Context dependent associations between variation in risk of ischemic heart disease and variation in the 5′ promoter region of the apolipoprotein E gene in Danish females

Jari H. Stengård, M.D., Ph.D.1, Greg Dyson, Ph.D.1, Ruth Frikke-Schmidt, M.D., Ph.D.2, Anne Tybjærg-Hansen, M.D., D.M.Sc.2,4, Borge G. Nordestgaard, M.D., D.M.Sc.3,4 and Charles F. Sing, Ph.D.1

1Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA, 2Department of Clinical Biochemistry, Section for Molecular Genetics, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark, 3Department of Clinical Biochemistry, Herlev University Hospital, Herlev, Denmark, 4The Copenhagen City Heart Study, Bispebjerg University Hospital, Copenhagen, Denmark

Correspondence to:
Charles F. Sing
Department of Human Genetics
University of Michigan
1241 E. Catherine Street, 5928 Buhl Building
Ann Arbor, MI 48109-5618, USA
Ph: 734-764-5492; Fax: 734-763-5277; Email: csing@umich.edu

Journal Subject Codes: [89] Genetics of cardiovascular disease; [135] Risk Factors
ABSTRACT

Background  Variations in the non-coding single nucleotide polymorphisms (SNPs) at positions 560 and 832 in the 5′ promoter region of the apolipoprotein E (APOE) gene define genotypes that distinguish between high and low concentrations of plasma total and high-density lipoprotein (HDL) cholesterol and triglycerides. We address whether these genotypes improve the prediction of ischemic heart disease (IHD) in sub-samples of individuals defined by traditional risk factors and the genotypes defined by the ε2, ε3 and ε4 alleles in exon four of the APOE gene.

Methods and Results  In a sample of 3,686 female and 2,772 male participants of the Copenhagen City Heart Study (CCHS) who were free of IHD events, 576 individuals (257 females, 7.0 % and 319 males, 11.5 %) were diagnosed as having developed IHD in 6.5 years of follow up. Using a stepwise Patient Rule-Induction Method modeling strategy that acknowledges the complex pathobiology of IHD, we identified a sub-sample of 764 elderly females (≥ 65 years) with hypertriglyceridemia who had either a history of smoking, a history of hypertension or a history of both in which the A560T832/A560T832 and A560T832/A560G832 5′ two-SNP genotypes had a higher cumulative incidence of IHD (172/1,000) compared to the incidence of 70/1,000 in the total sample of females.

Conclusions  Our study validates that 5′ APOE genotypes improve the prediction of IHD and documents that the improvement is greatest in a subset defined by a particular combination of traditional risk factors in CCHS females. We discuss the utility of these genotypes in medical risk assessment of IHD in the population represented by the CCHS.

Key Words: atherosclerosis; genetics; risk factors
INTRODUCTION

Cholesterol accumulation in arterial walls is an important contributing factor in the development of ischemic heart disease (IHD).\textsuperscript{1} A plethora of variations in genes and their expressed products involved in lipid metabolism have been characterized.\textsuperscript{2-7} Fast and inexpensive gene-measurement technologies can be used to identify genetic variations that predict inter-individual variation in measures of lipid metabolism and risk of IHD.\textsuperscript{8-11} The promise is that the identified variants would then be used as additional information in risk assessment to guide the selection of non-pharmacological and pharmacological interventions to prevent initiation, progression and/or severity of IHD.\textsuperscript{11}

Variations in the gene coding the apolipoprotein E (apoE) protein have been implicated in predicting variation in plasma lipids and risk of IHD. ApoE is a constituent of many atherogenic lipoprotein particles, such as triglyceride (TG) -rich chylomicrons and high-density lipoproteins (HDLs). The apoE molecule has three common isoforms, E2, E3 and E4 encoded by variation in three common alleles, $\varepsilon_2$, $\varepsilon_3$ and $\varepsilon_4$, defined by two single nucleotide polymorphisms (SNPs) in exon four of the apolipoprotein E (APOE) gene. Studies assessing the role of the APOE gene in lipid metabolism have repeatedly demonstrated that variations in plasma total cholesterol (T-C) and TG levels are associated with variation among genotypes defined by the $\varepsilon_2$, $\varepsilon_3$ and $\varepsilon_4$ alleles.\textsuperscript{12,13} Individuals carrying the $\varepsilon_4$ allele are also at higher risk for developing atherosclerosis both in animals and humans.\textsuperscript{14-17}

In a previous study we demonstrated that common variations in the non-coding SNPs located at positions 560 and 832 in the 5' promoter region of the APOE gene (see Stengård et al\textsuperscript{15} for details) define three genotypes, $A_{560}T_{832}/A_{560}T_{832}$, $A_{560}T_{832}/A_{560}G_{832}$ and $A_{560}T_{832}/T_{560}T_{832}$, that distinguish between high and low concentrations of plasma T-C, HDL-C and TG in four
independent samples ascertained to represent the human population at large.\textsuperscript{15} This observation led us to hypothesize that inter-individual differences in the risk of developing IHD may be associated with variation in these 5’ genotypes. A test of this hypothesis using a large population-based sample ascertained by the Copenhagen City Heart Study (CCHS) established an increased hazard of developing IHD in females carrying the $A_{560}T_{832}/T_{560}T_{832}$ genotype.\textsuperscript{18} Although quantitatively not very large, the estimated hazard remained statistically significant after the effects of dyslipidemia, other established risk factors and the genotypes defined by the $\varepsilon_2$, $\varepsilon_3$ and $\varepsilon_4$ alleles were considered in the prediction model.

