and non-cases. This assumption in conjunction with the results from the validity analysis of the estimated annual mean air pollution exposures (Johansson et al. 2006) led us to the decision not to correct the air pollution exposure variable for measurement error. The effect of non-differential measurement error for an exposure variable in an individual study has mostly been studied for the case of a dichotomous exposure and a dichotomous outcome, where the consequence of measurement error for the exposure variable on average is an attenuation of the association estimate. [Birkett 1992] For multiple categories and continuous exposures, the effect of the measurement error on the association estimate is more complex, which is also true for multivariate analyses that include covariates. [Brenner & Loomis 1994, Birkett 1992]

In Papers I and II, confounding has been dealt with by including confounders as covariates in the regression models. Assessment was carried out starting with a list of possible confounders from literature. Each potential confounder was entered into a logistic regression model for respective outcome along

with the exposure variable. If the inclusion of the potential confounder changed the effect estimate for the exposure of interest by 5% or more, the potential confounder was considered a confounder in the analysis at hand and included in the final regression model. Given the limited sample size, known risk factors were not included in analysis models unless they fulfilled the 5% criterion, in order to keep the models as parsimonious as possible. Hence, non-inclusion does not necessarily imply that variables were not risk factors in our population.

In the analysis results from Paper II, none of the regression models are adjusted for smoking, which may seem counterintuitive since the investigated pathological mechanism is systemic inflammation from pulmonary exposure. However, smoking was studied in depth as a potential confounder according to the 5% criterion above and was not considered a confounder. It was also studied in stratified analyses (effect modification analyses), where it did not show any conclusive effect modification, neither as a 3-level never/former/current smoker variable nor as a 2-level never/ever smoking variable. The effect modification results are presented in Appendix Table 1.

Despite careful evaluation of potential confounding, in general complete control of confounding is not possible due to lack of data or insufficient detail and some residual confounding may remain, although often this remainder is unlikely to be substantial. A potential example of possible residual confounding is the result from Paper II, which shows a non-significant effect of air pollution exposure in the "beneficial" direction with respect to hypertension, which may be considered counter-intuitive. One possible explanation for this finding is residual confounding by lifestyle and socioeconomic factors. To exemplify, people living in the center of Gothenburg, i.e. areas with higher traffic intensity and potentially higher air pollution exposure, also tend to have a higher socio-economic status, i.e. higher income and education, smoke less and have more regular health care contacts than individuals living in more rural areas. Similar patterns have been identified in other cities. [Forastiere et al. 2007, Zeka et al. 2006, Hoek et al. 2002] Thus, despite attempts to adjust for such factors in Paper II using an educational level variable, and indirectly also by adjustment for residential areas, and possibly also BMI, as well as assessing smoking as a confounder, what seems to be a potentially slight beneficial effect of air pollution may still represent a confounding effect of other risk factors.

Paper I used the collective diagnosis 'CHD' as an outcome along with hypertension, while in Paper II AMI was used instead of CHD. The reason why a more precise diagnosis was used for the air pollution analyses is that by using AMI cases, a definite date of event could be established and thus temporality of exposure (that assessed exposure occurred before time of event) could be ascertained. In addition, there was a suspicion that different susceptibilities and mechanisms may be active in the different cardiovascular outcomes contained in the umbrella term CHD, for example acute myocardial infarction, angina and coronary artery disease, regarding associations with air pollution exposure. Thus, to investigate the associations of a specific diagnosis with a clear onset, the AMI cases were selected among the CHD cases. Hypertension may be regarded as a risk factor for AMI in many cases, as an earlier step in the progression of CVD. One reason why hypertension was used as an outcome for Paper II was to evaluate whether air pollution exposure has a similar effect for different aspects of the CVD development.

In Paper III, the objective was to estimate a RERI for one dichotomous and one continuous variable that is both consistent with the original definition of additivity of risks and takes the characteristics of the data used into account. To obtain such a RERI, several assumptions were made. First, the assumption that the logistic regression model gives unconfounded effect estimates. Second, that the interval for each variable in which RERI is estimated, should be representative of the available data for the population regarding the exposure. For a dichotomous variable, this is straightforward, as there is only one baseline and one increment (exposed level) to choose from, and both main effects and interaction are naturally estimated for that one available contrast. For a continuous variable, there is generally no obvious baseline or increment, and in order to calculate RERI, these and the interval they define must be chosen. Paper III argues that a good choice for the sought contrast would be the effect close to a measure of location of the available data. Here, the center of the interval where RERI is estimated must be chosen carefully (focusing on a measure of location for the variable data, e.g. the population mean), while the increment ought to be small, approaching zero (in order to access the local regression slope which represents the effect), to best estimate the sought odds ratio. Due to the additive properties of RERI as conceptually defined [Rothman 1986], once the sought RERI is obtained for the small interval, it can be scaled back in an additive manner to the same unit that was used for the main effect estimate for the continuous variable.

Which measure of location (or measure of central tendency) is the most appropriate to center the RERI interval depends on the characteristics of the data. Often measures of location are compared in terms of robustness and resistance, two related properties which are founded on slightly different conditions [Andrews 1998], as well as power, i.e. the probability of detecting true associations between variables. To exemplify, if outliers are present, it will affect the mean in the direction of the outliers, meaning that the mean is not resistant. The mean is also sensitive to even small deviations from normal distribution of the data, which can inflate the standard error of the mean, which in turn reduces the power. [Wilcox 2014] The median, on the other hand, is resistant, because only the one or two most central ordered values are actually used to determine the median. However, precisely this characteristic reduces the power of the median. Regarding robustness, the median is more robust than the mean, because the shape of the tails of the variable distribution has a larger effect on the mean than on the median. Thus, neither the median nor the mean is an optimal measure of location in the sense that neither is unaffected by outliers or deviations from normal distribution. In the applications in this thesis, the mean is used as measure of location because the air pollution exposure variable meets the assumptions of no extreme outliers and approximately normal distribution, and trimming the data seemed wasteful when not explicitly called for. When the assumptions are not met, a trimmed mean with γ=0.2 has been recommended as a good compromise between power, resistance and robustness. [Wilcox 2014, Wilcox 1998] We consider it likely that this measure will also be useful in conjunction with our proposed pragmatic approach to obtain a RERI estimate for such data.

