ORIGINAL INVESTIGATION

Jari H. Stengård · Kenneth M. Weiss · Charles F. Sing

An ecological study of association between coronary heart disease mortality rates in men and the relative frequencies of common allelic variations in the gene coding for apolipoprotein E

Abstract Three common alleles, ε_2, ε_3, and ε_4, of the gene coding for apolipoprotein E (apoE) have been identified as predictors of interindividual variation in measures of lipid and lipoprotein metabolism, and ultimately risk of coronary heart disease (CHD), within many populations. Here we evaluated the utility of the geographic distribution of these alleles for prediction of interpopulation variation in average level of serum total cholesterol and other traditional risk factors, and CHD mortality rate. We employed published estimates of the relative frequencies of the three common *apoE* alleles, average levels of risk factors such as serum total cholesterol, systolic and diastolic blood pressure, body mass index, smoking prevalence and CHD mortality rate for nine population-based samples of middle-aged males studied by the international WHO MONICA Project. There was approximately a 10-fold difference between the highest and lowest CHD mortality rate. Of the traditional risk factors, variation in the average level of serum total cholesterol was the best predictor (approximately 33%) of the observed interpopulation variation in estimates of CHD mortality rate ($Pr=0.10$). Variation in the relative frequency of the ε_4 allele predicted approximately 50% of interpopulation variation in average serum total cholesterol level ($Pr=0.02$) and 75% of the variation in CHD mortality rate ($Pr=0.002$) when information about variation in the other risk factors and the ε_2 and ε_3 alleles is ignored. Furthermore, variation in the relative frequency of the ε_4 allele predicted approximately 40% of the variation in CHD mortali-

ty rate ($Pr=0.02$) after considering the contribution of variation in average serum total cholesterol level. Average serum total cholesterol level was estimated to increase by 0.114 mmol/l (4.405 mg/dl), and CHD mortality rate by 24.5/100 000, for an increase of 0.01 in the relative frequency of the ε_4 allele. The predictive utility of the ε_2 and ε_3 alleles was considerably less than that of the ε_4 allele. For the sample of populations considered, the geographic distribution of the *apoE* alleles can be a statistically significant predictor of interpopulation variation in both the average serum total cholesterol level and CHD mortality rate. In particular, the ε_4 allele may confer valuable ecological risk information.

Introduction

Coronary heart disease (CHD) is a worldwide health problem. Each year it claims approximately 2.4 million lives, which is about one-quarter of all annual deaths reported to WHO (WHO 1989). The health burden of CHD is not, however, uniformly distributed among different populations (Keys 1970; Tunstall-Pedoe et al. 1994; Zevallos et al. 1992). In populations where the disease takes its greatest toll, such as the province of North Karelia, Finland, and Glasgow, United Kingdom, CHD morbidity and mortality rates are about ten times higher than the rates in populations that are least prone to the CHD, such as Beijing, China, and Toulouse, France (Tunstall-Pedoe et al. 1994). Reasons for the observed regional differences in CHD have been under intensive study for decades.

There is accumulating evidence that variation in the levels of average serum total cholesterol and systolic blood pressure and smoking prevalence are associated with variation in CHD rates among populations (Fidanza and Kromhout 1993; Keys 1970; Zevallos et al. 1992). A fraction of the observed variation in average serum total cholesterol level is ascribed to interpopulation differences in saturated fatty acid consumption. Recent observations that the relative ranking of CHD rates among populations has per-

J.H. Stengård (✉)
National Public Health Institute, Department of Epidemiology and Health Promotion, Mannerheimintie 166,
FIN-00300 Helsinki, Finland
Tel.: +358-9-4744 8641, Fax: +358-9-4744 8338

K.M. Weiss
Pennsylvania State University, Department of Anthropology, University Park, PA 16802, USA

J.H. Stengård · C.F. Sing
University of Michigan Medical School, Department of Human Genetics, Ann Arbor, MI 48109, USA

sisted (Tunstall-Pedoe et al. 1994) despite the successes of various dietary interventions to lower risk of CHD in populations that have suffered the greatest burden of CHD (Dwyer and Hezel 1980; Pietinen et al. 1996) have suggested that dietary differences do not fully explain population differences in CHD rates. Attention has thus turned towards the role of genetic differentiation in determining the distribution of CHD within and between populations.

Genetic studies have focused on the genetics of serum total cholesterol and other risk factors that link genomic variation with variation in risk of CHD (Motulsky and Brunzel 1992; Sing and Moll 1989; Sing et al. 1996). Knowledge about an individual´s genotype is expected to improve prediction of his or her disease risk because (1) it captures information about the levels of and relationships between risk factors, (2) predictive utility of established risk factors for prediction of CHD may be genotype dependent, and (3) it may predict interindividual differences in response to alterations in the environment (Davignon 1993; Davignon et al. 1988; Haviland et al. 1995; Humphries et al. 1995; Kesäniemi et al. 1987; Reilly et al. 1994), or to biological changes associated with aging (Davignon 1993; Schächter et al. 1994). In this paper we employ published data to evaluate the utility of the geographic distribution of the three common alleles, ε_2, ε_3, and ε_4, of the gene coding for apolipoprotein E (apoE) for prediction of interpopulation variation in average levels of serum total cholesterol and the other traditional risk factors and CHD mortality rate among nine population based samples of middle-aged males studied by the international WHO MONICA Project. The apoE alleles were selected because the apoE molecule is important for uptake of triglycerides by the liver and transport of lipids in the circulation to target cells (Davignon 1993; Davignon et al. 1988; Mahley 1988).