Assumptions that are implicit in the application of the Cox proportional hazards model that was used to evaluate IHD risk in our previous study may limit the medical utility of the risk information obtained. Most important, employing a single model assumes that the expected relationship between disease status and variation in risk factor traits is the same for all individuals in the sample under study. The complex multifactorial nature of the pathobiology of IHD endpoints renders this assumption untenable. Development of IHD is an emergent property of interactions of many susceptibility genes and many environmental factors. There is no evidence that any of these factors act as independent agents whose phenotypic effects are additive, exchangeable with each other and the same for each individual in the population at large. Furthermore, because the combined number of interacting genes and environments is large, every incident IHD case cannot have experienced effects of the same combination of genetic variants and exposures to high-risk environments. These considerations suggest that the goal of a model building strategy is not which combination of risk factors in a single linear model predicts risk for every individual at risk; but how many models are necessary to best predict risk of disease? We explore in this paper whether variation in 5’ genotypes of the $APOE$
gene which contribute to predicting IHD when a single prediction model is employed makes a larger contribution to prediction when multiple models are employed to estimate risk of disease in sub-samples of individuals defined by particular combinations of risk factor values.

The augmented Patient Rule-Induction Method (PRIM) developed by Dyson et al\textsuperscript{19-21} is a novel model building strategy for evaluating risk which acknowledges the etiological heterogeneity of the disease. It is designed to identify combinations of risk factor values that characterize mutually exclusive subgroups of individuals that differ in average risk as measured by the cumulative incidence of the disease of interest. This analytical strategy addresses the question of which combination of alterations in which combination of risk factors best predicts the disease of interest in which subset of individuals of the population from which the sample under study was drawn.

In this paper we employ a stepwise application of PRIM to test the hypothesis that information about 5' \textit{APOE} genotypes significantly improves the prediction of IHD in particular sub-samples of individuals characterized by selected subsets of values of the traditional risk factors. We identified a sub-sample of 764 elderly females with hypertriglyceridemia who had either a history of smoking, a history of hypertension or a history of both in which the $A_{560}T_{832}/A_{560}T_{832}$ and $A_{560}T_{832}/A_{560}G_{832}$ 5' \textit{APOE} genotypes had a significantly higher cumulative incidence of IHD (172/1,000) compared to the cumulative incidence of 70/1,000 in the total sample of 3,686 females. The implications for the added value of this genetic information in the practice of medicine are discussed.

**METHODS**

**Study Participants**
The CCHS is a prospective study of the Danish population at large aged 20 years or older at entry into the study.\textsuperscript{22-24} The initial survey was carried out between 1976 and 1978. A follow-up survey was performed between 1991 and 1994. This follow-up survey serves as the baseline survey for the study reported here. Altogether, 16,563 individuals were invited to this survey, 10,135 participated (response rate 61\%) and 9,259 gave blood for DNA extraction. A sub-sample of middle-aged and elderly individuals who were at least 45 years old and free of IHD when they were seen in the follow-up survey was selected for our study. Clinical data and genotype information on the four \textit{APOE} SNPs considered in this study were available on 3,686 females and 2,772 males who satisfied the selection criteria. Informed consent was obtained from all participants. More than 99\% were Europeans of Danish descent. The study was approved by the Danish Ethics Committee for the City of Copenhagen and Frederiksberg (No. 100.2039/91).

\textbf{Variable Definitions}

All 6,458 participants of the CCHS ascertained for this study were free of IHD at baseline. IHD was evaluated during the period from baseline until December 31, 1999. Information to establish the diagnoses of IHD (World Health Organization International Classification of Diseases, 8\textsuperscript{th} edition, codes 410 to 414; 10\textsuperscript{th} edition, codes I20 to I25) was gathered from the Danish National Hospital Discharge Register, the Danish National Register of Causes of Death and medical records of general practitioners and hospitals. During the follow-up 576 individuals (257 of 3,686 females, 7.0 \% and 319 of 2,772 males, 11.5 \%) of the 6,458 participants were diagnosed as having developed IHD. The observed average follow-up time
until an IHD event or the censure date of December 31, 1999 was 6.5 years (range = 0.01 to 8.2). The total exposure to risk of developing IHD was 39,648 person-years.

Baseline plasma HDL-C, TG and T-C concentrations were measured by standard enzymatic assays (Boehringer Mannheim, GmbH Diagnostics, Mannheim, Germany) at the Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen. Recommendations of the National Cholesterol Education Program Expert Panel, National Institute of Health, USA, were used to define dyslipidemic subgroups (National Cholesterol Education Program National Heart, Lung, and Blood Institute 2002). Dyslipidemia was diagnosed when an individual’s plasma T-C concentration was ≥ 200 mg/dl (5.18 mmol/L), TG was ≥150 mg/dl (1.60 mmol/L) or HDL-C was < 40 mg/dl (1.04 mmol/L).