A γ-trimmed mean is

$$\widehat{X_t} = \frac{1}{n-2g}\left(X_{(g+1)} + \cdots + X_{(n-2g)}\right)$$

where $0 \leq \gamma \leq 0.5$; $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ are the observations written in ascending order and $g = [\gamma n]$ where $[\gamma n]$ is the value of γn rounded down to the nearest integer. [Wilcoxon 2014]

6 CONCLUSION

Overall, 58 SNPs in the three *NOS* genes, 7 SNPs in *GSTP1* and one in each of *GSTCD* and *GSTT1* were investigated regarding association with CHD, AMI or hypertension. Several of the *GST* SNPs were further analyzed to evaluate interaction between air pollution exposure and the genetic variants on the outcomes AMI and hypertension. An extension to a known method for estimating additive interaction was proposed, to support the use of continuous variables.

6.1 Paper-specific conclusions

I. Several SNPs in the *NOS1*, *NOS2* and *NOS3* genes were found to be significantly associated with either CHD or hypertension, including the *NOS1* SNP rs3782218, which was significantly associated with both outcomes. A haplotype analysis indicated that for CHD, the T-allele of this SNP is a main marker of the observed effect, whereas for hypertension another *NOS1* SNP seemed also to contribute to the effect. The results provide additional support for the biological rationale of the nitric oxide pathway in CVD. The *NOS1* findings are novel, although the gene has not been studied much previously in relation to CVD.

II. A significant increase in risk of AMI was found in association with long-term traffic-related air pollution exposure, which is consistent with previous findings regarding an association between long-term air pollution and AMI. For hypertension, no conclusions about a potential association with air pollution exposure could be drawn. On the other hand, variants in *GSTP1*, *GSTT1* and *GSTCD* showed no clear associations with AMI, but several SNPs were associated with hypertension. When the effect of long-term traffic-related air pollution was analyzed for AMI and hypertension in models stratified by genotypes of the most strongly associated SNPs, multiplicative interactions were not statistically significant, but results indicated that the effect of long-term air pollution exposure on the risk of AMI may vary by genotype, while no obvious effect modification was seen for hypertension. Although the interaction results were not statistically significant, the results are consistent with potential genetic susceptibility for air pollution exposure effects

on risk of AMI related to variants in antioxidant genes of the GST family, which may not involve hypertension.

III. In the methodological investigation of additive interaction, it was concluded that RERI varies with every parameter involved in its calculation. Of these, the beta coefficients (representing main effect and multiplicative interaction odds ratios) are determined by the data through logistic regression, while baseline and increment values must be selected by the investigator. A further conclusion was the pragmatic proposal that RERI ought to be estimated around a measure of location which is representative for the variable data, and with an increment approaching 0, in order to best estimate an unequivocal, interpretable measurement of additive interaction. Finally, a standardization using the inverse of the increment enables RERI to be scaled back in an additive fashion consistent with its original definition to the unit used for the main effects, facilitating interpretation.

7 FUTURE PERSPECTIVES

Due to the effort and time required to acquire the air pollution exposure data for this study material, some anticipated studies within the scope of this thesis remain to be done. One such study is that of investigating the 58 genotyped *NOS* SNPs for interaction with long-term traffic-related air pollution, similarly to what was done using SNPs from the *GST* gene family in Paper II. The approach laid out in Papers II and III can be used for this and similar analyses in the future.

For successful progress and discoveries in gene-environment interaction research, it is important that methods be further investigated. A relative measurement can be of interest if it is interpretable and can be compared across studies, which is only possible if all assumptions and characteristics of included variables are explicitly defined. An absolute measurement of additive interaction for other settings than that of two dichotomous variables has not yet been presented. It is hoped that the pragmatic proposal presented in Paper III can be a springboard for further investigations of the subject, since the approach tries to reconcile the estimates from the commonly used (but inherently multiplicative across a continuous variable) logistic regression with the conceptual additivity of risk in the original RERI definition. The details of the statistical implications of the approach remain to be elucidated, as well as validating the approach for different settings and exposure ranges.

As for the specific burden of CVD, a part of it is due to genetic susceptibility, which is important to investigate further. But at least as important is investigating human everyday exposures, including air pollution and lifestyle choices such as smoking, and maintenance of general health by prophylaxis as well as adherence to medication [Butler et al. 2002, Nichol, Venturini & Sung 1999]. Considering both genetic susceptibility and the individual everyday exposure, it appears that individual risk is the result of a complex equation involving both factors that can be manipulated as well as fixed, predetermined factors. Thus, while the focus from a public health perspective may be to identify common disease mechanisms and exposure patterns, the individual patient as encountered by primary care physicians may benefit from a more personalized approach which considers as many pieces of the puzzle as possible. It may also be valuable to realize that genetic factors may influence lifestyle, both adversely and beneficially.

However, a bit of caution ought to be applied when genetic research questions are formulated, investigated and the results shared with the public. The effects of individual genetic screening and mapping on society and the individual are not yet known. The opinions on genetic screening differ within and between populations. The ultimate objective for genetic exploration is often said to be personalized medicine, but the implications of such a development need to be considered.

