Am. J. Hum. Genet. 56:835–844, 1995

Epidemiological Evaluation of the Use of Genetics to Improve the Predictive Value of Disease Risk Factors

Muin J. Khoury[1] and Diane K. Wagener[2]

[1]Birth Defects and Genetic Diseases Branch, Division of Birth Defects and Developmental Disabilities, National Center for Environmental Health, Atlanta; and [2]Office of Analysis, Epidemiology, and Health Promotion, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD

Summary

The prevention of common diseases relies on identifying risk factors and implementing intervention in high-risk groups. Nevertheless, most known risk factors have low positive predictive value (PPV) and low population-attributable fraction (PAF) for diseases (e.g., cholesterol and coronary heart disease). With advancing genetic technology, it will be possible to refine the risk-factor approach to target intervention to individuals with risk factors who also carry disease-susceptibility allele(s). We provide an epidemiological approach to assess the impact of genetic testing on the PPV and PAF associated with risk factors. Under plausible models of interaction between a risk factor and a genotype, we derive values of PPV and PAF associated with the joint effects of a risk factor and a genotype. The use of genetic testing can markedly increase the PPV of a risk factor. PPV increases with increasing genotype–risk factor interaction and increasing marginal relative risk associated with the factor, but it is inversely proportional to the prevalences of the genotype and the factor. For example, for a disease with lifetime risk of 1%, if all the risk-factor effect is confined to individuals with a susceptible genotype, a risk factor with 10% prevalence and disease relative risk of 2 in the population will have a disease PPV of 1.8%, but it will have a PPV of 91.8% among persons with a genotype of 1% prevalence. On the other hand, genetic testing and restriction of preventive measures to those susceptible may decrease the PAF of the risk factor, especially at low prevalences of the risk factor and genotype. With advances in the Human Genome Project, medicine and public health should consider the feasibility of this approach as a new paradigm for disease prevention.

Received October 17, 1994; accepted for publication January 5, 1995.

Address for correspondence and reprints: Dr. Muin J. Khoury, Birth Defects and Genetic Diseases Branch, National Center for Environmental Health, Centers for Disease Control and Prevention, F45, 1600 Clifton Road, Atlanta, GA 30333.

0002-9297/95/5604-0004$02.00

Introduction

The prevention of common and chronic diseases, such as coronary heart disease, has been complicated by the multifactorial nature of these diseases (Badimon 1993). While the classical epidemiological paradigm of searching for "risk factors" and intervening in high-risk groups has enjoyed much success in controlling and preventing many infectious diseases, its success in preventing chronic diseases has been mixed. For example, >270 factors have been suggested as risk factors for coronary heart disease (Stehbens 1992). Even to date, there is continued discussion about the prevention usefulness of some well-established coronary-heart-disease risk factors, such as diet and hypercholesterolemia (LaRosa 1992; Allred 1993; Hamsten 1993; Woodard 1993). It has been shown that, for most risk factors for chronic diseases, their utility in predicting disease is limited (Khoury et al. 1985; Baron 1989). Also, the sensitivity of such factors could be small, limiting both their clinical relevance and their potential for prevention (Khoury et al. 1985).

The new era of molecular genetics and the Human Genome Project promises the identification of numerous genes and their allelic variants that, per se, may not cause disease but may interact with other genes and environmental factors in causing disease (i.e., they may be genetic risk factors). There is a growing interest in the concept of gene-environment interaction in disease causation (Balant et al. 1992; Hegele 1992; Khoury et al. 1993). To highlight the importance of gene-environment interaction in coronary heart disease, Hegele (1992, p. 177) states that "some vegetarians with 'acceptable' cholesterol levels suffer myocardial infarction in the 30's. Other individuals...seem to live forever despite personal stress, smoking, obesity, and poor adherence to a Heart Association-approved diet." To emphasize the importance of genetics in nutrition, Simopoulus et al. (1993, cover) state, in a recently published book on genetic nutrition, that "your genes can tell you what to eat and avoid—to live a longer, healthier life." Furthermore, it recently has been predicted that "the day of the personal DNA profile provided at birth, complete with calculated risks of various cancers, heart disease, and

many other conditions could be an actuality by the time that current first-year medical students begin to practice medicine" (Hoffman 1994, p. 130).

Even if all 50,000–100,000 human genes are cloned and their functions discovered, and even if sensitive and specific tests are available, it is still far from clear whether we can use genetics to predict disease risk, especially for multifactorial common disorders. One concern over the use of genetic testing in disease prediction is the notion that genetic tests will have poor disease-predictive ability as well as poor sensitivity (Khoury et al. 1985; Holtzman 1992). Another major concern is related to ethical issues in using genetic testing (Holtzman 1989; Suzuki and Knudston 1989; Hubbard and Wald 1993; Garver and Garver 1994). These concerns have led to the development of the Ethical, Legal, and Social Issues Program (ELSI) within the National Center for Human Genome Research (Hoffman 1994).

With advancing genetic technology and ongoing societal discussion regarding the use of genetic testing, will it be possible to refine the risk-factor approach to target preventive measures to individuals with risk factors who also carry disease-susceptibility allele(s)? In this article, we provide an epidemiological assessment of the value of using genetic tests to improve the predictive value of disease risk factors. Using simple epidemiological parameters and several plausible schemes of genotype–risk factor interaction, we show the impact of appropriate genetic testing on the positive predictive value (PPV) and the population-attributable fraction (PAF) associated with disease risk factors.