The definitions of smoking habit, glucose metabolism and blood pressure used in our study are described in a previous report of the CCHS. Briefly, each factor was dichotomized to define a high-risk group; a history of smoking (current smoker at any exam), a history of diabetes (self reported disease, use of insulin, use of oral hypoglycemic drugs and/or non-fasting plasma glucose ≥11.1 mmol/L at any exam) or a history of hypertension (systolic blood pressure ≥140 mmHg and/or diastolic blood pressure ≥90 mmHg and/or use of antihypertensive drugs at any exam).

The four SNPs in APOE were genotyped by PCR and restriction enzyme digestion as previously described by Frikke-Schmidt et al, Frikke-Schmidt et al and Hixson and Vernier.

Statistical Analyses

The chi-square statistic was used to test homogeneity of the relative frequencies of dichotomous risk factor traits between genders.
The PRIM algorithm: The PRIM algorithm was first introduced by Friedman and Fisher and augmented by Dyson et al for application to genetic studies. PRIM enables one to partition the total sample of individuals into multiple sub-samples, each defined by a subset of predictor variables. The augmented PRIM algorithm selects the subset of statistically significant terms, defined by values of one or more predictor variables, that maximize the mean outcome of a response variable of interest in a selected sub-sample of individuals. The selected sub-sample satisfies the minimum size criterion, denoted by the support parameter, $\beta$. $\beta$ defines the minimum proportion of the sample of individuals who have not been previously assigned to a sub-sample as a consequence of applying the PRIM algorithm who must be included in establishing a new sub-sample. The optimum value of $\beta$ for a particular application of PRIM is chosen according to the algorithm described in Dyson et al. Here we applied the PRIM algorithm using terms for the peeling and pasting processes defined by combinations of values of two predictor variables. The hypergeometric distribution was employed to derive the theoretical null distribution used to test whether the cumulative incidence of IHD associated with a particular peeling or pasting term was statistically significant. The multiple hypothesis tests conducted during the execution of PRIM each uses a nominal significance threshold of 0.023, which corresponds to an experiment-wise significance level of 0.05. Multiple mutually exclusive sub-samples of the total sample may be produced. Each sub-sample of the original sample includes individuals with the same values for a subset of predictor variables. The individuals that are not included in any of the sub-samples produced by the peeling and pasting processes are assigned to a remainder sub-sample.

Stepwise application of PRIM algorithm: The PRIM algorithm was first employed to identify sub-samples of individuals that are each characterized by different subsets of values for
age, the three plasma measures of dyslipidemia (T-C, HDL-C and TG), the three high-risk groups defined by three established risk factors (diabetes, smoking and hypertension) and the combined $\{e_{22}, e_{32}\}, \{e_{33}\}$ and $\{e_{42}, e_{43}, e_{44}\}$ APOE genotype groups. We carried out a second application of the PRIM algorithm in each of the sub-samples to test whether variation in the 5’ APOE genotypes improved the prediction of the cumulative incidence of IHD in any of the sub-samples. The improvement in prediction of the cumulative incidence of IHD using the 5’ APOE genotypes in one of the first step sub-samples and not in another is interpreted as evidence for non-additive relationships between the effects measured by the established risk factors that characterize the sub-samples and the genetic effects marked by the 5’ APOE genotypes. The values of the subset of risk factors that PRIM selects to define a sub-sample are expected to vary from sub-sample to sub-sample because of the heterogeneity of the relationships between the risk of IHD and etiological causes among individuals in a representative sample of the population at large.

RESULTS

A Descriptive Summary of the Sample

Summary statistics that describe the female and male samples are given in Table 1. Using a 5% test criterion, none of the single SNP genotypes deviated significantly from the Hardy-Weinberg expectations in either of the genders and the relative allele frequencies were not significantly different between genders. Relative frequencies of the four 5’ APOE genotypes and the three groups of traditional APOE genotypes defined by the $e_2$, $e_3$ and $e_4$ alleles are given in Table 2. There was also no statistically significant evidence for heterogeneity of relative frequencies of these two SNP genotypes between genders.
The widely acknowledged evidence that the natural history of IHD is gender–specific, combined with the statistically significant differences in the gender-specific cumulative incidence of IHD and the statistically significant differences in gender-specific frequency distributions of the outcomes of the proposed predictor variables (Table 1), justifies carrying out model building and hypothesis testing strategies separately in the female and male samples.

**Step One PRIM Analysis**

For females, the optimum value of the support parameter was $\beta = 0.20$. The selected risk factors and their values that characterized each of two statistically significant, mutually exclusive sub-samples (FS 1 and FS 2) and a remainder sub-sample (FS 3) based on the information obtained from the application of PRIM are given in Table 3. The estimated cumulative incidences of IHD in these sub-samples ranged from 33 cases to 139 cases/1,000 females at risk. The highest incidence of IHD in the sub-sample (FS 1) of elderly females is almost twice as large as the estimate in the total sample of females (139 versus 70 cases/1,000). Everyone in sub-sample FS 1 ($n=764$, 20.7 % of the total sample) was 65 years or older at baseline, had hypertriglyceridemia ($\text{TG} \geq 150\text{mg/dl}$) and either had a history of smoking, a history of having had hypertension or a history of both. The second sub-sample (FS 2, $n=839$, 22.8 % of the total sample), with the second highest estimate of cumulative incidence of IHD (99 cases/1,000), consisted of three subgroups. The first consisted of 799 elderly females with low TG (<150mg/dl) and a history of hypertension. The second consisted of six elderly females with low TG (<150mg/dl) who had a history of smoking and a history of diabetes, but no history of hypertension. The third subgroup of this second sub-sample consisted of 34 middle aged females (45 to 64 years) who had a history of smoking and a history of diabetes. The estimated
cumulative incidence of IHD in the remainder sub-sample (FS 3) of 2,083 individuals (56.5 %) who were not assigned to either of the two high-risk sub-samples was 50% smaller than the estimate in the total female sample (33 cases versus 70 cases/1,000).