On the other hand, gene-environment interaction studies can also be used to elucidate disease pathology mechanisms and further increase the knowledge of the human body's molecular functions. Personally, I believe that gene-environment studies on multiple comorbidities between patient groups can take us a long way towards understanding etiology at the causal and pathway levels. For example, by studying how individuals move through the four states of a comorbidity, i.e. unaffected, has disease A, has disease B and has both disease A and B, also considering which risk factors the diseases share and not share, the mechanisms of each disease may be clarified further and there is the opportunity to evaluate temporality for causal inference.

All in all, while John Donne as early as in 1624 noted that "No man [is] an island" [Donne 1624], future perspectives in CVD research are bound to focus on both individual susceptibility and environmental exposure in one form or another, i.e. interaction between individual characteristics and societal exposure.

ACKNOWLEDGEMENT

I would like to thank…

My main supervisor **Fredrik Nyberg**, who has taught me how to write the best possible manuscripts, to always be critical of everything and how to do epidemiology outside the textbook. For encouraging me to have an opinion, and for letting me travel to network and take courses.

My co-supervisors **Anna-Carin Olin**, **Annika Rosengren** and **Lena Björck** for your pep-talks, input on manuscripts and sharing your knowledge.

The **co-authors of my papers**, for all the discussions, input on data collection and thoughtful proof-reading.

My brilliant surgeon ophthalmologist **Maria Egardt** without whom I would be legally blind by now and who always asked how the PhD project was going (even though the waiting room was brimming with patients…)

My parents **Barbro** and **Anders**, who by always expecting the best made me do my best. My aunt **Birgitta** who has always looked out for me. My brothers **Christer** and **Ulf** and my extended family. We don't see each other very often, but I like to know that you're out there doing well.

My best friend **Lina** who always pretends to think that everything I do is cool, even though she really just thinks I should move back to Stockholm and be done with it.

My **friends** and **colleagues**, especially my **fellow PhD students**, here at **AMM**. If I begin citing names, the list would go on forever. You know who you are. For the talks and coffees and for always asking how I am when I show up with new bruises or worse from my "extreme-sporting". I suspect you all have a betting pool going for what will happen next and when…

One of those is my office roommate **Emilia**, who is the best roommate one could hope for, bringing a couch and fruit and reminding me to eat, go home and such important stuff during the production of this thesis :)

The equine creatures **Whisky** and **Rockan**. Compared to winning you guys over and keeping the peace, writing this thesis was really straightforward. You make me humble, and that's good for every academic, right?

Finally, my boyfriend **Magnus**. Thank you for all the help with the horses and for cooking dinners to make sure I eat. Also for putting up with my workaholic, horseaholic and sleepaholic tendencies (why do anything in moderation y'know…)

For **those of you in the PhD business**: The only way to successfully write a thesis is by having a good assortment of snacks by your desk. At all times.

THANKS!

The research in this thesis was funded by grants from the Västra Götaland County Council, the Swedish Council for Working Life and Social Research, the Swedish Research Council, the Swedish Research Council for Environment and Spatial Planning, the Swedish Heart and Lung Foundation, and AstraZeneca R&D Sweden.

REFERENCES

(2010) Special Report 17, Boston, MA, Health Effects Institute (HEI).

(2013) Fact Sheet N^{o} 317, Geneva, World Health Organisation (WHO).

(2013) Document Number: GEN1300666, to: 23andMe, Inc., warning letter, Silver Springs, MD, Food and Drug Administration (FDA), Department of Health and Human Services, 22 Nov 2013.

Ahlbom A & Alfredsson L (2005) Interaction: A word with two meanings creates confusion. Eur J Epidemiol 20: 563–564. doi: 10.1007/s10654-005-4410-4.

Ahlner J & Johansson J (1994) Kardiovaskulär sjukdom. Boehringer Mannheim Scandinavia, ISBN 91-972279-0-0.

Aminuddin F, Hackett TL, Stefanowicz D, Saferali A, Paré PD, Gulsvik A, Bakke P, Cho MH, Litonjua A, Lomas DA, Anderson WH, Beaty TH, Silverman EK, Sandford AJ (2013) Nitric oxide synthase polymorphisms, gene expression and lung function in chronic obstructive pulmonary disease. BMC Pulm Med 13:64. doi: 10.1186/1471-2466-13-64.

Andrews DF (1998) Robust Regression. In: Encyclopedia of Biostatistics, vol 5(6), John Wiley & Sons Ltd, ISBN 0-471-97576-1.

Annas GJ & Elias S (2014a) 23andMe and the FDA. N Engl J Med 370(11):985-988. doi: 10.1056/NEJMp1316367

Annas GJ & Elias S (2014b) Correspondence: 23andMe and the FDA. N Engl J Med 370(23):2248-9. doi: 10.1056/NEJMc1404692

Berg CM, Lappas G, Strandhagen E, Wolk A, Torén K, Wilhelmsen L, Rosengren A, Thelle DS (2005) Trends in blood lipid levels, blood pressure, alcohol and smoking habits from 1985 to 2002: results from INTERGENE and GOT-MONICA. J Cardiovasc Risk 12 (2):115–25.

Berg CM, Lappas G, Strandhagen E, Wolk A, Torén K, Rosengren A, Aires N, Thelle DS, Lissner L (2008) Food patterns and cardiovascular disease risk factors: The Swedish INTERGENE research program. Am J Clin Nutr 88(2): 289–297.