Materials and methods

Populations of inference and statistical methods

We considered nine middle-aged male populations participating in the WHO MONICA Project, from which we had complete and accurate information available on CHD mortality rates, on average population levels of the traditional risk factors, and on the estimates of the relative population frequencies of the three common apoE alleles. Our rationale for selecting the WHO MONICA populations were (1) the availability of good coverage of mortality registrations within the monitored areas, (2) a standardized diagnostic criterion for CHD mortality across populations, (3) a standardized risk factor data collection scheme in every population, and (4) a centralized quality control protocol for measurement of the quantitative intermediate risk factors (WHO 1989; WHO MONICA 1988).

The estimates of the apoE allele frequencies for the WHO MONICA populations are ascertained either from recent review-type publications (Gerdes et al. 1992; Hallman et al. 1991; Xhignesse et al. 1991) or from Medline. Our goal was to identify all publications that report estimates of the relative apoE allele frequencies using population-based samples collected from the very same geographically defined areas from which the WHO MONICA samples were drawn.

Estimates of relative frequencies of the apoE alleles for two of the WHO MONICA populations (Iceland and Beijing, China) were obtained from one of the review publications (Gerdes et al. 1992), which summarizes the estimates of apoE allele frequencies for 45 populations representative of a broad range of ethnic and geographic strata. In this publication a population was included if the estimates of relative frequencies were obtained from at least 100 individuals who were sampled to be representative of the population at large.

Through Medline we identified a report from the European Atherosclerosis Research Study (EARS; Tiret et al. 1994) which provided estimates of the relative frequencies of the apoE alleles for one of the WHO MONICA populations, Glasgow, United Kingdom. Estimates of the relative apoE allele frequencies for four WHO MONICA populations (Lille, Strasbourg, and Toulouse, France, and Belfast, United Kingdom) were available from control groups of the Etude Cas-Temoins sur l'Infarctus du Myocarde study (ECTIM; Luc et al. 1994). Estimates of the relative apoE allele frequencies for the two Finnish MONICA populations were obtained from the Cardiovascular Risk in Young Finns study (CRYF; Lehtimäki 1992).

Phenotypes

The following risk factors were considered in our analysis: serum total cholesterol, systolic and diastolic blood pressure, smoking status, and body mass index (BMI), computed as weight divided by height squared (kg/m^2). Detailed information about the sampling strategy, measurement procedures and laboratory methods is given elsewhere (WHO 1989). Estimates of the average population levels of the risk factors were obtained from representative samples in the age range 35 to 64 years collected in connection with the first survey of the WHO MONICA Project performed between 1982 and 1989 (WHO 1989; WHO MONICA 1988). Population-specific annual CHD mortality rates are based on data collected between 1985 and 1987 (Tunstall-Pedoe et al. 1994).

Statistical methods

All analyses were carried out on data collected in males. The following four-step model-building strategy and hypothesis testing procedure was used to evaluate the utility of the geographic distribution of the three common apoE alleles, ε_2, ε_3, and ε_4, for prediction of interpopulation variation in average levels of serum total cholesterol and the other risk factors, and ultimately interpopulation variation in CHD mortality rate. First, we evaluated the utility of variation in the average levels of known risk factors in the prediction of variation in the CHD mortality rate among the nine MONICA populations by employing linear regression analysis. Second, a univariable linear regression analysis was employed to evaluate the utility of variation in the relative frequency of either the ε_2, ε_3, or the ε_4 allele in prediction of the variation in average population levels of the risk factors. Third, a linear regression analysis was employed to estimate the contribution of variation in the relative frequency of either the ε_2, ε_3, or the ε_4 allele to the prediction of variation in the CHD mortality rates among populations. Fourth, we carried out a bivariable linear regression analysis to assess whether knowledge about the relative frequency of either the ε_2, ε_3, or ε_4 allele significantly improved the prediction of variation in the CHD mortality rate among the populations after considering knowledge about variation in the average population level of any one of the risk factors identified in the first step of our analysis as a predictor. In this fourth step we considered only risk factors that were detected as significant predictors of CHD mortality rate in a univariable analysis at the $Pr=0.30$ level of test criterion. This relatively high probability level was selected because the use of a more traditional test criterion (such as 0.05) at this point could result in disregarding risk factors that can contribute significantly to the prediction of a disease outcome only when they are considered in a multivariable model (Hosmer and Lemeshow 1989). An F test was used to judge whether the improvement in prediction attributable to the variation in the relative frequency of either the ε_2, ε_3, or the ε_4 allele was statistically significant. It compares the sum of squares attributable to information about both the relative frequency of either the three apoE alleles and the CHD risk factor being considered (complete linear regression model) with the sum of squares attributable to inclusion of only the risk factor (a reduced linear regression model).