For males the optimum value of the support parameter was $\beta = 0.05$. The selected risk factors and their values that characterized each of the three statistically significant, mutually exclusive sub-samples (MS 1, MS 2 and MS 3) and a remainder sub-sample (MS 4) based on the information obtained from the application of PRIM are given in Table 4. The estimated cumulative incidence of IHD in these four sub-samples ranged from 65 to 209 cases/1,000. On average, the estimates in the first three sub-samples are three times larger (mean incidence = 194 cases/1,000) than the estimated cumulative incidence of IHD in the remainder sub-sample (65 cases/1,000).

The majority of males assigned to each of the three high-risk sub-samples were older than 65 years (985 of 1,074 males). Sub-sample MS 1 consisted of 234 elderly males (8.4 % of the total sample) with low plasma HDL-C concentration. Sub-sample MS 2 consisted of 741 elderly males (26.7 % of the total sample) with HDL-C $\geq$ 40 mg/dl and a history of hypertension. Sub-sample MS 3 (n=99, 3.6 % of the total sample) includes two subgroups. The first subgroup consisted of 10 elderly males with HDL-C $\geq$ 40 mg/dl who had a history of diabetes, were one of the four genotypes $\epsilon_{33}$, $\epsilon_{42}$, $\epsilon_{43}$ or $\epsilon_{44}$ and did not have a history of hypertension. The second subgroup included 89 middle-aged males (45 to 64 years) who had a history of diabetes and who had one of the four genotypes $\epsilon_{33}$, $\epsilon_{42}$, $\epsilon_{43}$ or $\epsilon_{44}$. The MS 4 sub-sample of 1,698 individuals (61.3 % of total male sample) included those who were not assigned to MS 1, MS 2 or MS 3.

**Step Two PRIM Analysis**
We next applied PRIM to select combinations of 5’ APOE genotypes that identify statistically significant genetic subgroups of each of the two sub-samples and the remainder sub-sample in females and each of the three sub-samples and the remainder sub-sample in males that were identified in the application of PRIM to the predictor variables considered in step one. A statistically significant high-risk genetic subgroup of the first female sub-sample (FS 1, Table 5) was identified. It included elderly females with hypertriglyceridemia who had a history of smoking, a history of having had hypertension or a history of both and were carriers of one of the two 5’ genotypes, A560T832/T560T832 or A560T832/A560G832. The cumulative incidence of IHD in this genetic subgroup of 308 females is 172 cases/1,000 compared to 116 cases/1,000 in the subgroup with the A560T832/A560T832 or a genotype in the OTHERS group of genotypes. There was no statistically significant evidence that 5’ genotypes improved the prediction of IHD in the FS 2 and FS 3 sub-samples (Table 5).

We did not detect a statistically significant genetic subgroup of any of the three sub-samples or the remainder sub-sample identified in the first step PRIM analysis of the male sample.

**DISCUSSION**

**A Prevailing Prevention Paradigm in Cardiology**

A major goal of medical research is to identify sub-populations of individuals at increased risk of disease in order to efficiently allocate limited resources in a way that will maximize the reduction of individual suffering. Cardiovascular research has a long history of establishing the information collected on individuals that is useful in medical practice to identify those who are at increased risk of developing clinical symptoms of disease\(^8,28-30\) and those who
would benefit from various pharmacological interventions to prevent, or even reverse, the progression of atherosclerosis. However, for most common chronic diseases having a complex multifactorial etiology (including IHD) only a fraction of individuals who develop disease are identified by the established risk factors. It is widely accepted that the ability to accurately predict those at risk may be improved significantly by considering genomic information.

The immensity of the amount of genomic variation that may be considered and the complexity of how such variation may influence the initiation, progression and/or severity of IHD makes identifying relevant variations a daunting task. Etiological heterogeneity among those with disease, and the primary role that interactions between genetic elements, and between genetic elements and environmental exposures, play in determining risk, are biological realities that make the search for genetic variations that may have predictive value a challenge that has not been adequately addressed. The applications of traditional multivariable linear regression to case-control data or Cox proportional hazards modeling of longitudinal data to identify predictors to evaluate the contribution of interactions of those predictors to risk of disease assume etiological homogeneity among all individuals at risk and no correlations between predictor variables. These assumptions are untenable when analyzing data from observational studies designed to represent the population at large. The consequence of applications of such single-model approaches has been that genetic variations have made very small improvements in the prediction of common complex disease endpoints such as IHD. In the study reported here we employed a complementary analytical strategy designed to take etiological heterogeneity and non-additivity of predictor variable effects into consideration when evaluating the utility of genetic variation in the identification of individuals at increased risk of IHD.
Biological Plausibility of the Risk Information Provided by PRIM Analyses