Berger K, Stögbauer F, Stoll M, Wellmann J, Huge A, Cheng S, Kessler C, John U, Assmann G, Ringelstein E, Funke H (2007) The glu298asp polymorphism in the nitric oxide synthase 3 gene is associated with the risk of ischemic stroke in two large independent case–control studies. Hum Genet 121(2): 169-78. doi: 10.1007/s00439-006-0302-2

Bessa SS, Ali EMM, Hamdy SM (2009) The role of glutathione S-transferase M1 and T1 gene polymorphisms and oxidative stress-related parameters in Egyptian patients with essential hypertension. European Journal of Internal Medicine 20: 625–630. doi: 10.1016/j.ejim.2009.06.003

Birkett NJ (1992) Effect of nondifferential misclassification on estimates of odds ratios with multiple levels of exposure. Am J Epidemiol 136(3): 356-62.

Brenner H & Loomis D (1994) Varied forms of bias due to nondifferential error in measuring exposure. Epidemiology 5:510-517.

Brook RD & Rajagopalan S (2009) Particulate matter, air pollution, and blood pressure. J Am Soc Hypertens 3:332-350.

Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, Holguin F, Hong Y, Luepker RV, Mittleman MA, Peters A, Siscovick D, Smith SC Jr, Whitsel L, Kaufman JD, on behalf of the American Heart Association Council on Epidemiology and Prevention, Council on the Kidney in Cardiovascular Disease, and Council on Nutrition, Physical Activity and Metabolism (2010) Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement From the American Heart Association. Circulation 121: 2331–2378. doi: 10.1161/cir.0b013e3181dbece1

Brunekreef B (2007) Health effects of air pollution observed in cohort studies in Europe. J Expos Sci Environ Epidemiol 17: S61–S65. doi: 10.1038/sj.jes.7500628

Butler J, Arbogast PG, BeLue R, Daugherty J, Jain MK, Ray WA, Griffin MR (2002) Outpatient adherence to beta-blocker therapy after acute myocardial infarction. J Am Coll Cardiol 40(9):1589-1595. doi:10.1016/S0735-1097(02)02379-3

Carere DA, Couper MP, Crawford SD, Kalia SS, Duggan JR, Moreno TA, Mountain JL, Roberts JS, Green RC; PGen Study Group (2014) Design, methods, and participant characteristics of the Impact of Personal Genomics (PGen) Study, a prospective cohort study of direct-to-consumer personal genomic testing customers. Genome Med 6(12):96. doi: 10.1186/s13073-014-0096-0.

Casas JP, Cavalleri GL, Bautista LE, Smeeth L, Humphries SE, Hingorani AD (2006) Endothelial nitric oxide synthase gene polymorphisms and cardiovascular disease: a HuGE review. Am J Epidemiol 164(10): 921-935. doi: 10.1093/aje/kwj302

Campen MJ, Lund A, Rosenfeld M (2012) Mechanisms linking traffic-related air pollution and atherosclerosis. Current Opinion in Pulmonary Medicine 18: 155–160. doi: 10.1097/mcp.0b013e32834f210a

Chow CK, Teo KK, Rangarajan, S, Islam S, Gupta R, Avezum A, Bahonar A, Chifamba J, Dagenais G, Diaz R, Kazmi K, Lanas F, Wei L, Lopez-Jaramillo P, Fanghong L, Ismail NH, Puoane T, Rosengren A, Szuba A, Temizhan A, Wielgosz A, Yusuf R, Yusufali A, McKee M, Liu L, Mony P, Yusuf S, for the PURE (Prospective Urban Rural Epidemiology) Study investigators (2013) Prevalence, Awareness, Treatment, and Control of Hypertension in Rural and Urban Communities in High-, Middle-, and Low-Income Countries. JAMA 310(9): 959-968. doi:10.1001/jama.2013.184182

Coogan PF, White LF, Jerrett M, Brook RD, Su JG, Seto E, Burnett R, Palmer JR, Rosenberg L (2012) Air Pollution and Incidence of Hypertension and Diabetes in African American Women Living in Los Angeles. Circulation 125(6):767-772. doi:10.1161/CIRCULATIONAHA.111.052753.

Cooper C (2014) £125 genetic test kit backed by Google arrives in Britain – with a health warning. In: the Independent 2 Dec 2014, accessed online 10 Jan 2015 at http://www.independent.co.uk/life-style/gadgets-and-tech/news/googles-125-cancer-test-arrives-in-britain--but-with-a-health-warning-9896684.html.

Dahgam S, Nyberg F, Modig L, Naluai AT, Olin AC (2012) Single nucleotide polymorphisms in the NOS2 and NOS3 genes are associated with exhaled nitric oxide. J Med Genet 49(3):200-5. doi: 10.1136/jmedgenet-2011-100584.

Donne J (1624) Meditation XVII. In: Devotions upon emergent occasions, Kingdom of England.

Du Y, Wang H, Fu X, Sun R, Liu Y (2012) GSTT1 null genotype contributes to coronary heart disease risk: a meta-analysis. Mol Bio Rep 39:8571-8579. doi: 10.1007/s11033-012-1691-z

Dunning AM, Healey CS, Pharoah PDP, Teare MD, Ponder BAJ, Easton DF (1999) A Systematic Review of Genetic Polymorphisms and Breast Cancer Risk. Cancer Epidemiol Biomarkers Prev 8(10):843-54.

Forastiere F, Stafoggia M, Tasco C, Picciotto S, Agabiti N, Cesaroni G and Perucci CA (2007) Socioeconomic status, particulate air pollution, and daily mortality: Differential exposure or differential susceptibility. Am J Ind Med 50: 208–216. doi: 10.1002/ajim.20368.

Förstermann U & Sessa WC (2012) Nitric oxide synthases: regulation and function. Eur Heart J 33(7): 829-837. doi: 10.1093/eurheartj/ehr304.