Table 1 Population-specific annual, age-adjusted coronary heart disease (CHD) mortality rate in middle-aged men, their average population levels of CHD risk factors, and relative population frequencies of the *apoE* alleles in nine populations participating in the WHO MON-ICA Project (N_1 the size of the sample from which average levels of CHD risk factors have been estimated, N_2 the size of the sample from which relative *apoE* allele frequencies have been estimated)

| Study population | CHD mortality rate/ 100 000 | N_1 | Risk factor | | | | | Relative *apoE* allele frequency | | | |
			Serum total cholesterol mmol/l : mg/dl	Systolic blood pressure mmHg	Diastolic blood pressure mmHg	Body mass index kg/m²	Proportion of current smokers %	N_2	ε_2	ε_3	ε_4
United Kingdom											
Glasgow	364	464	6.3 : 243.6	138	87	25.8	52.4	114	0.057	0.769	0.184
Belfast	279	916	6.0 : 232.0	135	84	25.9	34.0	175	0.103	0.754	0.143
Finland											
Kuopio	349	977	6.4 : 247.5	147	90	26.8	33.0	421	0.034	0.791	0.175
Turku-Loimaa	301	1205	6.2 : 239.8	144	88	26.9	30.3	289	0.043	0.758	0.199
Iceland	186	656	6.2 : 239.8	126	83	25.9	26.3	185	0.068	0.767	0.165
France											
Strasbourg	95	666	5.6 : 216.6	145	93	27.4	34.3	159	0.085	0.808	0.107
Lille	90	570	6.5 : 251.4	140	91	25.8	38.6	139	0.119	0.741	0.140
Toulouse	67	659	5.9 : 228.2	131	86	25.7	36.5	171	0.056	0.863	0.081
China											
Beijing	38	619	4.2 : 162.4	130	86	23.4	50.0	507	0.124	0.806	0.070
Unweighted mean	197	748	5.9 : 228.2	137	87	25.8	37.3	240	0.076	0.784	0.140

All statistical analyses were performed with the SAS Institute statistical software package (SAS 1990). Statistical significance was taken to be $Pr \leq 0.05$ unless otherwise noted.

Results

Characteristics of the populations

The CHD mortality rate for the samples of middle-aged males in the nine populations ranged from almost 400/100 000 in Glasgow, United Kingdom to less than 50/100 000 in Beijing, China (Table 1). The CHD mortality rates for the remaining seven populations are uniformly distributed between these two extremes. These estimates span the range of CHD rates observed worldwide among men in the 35- to 64-year age group.

Populations with the highest CHD mortality rate also tend to have the most atherogenic risk factor profiles. Of them, Glasgow is characterized by a particularly high proportion of current smokers and a high average population level of serum total cholesterol (Table 1). Similarly, in the province of Kuopio, Finland, which has the second highest CHD mortality rate of the nine populations considered here, average population levels of four risk factors, serum total cholesterol, systolic and diastolic blood pressure, and BMI were all high. The lowest average population levels of most of the risk factors were observed in Beijing, China with the exception of the relative frequency of smoking.

The distributions of the estimates of the relative frequencies of the three common *apoE* alleles varied significantly among the populations ($\chi^2 = 153.971$, $df = 16$, $Pr = 0.001$). The relative frequency of the ε_2 allele ranged from 0.03 to 0.12, the ε_3 allele ranged from 0.74 to 0.86, and the ε_4 allele ranged from 0.07 to 0.20. The relative frequencies of the ε_2 and ε_4 alleles are inversely correlated (Spearman rank correlation = −0.633, $Pr = 0.06$). Those populations with a high CHD mortality rate, such as Glasgow, United Kingdom and the province of Kuopio, Finland, are characterized by a high relative frequency of the ε_4 allele and low relative frequency of the ε_2 allele (Table 1). In contrast, in Beijing, China, where the CHD mortality rate is extremely low, the ranking of these two alleles is reversed. The populations with the lowest CHD mortality rates tend to have the highest relative frequencies of the ε_3 allele. There tended to be a stronger inverse Spearman rank correlation of relative frequency of the ε_3 allele with the relative frequency of the ε_4 allele (−0.52, $Pr = 0.15$) than with the ε_2 allele (−0.20, $Pr = 0.61$).

Table 2 Contribution of variation in the average population levels of established risk factors to prediction of interpopulation variation in the CHD mortality rate in men

Risk factor	$R^2 \times 100$	Pr
Serum total cholesterol	33	0.10
Systolic blood pressure	15	0.29
Diastolic blood pressure	0	0.83
Body mass index	18	0.25
Smoking %	0	0.83

Traditional risk factors as predictors of CHD mortality rate

Although the CHD mortality rate of a particular population tends to reflect the risk factor profile in that population, average serum total cholesterol was the only traditional risk factor that was close to providing a statistically significant result in our analyses (Table 2). Approximately one-third of interpopulation variation in estimates of CHD mortality rate was associated with variation in estimates of average population level of serum total cholesterol at the $Pr=0.10$ level of statistical significance in a univariable model where the contribution of other risk factors was ignored. The estimate of the expected change in CHD mortality rate was 106/100 000 for a change of 1 mmol/l (2.7/100 000 for a change of 1 mg/dl) in average serum total cholesterol. Of the other traditional risk factors, variation in estimates of average population levels of BMI and systolic blood pressure predicted approximately 15% to 20% of interpopulation variation in CHD mortality rate ($Pr<0.30$). Interpopulation variation in average diastolic blood pressure, or in the proportion of current smokers, predicted less than 0.01% of the variation in CHD mortality rate ($Pr>0.80$).