Using the traditional risk factors for atherosclerosis and three groups of traditional APOE genotypes defined by the ε2, ε3 and ε4 alleles we found two statistically significant, mutually exclusive high-risk sub-samples in females and three sub-samples in males defined by different combinations of predictors (Tables 3 and 4). These sub-samples are consistent with common knowledge that all of the high-risk values for all of the risk factors, which are involved in determining risk of IHD in the population at large, are not expected to be present in all individuals at risk.8 Our finding that different combinations of risk factors and their values are associated with different high-risk sub-samples in females and males is consistent with well established knowledge that genders differ in the natural history of the development of IHD.36-38

Variations in the values of a number of established risk factors, which may be internal (e.g. plasma concentrations of T-C, TG and HDL-C) or external (e.g. exposure to tobacco smoking) to an individual, are hypothesized to combine with variations in the products of hundreds of genes to determine inter-individual differences in initiation, progression and/or severity of atherosclerosis; and these relationships are dynamic over the life cycle.8 The APOE gene is one of the few extensively studied candidate genes that has been repeatedly implicated in contributing to the determination of risk of IHD.7,17 In most studies the contribution of the traditional genotypes defined by the ε2, ε3 and ε4 alleles add a small improvement in prediction of risk of IHD beyond the traditional risk factors over a wide range of environmental and genetic backgrounds. We found in our previous study18 of the participants of the CCHS that the traditional genotypes defined by the ε2, ε3 and ε4 alleles do not statistically significantly improve the prediction of IHD in either gender. However, in the study reported here using the same
CCHS sample, when multiple models were considered the traditional APOE genotypes improved the prediction of IHD in males in two of the four sub-samples identified by PRIM (Table 4).

In our earlier studies of the CCHS sample considered here we established that particular two-SNP 5′ genotypes influenced variability in plasma measures of lipid metabolism and variation in risk of IHD beyond that contributed by the traditional risk factors and the traditional APOE genotypes, particularly in females.15,18 This inference is supported by studies that report that there are estrogen response elements marked by the 5′ SNPs that modulate the response of the gene to estrogen.39 Our current study further suggests that the small contribution of selected 5′ genotypes to improving risk prediction in the total sample of females is attributable to only one of the three sub-samples obtained from the application of PRIM (Table 6), as evidenced by a considerably larger statistically significant hazard ratio. That this validated effect is greatest in a sample of individuals with high TG suggests that the 5′ genotypes mark genetic elements that have pleiotropic effects on unmeasured or unknown intermediate traits that influence risk of IHD.

Consistent with the conceptual/theoretical biological modeling of the role of genetic variation in determining the risk of having a complex multifactorial disease,8,40-43 we find that the added value of genetic information for prediction depends on the genetic background and environmental contexts which are characterized by the sub-samples identified by PRIM. This result supports the argument that traditional statistical approaches to evaluating genetic variation that estimate only marginal, context independent effects are inconsistent with the ubiquity of biological interactions that determine the pathobiology of IHD.8,11,19,20,44-47 We next turn to a discussion of the implications of using genetic information for evaluating risk that come from our study of IHD.
Utility of 5' Genotypes of the APOE in Medical Risk Stratification

Any medical act requires the establishment of relevant information and a rationale for using it in making a decision. A traditional act in everyday medical practice is one in which the clinician, after logically considering all information available, including signs, symptoms and laboratory results, assigns an individual under investigation to either a high-risk group or to a low-risk group.48,49 The stepwise PRIM algorithm used here is consistent with, and is complementary to, such a clinical decision making strategy. In the samples under study, females can first be stratified into three, and males into four, sub-samples that differ in their cumulative incidence of IHD. Because of the historical precedence established by clinical practice, these strata would be considered first in the evaluation of risk followed by consideration of the inclusion of genetic information. In our particular study we found that one high-risk sub-sample of females (FS 1) can be further stratified into two subgroups based on their 5' APOE genotype. This is the kind of context dependent genetic information that is expected to improve the traditional act of assignment of risk of IHD.

Several kinds of information may be used in making a decision as to whether the statistically significant stratifications based on the 5' APOE genotypes should be incorporated into the medical act of making a decision about whether a female is at high-risk of IHD; 1) age-specific propensity, 2) sensitivity and specificity of the stratification into high- and low-risk groups and 3) the positive predictive value of the decision based on the stratification strategy.

The age-specific propensities for developing IHD for individuals in a particular stratum can be derived from the probability of surviving free of IHD to a particular age. In Figure 1 we present survival curves for elderly females who are included in sub-samples FS 1, FS 2 and FS 3.
Those in FS 1 who have high TG (≥ 150mg/dl) and who either have a history of smoking, a history of having had hypertension or a history of both have approximately a 0.50 probability of developing IHD by 90 years of age compared to 0.20 among the elderly females in the low-risk sub-population (FS 3). If the females in the FS 1 high-risk sub-population carry either the \(A_{560}T_{832}/T_{560}T_{832}\) or the \(A_{560}T_{832}/A_{560}G_{832}\) two-SNP \(5' APOE\) genotype (genetic subgroup FS 1-1), their propensity for developing IHD by 90 years of age is approximately 0.65 compared to less than 0.40 in the females who do not carry either of the proposed high-risk genotypes (genetic subgroup FS 1-2). The estimated age-dependent propensity of IHD in elderly females in the FS 2 sub-sample falls between the propensities observed for those elderly females in the FS 1 high-risk and those in the FS 3 low-risk sub-samples. The observed variations in the gender, age and genotype dependent propensities of IHD among these sub-samples serves as a compelling rationale for embracing a risk stratification algorithm that includes genotype information in making clinical decisions about the prevention of IHD.