Gonzales-Gay MA, Llorca J, Palomino-Morales R, Gomez-Acebo I, Gonzalez-Juanatey C, Martin J (2009) Influence of nitric oxide synthase gene polymorphisms on the risk of cardiovascular events in rheumatoid arthritis. Clin Exp Rheumatol 27(1): 116-119.

Greenland S (2009) Interactions in Epidemiology: Relevance, Identification, and Estimation. Epidemiology 20(1): 14-17. doi: 10.1097/EDE.0b013e318193e7b5.

Greenland S & Morgenstern H (1989) Ecological Bias, Confounding, and Effect Modification. Int J Epidem 18: 269-274.

Greenland S & Thomas DC (1982) On the need for the rare disease assumption in case-control studies. Am J Epidem 116(3): 547-553.

Gustavsson J, Mehlig K, Leander K, Strandhagen E, Björck L, Thelle DS, Lissner L, Blennow K, Zetterberg H, Nyberg F (2012) Interaction of apolipoprotein E genotype with smoking and physical inactivity on coronary heart disease risk in men and women. Atherosclerosis 220(2):486-92. doi: 10.1016/j.atherosclerosis.2011.10.011.

Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YMTA, Chen T, Barr RG, Schabath MB, Couper,DJ, Brusselle GG, Psaty BM, van Duijn CM, Rotter JI, Uitterlinden AG, Hofman A, Punjabi NM, Rivadeneira F, Morrison AC, Enright PL, North KE, Heckbert SR, Lumley T, Stricker BHC, O'Connor GT, London SJ (2010) Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat Genet 42(1): 45-52.

Hingorani AD, Liang CF, Fatibene J, Lyon A, Monteith S, Parsons A, Haydock S, Hopper RV, Stephens NG, O'Shaughnessy KM, Brown MJ (1999) A Common Variant of the Endothelial Nitric Oxide Synthase (Glu298⇒Asp) Is a Major Risk Factor for Coronary Artery Disease in the UK. Circulation 100(14): 1515-1520.

Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA (2002) Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. The Lancet 360(9341): 1203-1209. doi: 10.1016/S0140-6736(02)11280-3

Hosmer DW & Lemeshow S (1992) Confidence interval estimation of interaction. Epidemiology 3: 452–456. doi: 10.1097/00001648-199209000-00012

Ikonomidis I, Michalakeas CA, Parissis J, Paraskevaidis I, Ntai K, Papadakis I, Anastasiou-Nana M, Lekakis J (2012) Inflammatory markers in coronary artery disease. Biofactors 38(5):320-8. doi: 10.1002/biof.1024.

Iwai N, Tago N, Yasui N, Kokubo Y, Inamoto N, Tomoike H, Shioji K (2004) Genetic analysis of 22 candidate genes for hypertension in the Japanese population. J Hypertens 22(6): 1119-1126.

Jàchymovà M, Horky K, Bultas J, Kozich V, Jindra A, Peleska J, Martásek P (2001) Association of the Glu298Asp Polymorphism in the Endothelial Nitric Oxide Synthase Gene with Essential Hypertension Resistant to Conventional Therapy. Biochem Biophys Res Comm 284(2): 426-430. doi: 10.1006/bbrc/2001.5007

Janssen W, Pullamsetti SS, Cooke J, Weissmann N, Guenther A, Schermuly RT. (2013) The role of dimethylarginine dimethylaminohydrolase (DDAH) in pulmonary fibrosis. J Pathol 229(2):242-9. doi: 10.1002/path.4127.

Johansson C, Sällsten G, Bouma H, Johannesson S, Gustafsson S, Eneroth K, Norman M, Kruså M, Tinnerberg H, Bellander T (2006) EXPOSE: Exposure – comparison between measurements and calculations based on dispersion modelling. Swedish National Air Pollution and Health Effects Program (SNAP); Stockholm, Sweden.

Johnson T, Gaunt TR, Newhouse SJ, Padmanabhan S, Tomaszewski M, Kumari M, Morris RW, Tzoulaki J, O'Brien ET, Poulter NR, Sever P, Shields DC, Thom S, Wannamethee SG, Whincup PH, Brown MJ, Connell JM, Dobson RJ, Howard PJ, Mein CA, Onipinla A, Shaw-Hawkins S, Zhang Y, Smith GD, Day INM, Lawlor DA, Goodall AH, Fowkes FG, Abecasis GR, Elliott P, Gateva V, Braund PS, Burton PR, Nelson CP, Tobin

MD, van der Harst P, Glorioso N, Neuvrith H, Salvi E, Staessen, JA, Stucchi A, Devos N, Jeunemaitre X, Plouin P-F, Tichet J, Juhanson P, Org E, Putku M, Sõber S, Veldre G, Viigimaa M, Levinsson A, Rosengren A, Thelle DS, Hastie CE, Hedner T, Lee WK, Melander O, Wahlstrand B, Hardy R, Wong A, Cooper JA, Palmen J, Chen L, Stewart AFR, Wells GA, Westra HJ, Wolfs MGM, Clarke R, Franzosi MG, Goel A, Hamsten A, Lathrop M, Peden JF, Seedorf U, Watkins H, Ouwehand WH, Sambrook J, Stephens J, Casas J-P, Drenos F, Holmes MV, Kivimaki M, Shah S, Shah T, Talmud PJ, Whittaker J, Wallace C, Delles C, Laan M, Kuh D, Humphries SE, Nyberg F, Cusi D, Roberts R, Newton-Cheh C, Franke L, Stanton AV, Dominiczak AF, Farrall M, Hingorani AD, Samani NJ, Caulfield MJ, Munroe PB (2011) Blood Pressure Loci Identified with a Gene-Centric Array. Am J Hum Genet 89(6): 688-700.