The *apoE* alleles as predictors of risk factor profile

Approximately one-half of the interpopulation variation in the estimates of average population level of serum total cholesterol was associated with variation in the estimates of

relative population frequencies of the ε_4 allele ($Pr=0.02$) when the contributions of the other *apoE* alleles and other risk factors are ignored (Table 3). In contrast, variation in the relative frequency of the ε_2 allele predicts only approximately one-quarter of interpopulation variation in the average population level of serum total cholesterol level ($Pr=0.15$). Variation in the ε_3 allele provided little information about variation in traditional risk factors. The observed association between average serum total cholesterol level and the relative frequencies of the *apoE* alleles was positive when the ε_4 allele was considered, (the expected increase in average serum total cholesterol level was 0.114 mmol/l, or 4.405 mg/dl, for an increase of 0.01 in relative frequency) and negative when the ε_2 allele was considered (the expected decrease in average serum total cholesterol level was 0.112 mmol/l, or 4.334 mg/dl, for an increase of 0.01 in relative frequency). Variation in the relative frequencies of both the ε_4 allele ($Pr=0.13$) and the ε_2 allele ($Pr=0.09$) also predicted approximately one-third of interpopulation variation in average BMI. Interpopulation differences in average levels of systolic and diastolic blood pressure were not significantly associated with variation in the relative frequency of any of the three *apoE* alleles considered in these analyses.

The *apoE* alleles as predictors of CHD mortality rate

Consistent with the observation that the geographic distribution of the *apoE* alleles predicts variation in the average population level of one of the major traditional risk factors, serum total cholesterol, we found evidence that it also predicts interpopulation differences in the CHD mortality rate. Variation in the relative frequency of the ε_4 allele made a statistically significant contribution to prediction of the geographic variation of CHD mortality ($Pr=0.002$) when the contributions of other *apoE* alleles and traditional risk factors are ignored (Table 4). Similarly, the relative frequency of the ε_2 allele was a significant predictor of CHD mortality rate ($Pr=0.07$) when the contributions of other *apoE* alleles and traditional risk factors are ignored. The estimated fraction of the interpopulation variation in CHD mortality rate associated with variation in the relative frequencies of the ε_2 allele (39%) was, however, approximate-

Table 3 The separate contribution of interpopulation variation in the relative frequency of the ε_2, ε_3, or ε_4 allele to prediction of variation in the average population level of each established risk factor

Risk factor	*apoE* allele					
	ε_2		ε_3		ε_4	
	$R^2 \times 100$	Pr	$R^2 \times 100$	Pr	$R^2 \times 100$	Pr
Serum total cholesterol	28	0.15	19	0.25	55	0.02
Systolic blood pressure	11	0.39	3	0.65	14	0.32
Diastolic blood pressure	0	0.90	0	0.90	0	0.98
Body mass index	35	0.09	2	0.73	29	0.13

Table 4 The separate contribution of interpopulation variation in the relative frequency of the ε₂, ε₃, or ε₄ allele to prediction of variation of CHD mortality rate among populations before (A) and after the inclusion of each of the indicated risk factors is considered (B–D)

	Risk factor considered in bivariable analysis			
	A Univariable analysis	B Serum total cholesterol	C Systolic blood pressure	D Body mass index
apoE allele				
ε₂				
$R^2 \times 100$	39	15	28	21
Pr	0.07	0.24	0.12	0.18
ε₃				
$R^2 \times 100$	25	8	19	20
Pr	0.17	0.40	0.23	0.21
ε₄				
$R^2 \times 100$	75	43	60	57
Pr	0.002	0.02	0.01	0.01

ly half of that associated with variation in the relative frequency of the ε₄ allele (75%). The estimate of the expected association between CHD mortality rate and relative frequencies of the *apoE* alleles was positive when the ε₄ allele was considered (estimated slope = 24.5/100 000 for an increase of 0.01 in relative frequency, see Fig. 1A) and negative when the ε₂ allele was considered (estimated slope = -24.6/100 000 for an increase of 0.01 in relative frequency, see Fig. 1B). Variation in the relative frequency of the ε₃ allele was not significantly associated with variation in CHD mortality rates.

When information about relative frequency of the ε₄ allele was entered into a model that already included average level of either serum total cholesterol, systolic blood pressure, or BMI, which were significant predictors of CHD mortality rate in univariable analyses at *P*r=0.30 level of statistical significance, a highly significant improvement in the prediction of the interpopulation differences in CHD mortality rate was observed (Table 4). The estimated fraction of interpopulation variation in CHD mortality rate associated with variation in the relative frequency of the ε₄ allele was approximately 40% (*P*r=0.02) when the contribution of average serum total cholesterol was first considered in a bivariable model, and approximately 60% (*P*r=0.01) when average systolic blood pressure or BMI was first considered in a bivariable model. No statistically significant improvement in prediction could be detected when information about relative frequency of the ε₂ or ε₃ allele was entered into a model that already included estimates of the average population levels of either average serum total cholesterol, systolic blood pressure, or BMI (Table 4).