Sensitivity and specificity are properties of a risk stratification strategy that need to be taken into account when selecting a risk stratification algorithm in medical practice.\(^{50,51}\) From a clinical decision making point of view, it is also important to know the probability that individuals in the high-risk stratum actually are at increased risk. The sensitivity and specificity do not give this information. Instead the interpretation of the risk analysis needs to consider the positive predictive value (+PV). The +PV is related to the sensitivity and specificity of a risk stratification algorithm and the prevalence of the disease of interest in the population from which the individuals are coming from through a mathematical formula derived from an application of the Bayes’ theorem of conditional probabilities.\(^{50}\)
The sensitivity and specificity estimates for an algorithm that assigns females in the FS 1 and FS 2 to a high-risk stratum and those in the FS 3 sub-sample into a low-risk stratum are modest, 0.735 and 0.588, respectively (Table 7). An estimate of +PV for the high-risk stratum is 0.118, which is approximately 1.5 times higher than cumulative incidence of IHD of 0.070 (70 cases/1,000 individuals at risk) in the overall sample of females. If the high-risk stratum includes the FS 1 sub-sample of females only, sensitivity of the stratification algorithm decreases but its specificity increases to 0.808 and the estimate of the +PV, 0.139, is slightly larger than the estimate for the stratum that includes both the FS 1 and the FS 2 sub-samples.

If the high-risk stratum includes only the FS 1-1 genetic subgroup of elderly hypertriglyceridemic females who either have a history of smoking, a history of hypertension or a history of both, and who also carry either the $A_{560}T_{832}/T_{560}T_{832}$ or the $A_{560}T_{832}/A_{560}G_{832}$ 5' genotype in the $APOE$ promoter region, the sensitivity is low (0.206), but the specificity is highest (0.926). The estimated +PV for this FS 1-1 high-risk stratum, 0.172, is higher than the estimate for the FS 1 high-risk stratum ignoring genotype information, 0.139, and approximately 2.5 times higher than the estimate of 0.070 in overall female sample when assignments of individuals to sub-samples are ignored (Table 7).

CONCLUSIONS

Clinicians and patients need to know which trait(s) and which interventions work best in particular sub-samples of individuals.\textsuperscript{8,52} The evaluation of the added value of a genetic variation using the traditional “one model describes all” approach provides information that has limited utility in clinical practice.\textsuperscript{8,41,43} The stepwise PRIM algorithm we employed to evaluate the improvement in IHD prediction attributable to information about 5' $APOE$ genotypes
acknowledges the biological reality of etiological heterogeneity and non-additive, context
dependent genetic effects that are consistent with what we know about the etiology of IHD. We
propose that in Denmark this genetic information has added value when predicting an
individual’s propensity to develop IHD only in a subpopulation of elderly females who have
hypertriglyceridemia and either a history of smoking, a history of hypertension or a history of
both. Intervention studies in Denmark are now needed to determine whether this context
dependent genetic information improves our ability to deliver the right treatment to the right
patients at the right time, which is the primary goal of a modern health care system.

**ACKNOWLEDGEMENTS**: We wish to thank Ken Weiss for his devotion to carrying out the
statistical analyses and Christine Lusk, Tom Rea and Kim Zerba for their critical review and
suggestions for improving earlier drafts of this manuscript.

**FUNDING SOURCES**: This work was supported by National Institutes of Health grants
GM065509 and HL072905.

**DISCLOSURES**: None.
REFERENCES


3. Nickerson DA. Department of Genome Sciences, University of Washington; Nickerson Group, MDECODE; SNP data; http://droog.gs.washington.edu/mdecode.


18. Stengård JH, Frikke-Schmidt R, Tybjærg-Hansen A, Nordestgaard BG, Sing CF. Variation in 5’ promoter region of the APOE gene contributes to predicting ischemic heart disease...


**Table 1:** A Description of the CCHS Samples.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Females (N=3,686)</th>
<th>Males (N=2,772)</th>
<th>Prob. ( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative Incidence of IHD/1,000 individuals at risk</td>
<td>70</td>
<td>115</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Average Follow Up Time (years)</td>
<td>6.3(^a)</td>
<td>5.9(^a)</td>
<td></td>
</tr>
<tr>
<td>Predictors of IHD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age over 65 years (%)</td>
<td>51</td>
<td>43</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Blood Measures of Dyslipidemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDL-C &lt; 40 mg/dl (%)</td>
<td>6</td>
<td>20</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>TG &gt; 150 mg/dl (%)</td>
<td>41</td>
<td>52</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>T-C &gt; 200 mg/dl (%)</td>
<td>89</td>
<td>80</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Other Established Risk Factor Characters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking (%)</td>
<td>57</td>
<td>70</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>4</td>
<td>8</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Hypertension (%)</td>
<td>62</td>
<td>69</td>
<td>0.001</td>
</tr>
</tbody>
</table>