Jorde LB, Carey JC, Bamshad MJ (2005) Medical Genetics, 3rd ed. Mosby Elsevier, St Louis, ISBN 0-323-04035-7.

Katsoulis M & Bamia C (2014) Additive Interaction between Continuous Risk Factors Using Logistic Regression. Epidemiology 25(3): 462-464. doi: 10.1097/EDE.0000000000000083.

Kaufman JS (2009) Interaction Reaction. Epidemiology 20(2):159-60. doi: 10.1097/EDE.0b013e318197c0f5.

Khoury MJ (1997) Genetic Epidemiology and the Future of Disease Prevention and Public Health. Epidemiol Rev 19(1): 175-180.

Khoury MJ & Wagener DK (1995) Epidemiological evaluation of the use of genetics to improve the predictive value of disease risk factors. Am J Hum Genet 56:835-844.

Kim EH, Carrigan TP, Menon V (2008) Enrollment of Women in National Heart, Lung, and Blood Institute-Funded Cardiovascular Randomized Controlled Trials Fails to Meet Current Federal Mandates for Inclusion. J Am Coll Cardiol 52(8): 672-673. doi:10.1016/j.jacc.2008.05.025.

Knol MJ, van der Tweel I, Grobbee DE, Numans ME, Geerlings MI (2007) Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int J Epidemiol 36(5): 1111-1118.

Knol MJ, VanderWeele TJ, Groenwold RHH, Klungel OH, Rovers MM, Grobbee DE (2011) Estimating measures of interaction on an additive scale for preventive exposures. Eur J Epidemiol 26(6): 433-438. doi:10.1007/s10654-011-9554-9.

Knol MJ & VanderWeele TJ (2012) Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol 41(2): 514-20. doi: 10.1093/ije/dyr218.

LaBaer J (2002) Genomics, proteomics, and the new paradigm in biomedical research. Genet Med 4 (S6): 2S-9S.

Lettre G, Lange C, Hirschhorn J (2007) Genetic model testing and statistical power in population-based association studies of quantitative traits. Genet Epidemiol 31: 358–362. doi: 10.1002/gepi.20217

Marchenko YV, Carroll RJ, Lin DY, Amos CI, Gutierrez RG (2008) Semiparametric analysis of case-control genetic data in the presence of environmental factors. Stata Journal 8 (3): 305-333.

Meschia JF, Nalls M, Matarin M, Brott TG, Brown RD, Hardy J, Kissela B, Rich SS, Singleton A, Hernandez D, Ferrucci L, Pearce K, Keller M, Worrall BB for the Siblings With Ischemic Stroke Study Investigators (2011) Siblings With Ischemic Stroke Study. Stroke 42(10): 2726-2732. doi: 10.1161/strokeaha.111.620484

Minelli C, Granell R, Newson R, Rose-Zerilli MJ, Torrent M, Ring SM, Holloway JW, Shaheen SO, Henderson JA (2009) Glutathione-S-transferase genes and asthma phenotypes: a Human Genome Epidemiology (HuGE) systematic review and meta-analysis including unpublished data. Int J Epidemiol. doi:10.1093/ije/dyp337.

Mordukhovich I, Wilker E, Suh H, Wright R, Sparrow D, Vokonas PS, Schwartz J (2009) Black Carbon Exposure, Oxidative Stress Genes, and Blood Pressure in a Repeated-Measures Study. Env Health Perspect 117(11): 1767-72.

Nichol MB, Venturini F, Sung JC (1999) A critical evaluation of the methodology of the literature on medication compliance. Pharmacother 33: 531-540.

Nilsson J (2010) Åderförkalkning och inflammation. In Akut kranskärlssjukdom, ed. Wallentin L & Lindahl B, Liber AB, Stockholm. ISBN 978-91-47-09388-5

Niu W & Qi Y (2011) An Updated Meta-Analysis of Endothelial Nitric Oxide Synthase Gene: Three Well-Characterized Polymorphisms with Hypertension. PLoS ONE 6(9): e24266. doi: 10.1371/journal.pone.0024266

Nørskov MS (2013) GSTT1 null genotype and risk of coronary heart disease. Mol Bio Rep 40:2015-2017. doi 10.1007/s11033-012-2260-1

Ottman R (1995) Gene-Environment Interaction and Public Health. Am J Hum Genet 56:821-823.

Padmanabhan S, Menni C, Lee WK, Laing S, Brambilla P, Sega R, Perego R, Grassi G, Cesana G, Delles C, Mancia G, Dominiczak AF (2010) The effects of sex and method of blood pressure measurement on genetic associations with blood pressure in the PAMELA study. J Hypertens 28(3): 465-477.

Pearce N (1993) What does the odds ratio estimate in a case-control study? Int J Epidemiol 22(6): 1189-1192.

Pemble S, Schroeder KR, Spencer SR, Meyer DJ, Hallier E, Bolt HM, Ketterer B, Taylor JB (1994) Human glutathione S-transferase theta

(GSTT1): cDNA cloning and the characterization of a genetic polymorphism. Biochem J 300: 271-276.

Persson S (1986) Kardiologi – hjärtsjukdomar hos vuxna, 2nd ed. Studentlitteratur, Lund. ISBN 91-44-17232-X.

Peters A (2005) Particulate matter and heart disease: Evidence from epidemiological studies. Toxicology and Applied Pharmacology 207: 477–482. doi: 10.1016/j.taap.2005.04.030

Peters A, Veronesi B, Calderon-Garciduenas L, Gehr P, Chen L, et al. (2006) Translocation and potential neurological effects of fine and ultrafine particles a critical update. Particle and Fibre Toxicology 3: 13. doi: 10.1186/1743-8977-3-13

Pigliucci M (2001) Epilogue: beyond nature and nurture. In: Phenotypic Plasticity: Beyond Nature and Nurture. Johns Hopkins University Press. ISBN 978-0801867880.