Discussion

### Traditional risk factors as ecological predictors

Ecological studies have had a considerable impact on CHD medicine in the past. One of the most important findings in these studies has been that up to two-thirds of interpopulation differences in CHD morbidity and mortality rates may be attributable to variation in average consumption of saturated fatty acids, and ultimately in average population level of serum total cholesterol (Keys 1970; Fidanza and Kromhout 1993; McGill 1996; Zevallos et al. 1992). Numerous dietary intervention programs have been implemented based on these findings. The programs have reduced CHD morbidity and mortality up to a 60% in high risk populations (Dwyer and Hezel 1980; Pietinen et al. 1996). Furthermore, the fraction of the interpopulation differences in CHD rates that can be predicted by variation in average levels of traditional risk factors has halved, to about one-third (Stewart et al. 1994). Approximately 15% to 30% of the observed interpopulation variation in the CHD mortality rate in our analyses was attributable to variation in the average population level of either serum total cholesterol (33%), BMI (18%), or systolic blood pressure (15%) when the contributions of other risk factors are ignored. A fraction of the remaining 67% to 85% of interpopulation variation that cannot be predicted by the established risk factors may be ascribed to genetic differentiation of populations. A fraction of the unexplained variation in deviations from expectations will be ascribable to intrinsic stochastic processes that will not yield to explanation by traditional linear modeling (Ekland 1993).

ApoE alleles as ecological predictors

Knowledge about the contribution of genetic differentiation of populations to the determination of the regional differences in the CHD morbidity and mortality rates is scanty. Observations that the relative frequency of the ε₄ allele, which nearly always predicts elevated serum total cholesterol levels in individuals (Davignon 1993; Davignon et al. 1988), tends to be high in populations that have high CHD rates, and low in populations where the rates are low, have suggested the hypothesis that the three common *apoE* alleles are involved in determining the heterogeneity of CHD rates among populations (Davignon 1993; Davignon et al. 1988; Sing and Moll 1989). In our analyses approximately half of the interpopulation variation in the average population level of serum total cholesterol and approximately three-quarters of the interpopulation variation in the CHD mortality rate were associated with variation in the relative frequency of the high risk ε₄ allele across the nine WHO MONICA populations considered in this study. These results further support the hypothesis that variation in the relative frequencies of the ε₂, ε₃ and ε₄ alleles confer information about the variation in average serum total cholesterol level, and ultimately CHD rate, among populations

Fig 1 Linear regression relationships between CHD mortality rate and relative frequency of the ε4 (A) or ε2 allele (B) among nine middle-aged male populations studied in connection with the WHO MONICA Project.
Fi1 = the province of Kuopio, Finland
Fi2 = the Turku-Loimaa region, Finland
UK1 = the city of Glasgow, United Kingdom
UK2 = the city of Belfast, United Kingdom
Ice = Iceland
Fr1 = the city of Lille, France
Fr2 = the city of Strasbourg, France
Fr3 = the city of Toulouse, France
Chn = the city of Beijing, China