\( ^a \) The range of follow up time was 0.01 – 8.2 years in both genders

**Table 2:** Relative Genotype Frequencies in the CCHS Samples.

<table>
<thead>
<tr>
<th>Genotype Groups</th>
<th>Relative Frequency (N)</th>
<th>Females</th>
<th>Males</th>
<th>Prob. ( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' Promoter Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_{560T} T_{832} / A_{560G} G_{832} )</td>
<td>0.35 (1,290)</td>
<td>0.36 (1,010)</td>
<td>0.438</td>
<td></td>
</tr>
<tr>
<td>( A_{560T} T_{832} / A_{560T} T_{832} )</td>
<td>0.16 (583)</td>
<td>0.15 (416)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_{560T} T_{832} / T_{560T} T_{832} )</td>
<td>0.06 (213)</td>
<td>0.06 (174)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTHERS</td>
<td>0.43 (1,600)</td>
<td>0.43 (1,172)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 4 Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( { E_{33} } )</td>
<td>0.56 (2,066)</td>
<td>0.58 (1,596)</td>
<td>0.335</td>
<td></td>
</tr>
<tr>
<td>( { E_{22}, E_{32} } )</td>
<td>0.13 (475)</td>
<td>0.13 (362)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( { E_{42}, E_{43}, E_{44} } )</td>
<td>0.31 (1,145)</td>
<td>0.29 (814)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table 3:** The Sub-Samples of Females (n=3,686, I=70)† Defined by the Step One PRIM Analysis.

<table>
<thead>
<tr>
<th>Sub-sample (n, %)</th>
<th>Cumulative Incidence (95% Confidence Interval)</th>
<th>Description*</th>
<th>Subgroup (n, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS 1 (764, 20.7%)</td>
<td>I=139 (119, 161)</td>
<td>(Age ≥ 65 and TG ≥ 150) and (SMK and/or HYT)</td>
<td>(764, 20.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Age ≥ 65 and TG &lt; 150 and HYT)</td>
<td>(799, 21.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Age ≥ 65 and TG &lt; 150 and SMK and No HYT and DIAB)</td>
<td>(6, 0.2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Age 45-64 and SMK and DIAB)</td>
<td>(34, 0.9%)</td>
</tr>
<tr>
<td>FS 2 (839, 22.8%)</td>
<td>I=99 (82, 128)</td>
<td>(Age ≥ 65 and No SMK and No HYT)</td>
<td>(177, 4.8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Age ≥ 65 and No SMK and No HYT)</td>
<td>(157, 4.3%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Age 45-64 and (No SMK and/or No DIAB)</td>
<td>(1,749, 47.4%)</td>
</tr>
</tbody>
</table>

† I=Number of cases / 1000 individuals at risk

* HYT = History of hypertension.
DIAB = History of diabetes.
SMK = History of smoking.
Table 4: The Sub-Samples of Males (n=2,772, I=115)† Defined by the Step One PRIM Analysis.

<table>
<thead>
<tr>
<th>Sub-sample (n, %)</th>
<th>Cumulative Incidence (95% Confidence Interval)</th>
<th>Description*</th>
<th>Subgroup (n, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS 1 (234, 8.4%)</td>
<td>I=209 (185, 273)</td>
<td>(Age ≥ 65 and HDL-C &lt; 40)</td>
<td>(234, 8.4%)</td>
</tr>
<tr>
<td>MS 2 (741, 26.7%)</td>
<td>I=189 (171, 248)</td>
<td>(Age ≥ 65 and HDL-C ≥ 40 and HYT)</td>
<td>(741, 26.7%)</td>
</tr>
<tr>
<td>MS 3 (99, 3.6%)</td>
<td>I=192 (123, 296)</td>
<td>(Age ≥ 65 and HDL-C ≥ 40 and DIAB and (E33 or E42 or E43 or E44)) and No HYT</td>
<td>(10, 0.4%)</td>
</tr>
<tr>
<td>MS 4 (1,698, 61.3%)</td>
<td>I=65 (54, 77)</td>
<td>(Age 45-64 and DIAB and (E33 or E42 or E43 or E44))</td>
<td>(213, 7.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Age ≥ 65 and HDL-C ≥ 40 and No HYT) and (No DIAB and/or (E22 or E32))</td>
<td>(1,485, 53.6%)</td>
</tr>
</tbody>
</table>

† I=Number of cases / 1000 individuals at risk.
* HYT = History of hypertension.
DIAB = History of diabetes.
SMK = History of smoking.
TG = Triglycerides (mg/dl).
HDL-C = HDL cholesterol (mg/dl).
Table 5: The Genetic Subgroups Defined by the Step Two PRIM Analysis in Females.

<table>
<thead>
<tr>
<th>Genetic Subgroup</th>
<th>Sub-Sample†</th>
<th>Overall</th>
<th>FS 1</th>
<th>FS 2‡</th>
<th>FS 3‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A_{560}T_{832}/T_{560}T_{832}) or (A_{560}T_{832}/A_{560}G_{832})</td>
<td>N=1,503, I=77</td>
<td>N=308, I=172 (134, 217)§ (FS 1-1)</td>
<td>N=348, I=98</td>
<td>N=847, I=34</td>
<td></td>
</tr>
<tr>
<td>(A_{560}T_{832}/A_{560}T_{832}) or OTHERS*</td>
<td>N=2,183, I=65</td>
<td>N=456, I=116 (82, 143) (FS 1-2)</td>
<td>N=491, I=100</td>
<td>N=1,236, I=32</td>
<td></td>
</tr>
</tbody>
</table>

Box indicates the significant genetic sub-group from sub-sample S 1.