Probst-Hensch NM, Imboden M, Dietrich DF, Barthélemy J-C, Ackermann-Liebrich U, Berger W, Gaspoz J-M, Schwartz J (2008) Glutathione S-Transferase Polymorphisms, Passive Smoking, Obesity, and Heart Rate Variability in Nonsmokers. Environ Health Perspect 116(11): 1494-1499. doi:10.1289/ehp.11402

Raza H (2011) Dual localization of glutathione S-transferase in the cytosol and mitochondria: implications in oxidative stress, toxicity and disease. FEBS Journal 278: 4243–4251. doi: 10.1111/j.1742-4658.2011.08358.x

Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat Me, Zhao JH, Ramasamy A, Zhai G, Vitart V, Huffman JE, Igl W, Albrecht E, Deloukas P, Henderson J, Granell R, McArdle WL, Rudnicka AR, Barroso I, Loos RJF, Wareham NJ, Mustelin L, Rantanen T, Surakka I, Imboden M, Wichmann HE, Grkovic I, Jankovic S, Zgaga L, Hartikainen A-L, Peltonen L, Gyllensten U, Johansson A, Zaboli G, Campbell H, Wild SH, Wilson JF, Glaser S, Homuth G, Volzke H, Mangino M, Soranzo N, Spector TD, Polasek O, Rudan I, Wright AF, Heliovaara M, Ripatti S, Pouta A, Naluai AT, Olin A-C, Toren K, Cooper MN, James AL, Palmer LJ, Hingorani AD, Wannamethee SG, Whincup PH, Smith GD, Ebrahim S, McKeever TM, Pavord ID, MacLeod AK, Morris AD, Porteous DJ, Cooper C, Dennison E, Shaheen S, Karrasch S, Schnabel E, Schulz H, Grallert H, Bouatia-Naji N, Delplanque J, Froguel P, Blakey JD, Britton JR, Morris RW, Holloway JW, Lawlor DA, Hui J, Nyberg F, Jarvelin M-R, Jackson C, Kahonen M, Kaprio J, Probst-Hensch NM, Koch B, Hayward C, Evans DM, Elliott P, Strachan DP, Hall IP, Tobin MD (2010) Genome-wide association study identifies five loci associated with lung function. Nat Genet 42(1): 36-44.

Richardson DB & Kaufman JS (2009) Estimation of Relative Excess Risk due to Interaction and associated confidence bounds. Am J Epidemiol 169 (6): 756-760.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic Structure of Human Populations. Science 298: 2381-2384. doi: 10.1126/science.1078311

Rothman KJ (1986) Modern Epidemiology. 1st Ed., Little, Brown & Company. Boston, MA.

Rothman KJ, Greenland S, Lash TL (2008) Modern Epidemiology, 3rd Ed., Lippincott Williams & Wilkins, Philadelphia, PA.

Shrey K, Suchit A, Deepika D, Shruti K, Vibha R (2011) Air pollutants: The key stages in the pathway towards the development of cardiovascular disorders. Environmental Toxicology and Pharmacology 31: 1–9. doi: 10.1016/j.etap.2010.09.002

Siegbahn A (2010) Trombosmekanismer. In: Akut kranskärlssjukdom, ed. Wallentin L & Lindahl B, Liber AB, Stockholm. ISBN 978-91-47-09388-5

Skrondal A (2003) Interaction as departure from additivity in case-control studies: a cautionary note. Am J Epidemiol 158(3): 251-258.

Steele BD (2014) The Ending of the Nature vs Nurture Debate, https://benjamindavidsteele.wordpress.com/2014/01/16/the-ending-of-the-nature-vs-nurture-debate/ accessed on 30th Dec 2014

Stephens JW, Bain SC, Humphries SE (2008) Gene-environment interaction and oxidative stress in cardiovascular disease. Atherosclerosis 200: 229-238.

Stramba-Badiale M (2009) Red Alert on Women's Hearts: Women and Cardiovascular Research in Europe. European Society of Cardiology (European Heart Health Strategy, EuroHeart Project): Work Package 6 Women and Cardiovascular Diseases.

Strandhagen E, Berg C, Lissner L, Nunez L, Rosengren A, Torén K, Thelle DS (2010) Selection bias in a population survey with registry linkage: potential effect on socioeconomic gradient in cardiovascular risk. Eur J Epidemiol 25 (3): 163-172.

Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 3(5): 391-397.

Talmud PJ (2007) Gene–environment interaction and its impact on coronary heart disease risk. Nutr Metab Cardiovasc Dis. 17(2):148-52. doi: 10.1016/j.numecd.2006.01.008.

Tepliakov ATSS, Berezikova EN, Iakovleva NF, Maianskaia SD, Popova AA, Luksha EB, Voronina EN, Torim IuIu, Karpov RS (2010) Polymorphism of eNOS and iNOS Genes and Chronic Heart Failure in Patients With Ischemic Heart Disease. Kardiologiia 50(4): 23-30.

VanderWeele TJ (2011) A word and that to which it once referred: assessing "biologic" interaction. Epidemiology 22(4): 612-613.

Vermylen J, Nemmar A, Nemery B, Hoylaerts MF (2005) Ambient air pollution and acute myocardial infarction. J Thromb Haemost 3: 1955–61.