$\beta = 24.55/100,000/0.01$
$Pr < 0.01$
$R^2 = 0.76$

$\beta = -24.58/100,000/0.01$
$Pr = 0.07$
$R^2 = 0.39$
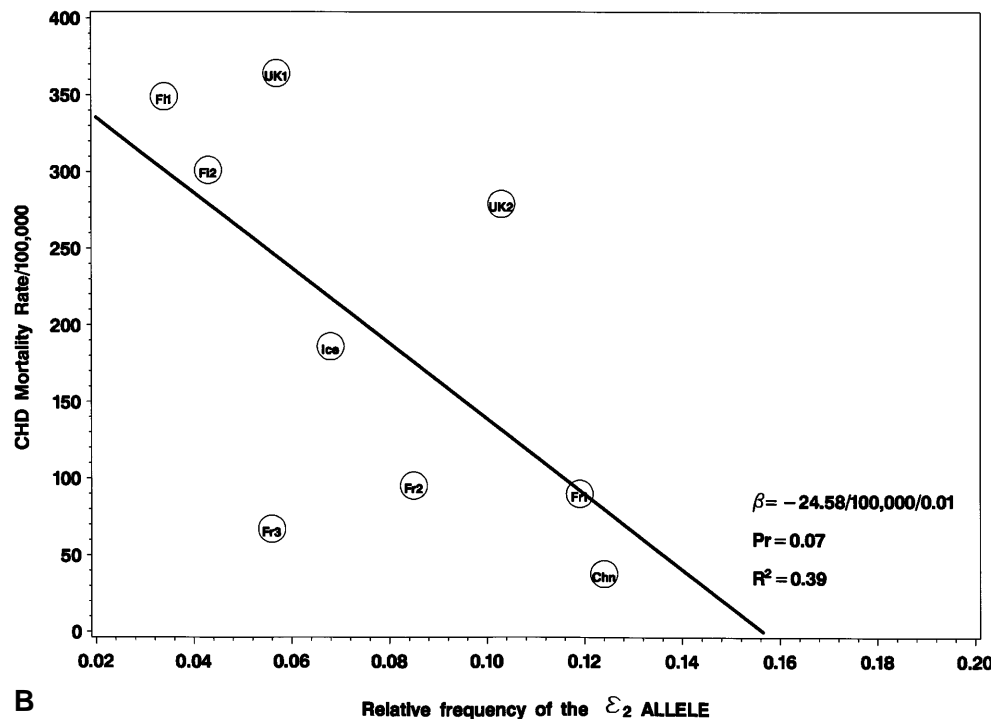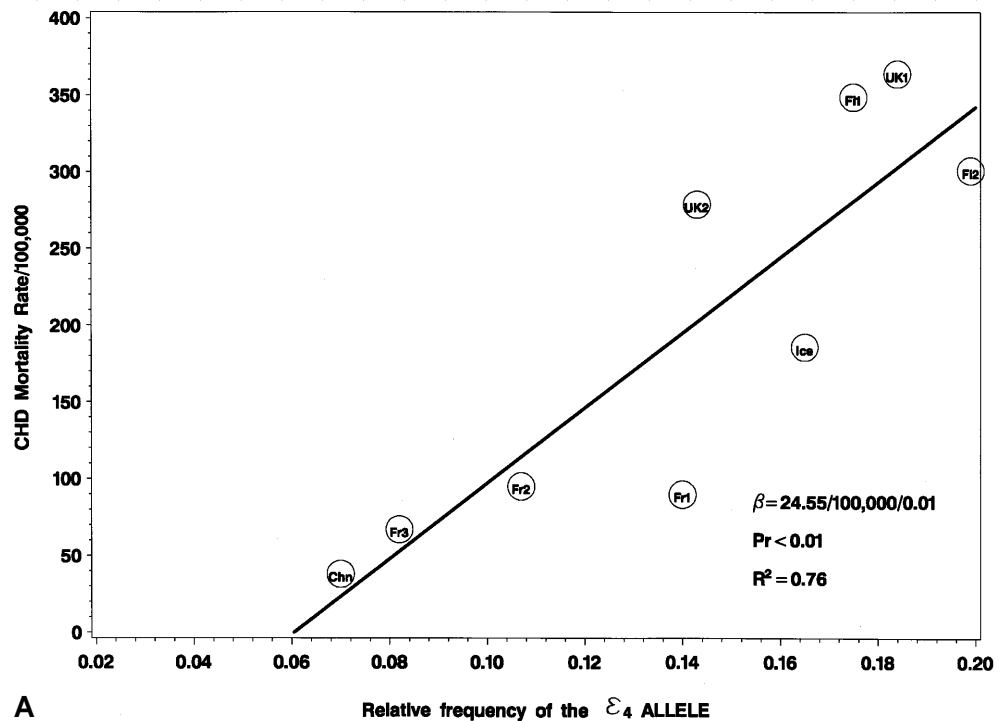
(Davignon 1993; Davignon et al. 1988; Sing and Moll 1989).

The causal relationships between genomic variations and interindividual variation in initiation, progression, and severity of a complex disease such as CHD involve connections through multiple networks of intermediate biochemical and physiological agents (Ferrannini 1991; Sing et al. 1996). The complexity of the connections between causal agents in these networks across the hierarchy from genomic variation to variation in the risk of CHD argues that the utility of genomic information for prediction of risk of disease may not persist after the contribution of intermediate

risk factors is considered in risk assessment. Our findings argue that estimates of the relative frequency of the ε_4 allele confer ecological risk information about CHD mortality rate that is not solely mediated through its effect on known intermediate risk factors.

Another related question is whether knowledge about the distribution of the relative frequency of the ε_2 allele among populations provides complementary information about variation in average serum total cholesterol or CHD mortality rate among populations (Davignon 1993; Davignon et al. 1989). In our analyses the additional information about relative frequency of the ε_2 allele in a bivariable model where relative frequency of the ε_4 allele was also included did not improve our ability to predict either a population's average serum total cholesterol level or its CHD mortality rate (data not shown). This was also the case when any of the other average risk factor levels were included in a bivariable model. Similarly, we detected only marginally significant ($P=0.07$) predictive value for the relative frequency of the ε_2 allele in any univariable analyses where the contribution of the ε_4 allele was ignored. These observations, together with recent meta-analyses by Wilson et al. (1996), argue that most of the ecological risk information in the *apoE* polymorphism resides in the ε_4 allele when the prediction of CHD mortality is considered.

Clinical and public health utility of ecologic risk factors

Individual-based studies have established that information about an individual's *apoE* genotype may improve our ability to assess his or her risk of CHD (Stengård et al. 1996). Furthermore, knowledge about an individual´s *apoE* genotype is expected to predict how he or she will respond to alterations in the environment defined by diet, exercise, smoking, gender and drug therapy (Davignon 1993; Davignon et al. 1988; Haviland et al. 1995; Humphries et al. 1995; Kesäniemi et al. 1987; Reilly et al. 1994). Both of these properties are expected to provide tools for clinicians and public health professionals to promote individually tailored intervention activities among those with the greatest potential for health benefits.