† I=Number of cases / 1,000 individuals at risk.

§ I (confidence interval).

‡ A confidence interval cannot be calculated because PRIM did not define this contrast in these sub-samples (See Dyson et al19 for more details).

* OTHERS is the group of 5′ genotypes which does not include \(A_{560}T_{832}/T_{560}T_{832}\), \(A_{560}T_{832}/A_{560}G_{832}\) or \(A_{560}T_{832}/A_{560}T_{832}\).

Table 6: Hazard Ratio (p-value) for Genetic Subgroups Defined by the Step Two PRIM Analysis in Females.

<table>
<thead>
<tr>
<th>Genetic Subgroup</th>
<th>Overal</th>
<th>FS 1</th>
<th>FS 2‡</th>
<th>FS 3‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hazard Ratio (ignoring established risk factors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_{560}T_{832}/T_{560}T_{832}) or (A_{560}T_{832}/A_{560}G_{832})</td>
<td>1.242 (0.085)</td>
<td>1.524 (0.031) (FS 1-1)</td>
<td>0.998 (0.994)</td>
<td>1.164 (0.538)</td>
</tr>
<tr>
<td>(A_{560}T_{832}/A_{560}T_{832}) or OTHERS*</td>
<td>1.0</td>
<td>1.0 (FS 1-2)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Hazard Ratio (conditional on fitting established risk factors)

| Hazard Ratio (conditional on fitting established risk factors) | | | | |
| \(A_{560}T_{832}/T_{560}T_{832}\) or \(A_{560}T_{832}/A_{560}G_{832}\) | 1.310 (0.038) | 1.517 (0.037) FS 1-1 | 1.326 (0.243) | 1.066 (0.801) |
| \(A_{560}T_{832}/A_{560}T_{832}\) or OTHERS* | 1.0 | 1.0 FS 1-2 | 1.0 | 1.0 |

* OTHERS is the group of 5′ genotypes which does not include \(A_{560}T_{832}/T_{560}T_{832}\), \(A_{560}T_{832}/A_{560}G_{832}\) or \(A_{560}T_{832}/A_{560}T_{832}\).
**Table 7:** Epidemiological Summaries of the Clinical Predictive Value of the PRIM-Defined Sub-Samples.

<table>
<thead>
<tr>
<th>Sub-sample</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>+PV^A</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-Risk</td>
<td>Low-Risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FS 2 and FS 1</td>
<td>FS 3</td>
<td>0.735</td>
<td>0.118</td>
</tr>
<tr>
<td>FS 1</td>
<td>FS 3 and FS 2</td>
<td>0.412</td>
<td>0.139</td>
</tr>
<tr>
<td>FS 1-1</td>
<td>FS 3 and FS 2 and FS 1-2</td>
<td>0.206</td>
<td>0.172</td>
</tr>
</tbody>
</table>

**Marginal Estimates**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTT/ATAG</td>
<td>ATAT/OTHERS</td>
<td>0.451</td>
<td>0.077</td>
</tr>
<tr>
<td>Smoke†</td>
<td>No smoke</td>
<td>0.642</td>
<td>0.079</td>
</tr>
<tr>
<td>Hypertension‡‡</td>
<td>No hypertension</td>
<td>0.763</td>
<td>0.085</td>
</tr>
<tr>
<td>TG ≥ 150</td>
<td>TG &lt; 150</td>
<td>0.529</td>
<td>0.090</td>
</tr>
<tr>
<td>Over 65</td>
<td>45 to 65</td>
<td>0.770</td>
<td>0.104</td>
</tr>
</tbody>
</table>

^A +PV (positive predictive value) reflects an expected proportion of IHD cases who are assigned to the related high-risk group. The estimate of +PV is related to sensitivity and specificity of a risk stratification algorithm that employs sub-samples as a tool to identify individuals who are at increased risk, and to the prevalence of the proposed high-risk sub-sample in the population of interest, through a mathematical formula that is derived from the application of the Bayes’ theorem of conditional probabilities (Fletcher and Fletcher^50).  

† History of smoking
‡‡ History of hypertension

**FIGURE LEGENDS:**

**Figure 1:** Age dependent propensity of being IHD free, separately for the three sub-samples and for the two genetic subgroups, of sub-sample 1, in elderly females.
Context Dependent Associations between Variation in Risk of Ischemic Heart Disease and Variation in the 5' Promoter Region of the Apolipoprotein E Gene in Danish Females
Jari H. Stengård, Greg Dyson, Ruth Frikke-Schmidt, Anne Tybjærg-Hansen, Borge G. Nordestgaard and Charles F. Sing

*Circ Cardiovasc Genet.* published online December 3, 2009;
*Circulation: Cardiovascular Genetics* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/early/2009/12/03/CIRCGENETICS.109.862748

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation: Cardiovascular Genetics* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Circulation: Cardiovascular Genetics* is online at:
http://circgenetics.ahajournals.org//subscriptions/