Wang XL, Rainwater DL, VandeBerg JF, Mitchell BD, Mahaney MC (2001) Atherosclerosis and Lipoproteins: Genetic Contributions to Plasma Total Antioxidant Activity. Arterioscler Thromb Vasc Biol 21: 1190-1195. doi: 10.1161/hq0701.092146

White DL, Li D, Nurgalieva Z, El-Serag HB (2008) Genetic Variants of Glutathione S-Transferase as Possible Risk Factors for Hepatocellular Carcinoma: A HuGE Systematic Review and Meta-Analysis. Am J Epidemiol 167(4): 377-389. doi: 10.1093/aje/kwm315.

Wilcox RR (1998) Trimming and Winsorization. In: Encyclopedia of Biostatistics, vol 6(6), John Wiley & Sons Ltd, ISBN 0-471-97576-1.

Wilcox RR (2014) Gaining a deeper and more accurate understanding of data via modern robust statistical techniques. J Psychol Clin Psychiatry 1 (2): 00012.

Young RP, Hopkins R, Black PN, Eddy C, Wu L, Gamble GD, Mills GD, Garrett JE, Eaton TE, Rees MI (2006) Chronic obstructive pulmonary disease: Functional variants of antioxidant genes in smokers with COPD and in those with normal lung function. Thorax 61(5): 394-399 doi: 10.1136/thx.2005.048512

Yuji K, Tanimoto T, Oshima Y (2014) Correspondence: 23andMe and the FDA. N Engl J Med 370(23):2248-2249. DOI: 10.1056/NEJMc1404692

Yusuf S, Hawken S, Ôunpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L, on behalf of the INTERHEART Study Investigators (2004) Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 364(9438): 937-952.

Zanobetti A, Baccarelli A, Schwartz J (2011) Gene–Air Pollution Interaction and Cardiovascular Disease: A Review. Prog Cardiovasc Dis 53:344-352.

Zeka A, Zanobetti A, Schwartz J (2006) Individual-level modifiers of the effects of particulate matter on daily mortality. Am J Epidemiol 163(9): 849-859. doi: 10.1093/aje/kwj116.

APPENDIX

Table 1. Analyses of effect of NO2 per 10µg/m3 on respective outcome AMI and hypertension, stratified for genotype and 3-level smoking variable.

				Effect estimate of NO_2 per 10µg/m^3 increase			
Outcome	***Gene:* SNP**	**Genotype**	**Smoking***	**OR**	**95% CI**		**p-value**
AMI	*GSTP1*	GG	0	1.00	0.37	2.73	1.00
	rs596603		1	1.57	0.74	3.30	0.24
			2	1.67	0.69	4.02	0.25
		GT + TT	0	1.79	0.73	4.43	0.21
			1	2.28	1.08	4.82	0.03
			2	3.22	1.39	7.44	0.006
Hypertension	*GSTP1*	CC	0	0.95	0.56	1.61	0.84
	rs1871042		1	0.92	0.56	1.53	0.76
			2	0.74	0.38	1.43	0.36
		TC + TT	0	0.97	0.57	1.67	0.92
			1	0.62	0.36	1.06	0.08
			2	0.60	0.32	1.12	0.11
	GSTP1	GG	0	0.95	0.56	1.61	0.85
	rs749174		1	0.95	0.58	1.57	0.84
			2	0.76	0.39	1.47	0.41
		AG + AA	0	1.02	0.60	1.74	0.94
			1	0.63	0.37	1.09	0.10
			2	0.62	0.33	1.16	0.14
	GSTP1	CC	0	0.76	0.42	1.37	0.36
	rs762803		1	0.83	0.48	1.44	0.50
			2	0.52	0.23	1.16	0.11
		CA + AA	0	1.16	0.70	1.91	0.58
			1	0.71	0.43	1.19	0.19
			2	0.69	0.39	1.23	0.21
AMI	*GSTT1*	Null	0	No cases in this stratum			
	rs2266637		1	1.19	0.22	6.34	0.84

			2	1.76	0.31	9.94	0.52
		Non-null	0	1.36	0.60	3.07	0.47
			1	1.83	0.99	3.41	0.06
			2	2.63	1.25	5.50	0.01
Hypertension		Null	0	1.00	0.43	2.35	1.00
			1	0.62	0.25	1.57	0.32
			2	0.47	0.16	1.34	0.16
		Non-null	0	1.03	0.64	1.65	0.91
			1	0.82	0.53	1.26	0.36
			2	0.71	0.41	1.23	0.22
AMI	*GSTCD*	AA	0	1.49	0.65	3.41	0.34
	rs10516526		1	2.06	1.09	3.91	0.03
			2	2.73	1.30	5.74	0.01
		AG + GG	0	0.57	0.08	3.95	0.57
			1	1.01	0.25	4.14	0.99
			2	1.45	0.27	7.87	0.67
Hypertension		AA	0	1.03	0.65	1.66	0.88
			1	0.84	0.54	1.30	0.42
			2	0.75	0.43	1.31	0.31
		AG + GG	0	1.04	0.46	2.34	0.93
			1	0.68	0.28	1.65	0.39
			2	0.39	0.11	1.37	0.14

* Smoking status: 0= never smoker, 1= former smoker, 2= current smoker

Erratum

Table 2. Corrected results for the GSTT1-stratified analysis of effect of long-term traffic-related air pollution on risk of AMI from Paper II.

		Effects per 10μg/m^3 of NO_2			
Gene:SNP	**Genotype**	**OR**	**95% CI**	**p-value**	**Interaction p-value**
GSTT1: rs2266637	Null	1.37	0.32 – 5.89	0.67	0.60
	Non-null	2.01	1.13 – 3.58	0.02	

Model is adjusted for age, age squared, sex, BMI, educational level (included correctly as a categorical variable, rather than incorrectly as a continuous variable) and residential area.

The estimates were only very marginally changed as compared to the published version (Table 4, Paper II), with no change in interpretation.