Ecological studies are expected to provide information for developing population-based intervention strategies to reduce the disease incidence (Rose 1992). Potentially, ecologically identified risk factors may also be employed in developing individual-based high-risk intervention strategies if an observed ecological relationship between the average population level of a variable and the average population level of its predictors are a direct consequence of the relationship between the levels of the variable and its predictors within a population. Our study demonstrates that quantitatively this may not be the case when a genetic polymorphism is considered. The principles of quantitative genetics (Falconer and Mackay 1996) suggest that if we assume that (1) there are only two functional alleles, (2) the Hardy-Weinberg relationship between the relative frequencies of these two alleles and genotype frequencies is true, (3) there are no effects due to dominance or epistasis with other loci,

(4) the effects of the alleles are independent of the influences of environmental variation, and (5) the relative frequencies of allele effects are uncorrelated with the relative frequencies of other gene and environmental effects, the mean of the ith population, μ_i, may be expressed as a linear function of the average phenotypic level of the homozygote, G_{11}, for the first allele and the product of the average effect of substitution of a second allele for the first (α) and the relative frequency of the second allele, p_i:

$$\mu_i = G_{11} + 2 * \alpha * p_i.$$

If we further assume that G_{11} represents the average serum total cholesterol level of individuals who do not carry the ε_4 allele, α represents the average effect of substitution of the ε_4 allele, p_i represents the relative frequency of the ε_4 allele in the ith population and there are no effects of other alleles of the *apoE* gene, the average serum total cholesterol level of the population is expected to increase $2*\alpha/100$ when the relative frequency of the ε_4 allele increases by 0.01. A study of Hallman et al. (1991) suggests that within-population estimates of G_{11} (intercept) may vary between 3.67 mmol/l (141.92 mg/dl) and 5.85 mmol/l (226.82 mg/dl). The population-specific estimates of α (slope) may vary between –0.03 mmol/l (–1.16 mg/dl) and 0.38 mmol/l (14.69 mg/dl). Estimates of the intercept and slope of the linear regression of average serum total cholesterol level on relative frequency of the ε_4 allele from our ecological study were 4.32 mmol/l (167.10 mg/dl) and 11.39 mmol/l (440.49 mg/dl), respectively. The large disparity between estimates of slope from our ecological study and the within-population estimates of slope is strong evidence that one or more of the above-mentioned assumptions is not true. Unraveling the responsible factors may be an unrealistic expectation from the application of linear statistical methods (Sing et al. 1996).

Genetic heterogeneity of CHD dictates that results of ecological studies that consider only one genetic polymorphism at a time, such as the *apoE* polymorphism in this study, need to be translated to public health activities with caution. Particular attention should be paid to the possibility that different sets of high-risk alleles may be operating in different populations. Documentation that we should be concerned about this possibility, at least for the rare alleles, comes from the observation that different mutations in the low density lipoprotein receptor gene leading to familial hypercholesterolemia, a well-known monogenic disease, are found exclusively in different populations in Israel (Leitersdorf and Friedlander 1995). Hence, the success of any proposed public health program to use genetic information to identify and treat subsets of individuals, or populations, susceptible to CHD depends critically on (1) an understanding of the distribution of genetic variation within and among populations and (2) knowing the extent to which the impact of a particular allelic variation on risk of CHD depends on the genetic and environmental contexts that vary among individuals and among populations.

Our analyses demonstrate the utility of the relative frequency of the ε_4 allele in predicting interpopulation variability of serum total cholesterol level and CHD mortality rate. Information about the distribution of the CHD mortal-

ity persists after interpopulation variation in other traditional risk factor levels is considered. Studies of the *apoE* polymorphism clearly document that public health programs may rely to a greater extent in the future on the distribution of measured genetic variations that influence the distribution of risk of CHD within and among populations. However, the disparity of the findings of our ecological study with the findings of association studies within populations regarding the relationship between serum total cholesterol and the *apoE* polymorphism suggests caution in relying on information from the study of a single gene in a particular population to develop a public health program for a common disease having complex multifactorial etiology.

References

Davignon J (1993) Apolipoprotein E polymorphism and arteriosclerosis. In: Born GVR, Schwartz CJ (eds) New horizons in coronary heart disease. Current Science/Science Press, London, pp 5.1–5.21

Davignon J, Gregg RE, Sing CF (1988) Apolipoprotein E polymorphism and atherosclerosis.Arteriosclerosis 8:1–21

Dwyer T, Hezel BS (1980) A comparison of trends of coronary heart disease mortality in Australia, USA and England and Wales with reference to three major risk factors – hypertension, cigarette smoking and diet. Int J Epidemiol 9:65–71

Ekland I (1993) The broken dice, and other mathematical tales of chance. University of Chicago Press, Chicago

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, Longman Group, Essex

Ferrannini E (1991) Metabolic abnormalities of hypertension: a lesson in complexity. Hypertension 18:636–639

Fidanza F, Kromhout D (1993) Dietary surveys in the seven countries study. In: Kromhout D, Menotti A, Blackburn H (eds) The seven countries study: a scientific adventure in cardiovascular disease epidemiology. Brouwer Offset, Utrecht, Netherlands, pp 151–157

Gerdes LU, Klausen C, Sihm I, Faergeman O (1992) Apolipoprotein E polymorphism in a Danish population compared to findings in 45 other study populations around the world. Genetic Epidemiol 9:155–167

Hallman DM, Boerwinkle E, Saha N, Sandholzer C, Menzel HJ, Czázár A, Utermann G (1991) The apolipoprotein E polymorphism: a comparison of allele frequencies and effects in nine populations. Am J Hum Genet 49:338–349

Haviland MB, Lussier-Cacan S, Davignon J, Sing CF (1995) Impact of apolipoprotein E genotype variation on means variances, and correlations of plasma lipid, lipoprotein, and apolipoprotein traits in octogenarians. Am J Med Genet 58:315–331

Hosmer DW, Lemeshow S (1989) Applied logistic regression. Wiley, Toronto

Humphries SE, Ye S, Talmud P, Bara L, Wihelmsen L, Tiret L, EARS (1995) European atherosclerosis research study: fenotype at the fibrinogen locus (G$_{-455}$ -A β-gene) is associated with differences in plasma fibrinogen levels in young men and women from different regions in Europe. Arterioscler Thromb Vas Biol 15:96–104

Kesäniemi YA, Ehnholm C, Miettinen TA (1987) Intestinal cholesterol absorption efficiency in man related to apoprotien E phenotype. J Clin Invest 80:578–581

Keys A (1970) Coronary heart disease in seven countries. Circulation 41[Suppl 1]:1–211

Lehtimäki T (1992) Apolipoprotein E phenotypes in Finnish children and young adults. Acta Univ Tamperensis Ser A 340

Leitersdorf E, Friedlander Y (1995) Familial hypercholesterolemia in Israel: molecular genetics, phenotype characteristics and pharmacogentics. Proceedings of the 65th Meeting of the European Atherosclerosis Society Genetics of Atherosclerotic Heart Disease, Denmark September 28-October 1, pp 10

Luc G, Bard J, Arveiler D, Evans A, Cambou J, Bingham A, Amouyel P, Schaffer P, Ruidavets J, Cambien F, Fruchart J, Ducimetiere P (1994) Impact of apolipoprotein E polymorphism on lipoproteins and risk of myocardial infarction: the ECTIM study. Arterioscler Thromb 14:1412–1419

Mahley RW (1988) Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. Science 240:622–630

McGill HC Jr (1996) Major risk factors and primary prevention, part 1. In: Ross VFR, Topol EJ (eds) Atherosclerosis and coronary artery disease. Lippincott-Raven, Philadelphia, pp 25–38

Motulsky AG, Brunzel J D (1992) The genetics of coronary atherosclerosis. In: The genetic basis of common disease. Oxford University Press, Oxford, pp 150–169

Pietinen PD, Vartiainen E, Seppänen D, Aro A, Puska P (1996) Changes in diet in Finland from 1972 to 1992: impact on coronary heart disease risk. Prev Med 25:243–250

Reilly SL, Ferrell RE, Sing CF (1994) The gender-specific apolipoprotein E genotype influence on the distribution of plasma lipids and apolipoproteins in the population of Rochester, MN. III. Correlations and covariances. Am J Hum Genet 55:1001–1018

Rose G (1992) The strategy of preventive medicine. Oxford University Press, Oxford

SAS Institute (1990) SAS user´s guide, version 6, 4th edn, vol. 1. SAS Institute, Cary, N.D.

Schächter F, Faure-Delanef L, Guénot F, Rouger H, Froguel P, Lesueur-Ginot L, Cohen D (1994) Genetic associations with human longevity at the APOE and ACE loci. Nat Genet 6:29–32

Sing CF, Moll PP (1989) Genetic of variability of CHD risk. Int J Epidemiol 18[Suppl 1]:S183-S195

Sing CF, Haviland MB, Reilly SL (1996) Genetic architecture of common multifactorial diseases. In: Chadwick D, Cardew G (eds) Variation in the human genome. (Ciba Foundation Symposia 197) Wiley, Chichester, pp 211–232

Stengård JH, Pekkanen J, Ehnolm C, Nissinen A, Sing CF (1996) Genotypes with the apolipoprotein ε4 allele are predictors of coronary heart disease in mortality in a longitudinal study of elderly Finnish men. Hum Genet 97:677–684

Stewart AW, Kuulasmaa K, Beaglehole R (1994) Ecological analysis of the association between mortality and major risk factors of cardiovascular disease. Int J Epidemiol 23:505–516

Tiret L, Knijff P de, Menzel H, Ehnholm C, Nicaud V, Havekes LM (1994) ApoE polymorphism and predisposition to coronary heart disease in youths of different European populations: the EARS study. Arterioscler Thromb 14:1617–1624

Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM, Pajak A (1994) Myocardial infarction and coronary deaths in the World Health Organization MONICA Project: registration procedures, events rates, and case-fatality rates in 38 populations from 21 countries in four continents. Circulation 9:583–612

Wilson PWF, Schaefer EJ, Larson MG, Ordovas JM (1996) Apolipoprotein E alleles and risk of coronary disease. Arterioscler Thromb Vasc Biol 16:1250–1255

WHO MONICA Project (1988) Geographical variation in the major risk factors of coronary heart disease in men and women aged 35–64 years. World Health Stat Q 41:115–140

World Health Organization (1989) The WHO MONICA Project: a worldwide monitoring system for cardiovascular diseases. World Health Statistics Annual

Xhignesse M, Lussier-Cacan S, Sing CF, Kessling AM, Davignon J (1991) Influences of common variants of apolipoprotein E on measures of lipid metabolism in a sample selected for health. Arterioscler Thromb 11:1100–1110

Zevallos JC, Chiriboga D, Hebert JR (1992) An international perspective on coronary heart disease and related risk factors. In: Ockene IS, Ockene JK (eds) Prevention of coronary heart disease. Little, Brown & Co., Boston, pp 147–170