Methods

Marginal Effects of a Risk Factor

Given a disease with a lifetime risk of d in the population, suppose that a risk factor (an exposure with prevalence e) is associated with the risk of the disease with a relative risk R (equivalent to an odds ratio for rare conditions). For this paper, the clinical and public health impacts of the factor on disease occurrence can be measured in two ways:

1. PPV of the factor—PPV(e): this refers to the risk of disease among individuals with the factor (the clinical impact). As shown in table 1, PPV(e) can be written as

$$\text{PPV}(e) = d \cdot R/[1 + e(R - 1)] . \qquad (1)$$

2. PAF of the factor—PAF(e): this refers to the proportion of cases that could be prevented if the factor was absent (the public health impact). By the method of Levin (1953), PAF(e) can be written as

Table 1

Parameters of a Genotype–Risk Factor Interaction Model of Disease Risk

Susceptibility Genotype (g)	Risk Factor (e)	Prevalence	Relative Risk
Absent	Absent	$(1 - g)(1 - e)$	1
Absent	Present	$(1 - g)e$	R_e
Present	Absent	$g(1 - e)$	R_g
Present	Present	ge	R_{ge}

NOTE.—Exposure and genotype are independent. $d = (1 - e) \times P(\text{D}|\text{no factor}) + eP(\text{D}|\text{factor})$; marginal relative risk associated with exposure is $R = P(\text{D}|\text{factor}|/P(\text{D}|\text{no factor})$; replacing the value of $P(\text{D}|\text{factor})$ from the second equation into the first equation, we obtain $P(\text{D}|\text{no factor}) = d/[(1 + e(R - 1)]$; $P(\text{D}|\text{factor})$ or PPV (e) $= dR/[1 + e(R - 1)]$; $R = [(1 - g)R_e + gR_{ge}][(1 - g) + gR_g]$. The apparent relative risk for genotype–risk factor combination is $R'_{ge} = [(1 - ge)R_{ge}]/[(1 - g)(1 - e) + (1 - g)eR_e + g(1 - e)R_g]$.

$$\text{PAF}(e) = e(R - 1)/[1 + e(R - 1)] . \qquad (2)$$

Throughout our illustrations, we use examples of diseases with lifetime risks ranging from 0.1% (such as for some specific birth defects) to 10% (such as for some specific cancers). For a disease with lifetime risk of 1%, if a risk factor with 10% prevalence has a relative risk of 2, then we can calculate that PPV(e) = 1.8% and PAF(e) = 9.1%. Such a risk factor, typical of many chronic-disease risk factors, is poorly predictive of the disease and also accounts for a small fraction of cases (etiologic heterogeneity). For simplicity, we assume that the risk factor is dichotomous (present/absent) although dose-response functions can be addressed.

Genotype–Risk Factor Interaction

We assume that, underlying the marginal effects of a risk factor, there exists a pattern of a genotype–risk factor interaction in producing disease. For simplicity, we assume a measurable genotype (which could be one allele at one locus or multiple alleles at multiple loci) with a prevalence g independent of the risk factor. The joint effects of risk factor and genotype on disease risk are shown in table 1. Three values of relative risk can be calculated relative to individuals with neither susceptibility genotype nor risk factor:

- R_e: relative risk of disease with factor alone (i.e., no susceptibility genotype);
- $*R_g$: relative risk of disease with genotype alone (i.e., no risk factor);
- $*R_{ge}$: relative risk of disease with both factor and genotype.

The background risk of disease for people with neither the factor nor the genotype is assumed to be >0, suggesting etiologic heterogeneity in the disease.

Effects of the Risk Factor with the Susceptibility Genotype

To evaluate how performing a genetic test on individuals with a risk factor can affect measures of PPV and PAF associated with the risk factor, we now consider "exposed" individuals only, i.e., those with the factor plus the susceptibility genotype. The prevalence of individuals with risk-factor/genotype combination is simply the product of g and e (under the assumption of independence). The apparent relative risk for the risk-factor/genotype combination (R'_{ge}) can be derived by comparing disease risk for these individuals with disease risk for all others (not simply those with neither genotype nor risk factor). As shown in table 1, R'_{ge} is a function of e, g, and the individual relative risks R_e, R_g, and R_{ge}.

$$R'_{ge} = [(1 - ge)R_{ge}]/[(1 - g)(1 - e) + (1 - g)eR_e + g(1 - e)R_g] . \quad (3)$$

Thus, by using ge and R'_{ge} in equations (1) and (2), instead of e and R, respectively, we can calculate the PPV and PAF for the risk factor in combination with the susceptibility genotype (denoted "PPV(ge)" and "PAF(ge)," respectively).

$$\text{PPV}(ge) = d \cdot R'_{ge}/[1 + ge(R'_{ge} - 1)] . \quad (4)$$

$$\text{PAF}(ge) = ge(R'_{ge} - 1)/[1 + ge(R'_{ge} - 1)] . \quad (5)$$

Models of Genotype–Risk Factor Interaction

The relationship between R_{ge}, R_g, and R_e depends on the underlying model of interaction. For purely additive effects of factors (no interaction), R_{ge} will be equal to $R_e + R_g - 1$. For multiplicative effects, R_{ge} will be equal to the product of R_g and R_e. In this paper, in addition to multiplicative effects, we illustrate three biologically plausible models of interaction, which are described in more detail elsewhere (Khoury et al. 1988; authors' unpublished data).

Type 1 interaction.—Disease risk is increased only in the presence of the susceptibility genotype and the risk factor (i.e., R_g and R_e are both equal to 1). For example, phenylketonuria is an autosomal recessive disease in which the cardinal clinical manifestation (mental retardation) depends on both the presence of the phenylalanine hydroxylase enzyme deficiency and the presence of phenylalanine in the diet (Scriver et al. 1989). Since there are other causes of mental retardation, the background risk is not zero.

Type 2 interaction.—Disease risk is increased by the risk factor alone but not by the genotype alone. Thus, the effect of the susceptibility genotype is restricted to exposed individuals. Hence, $R_g = 1$, but R_e and R_{ge} are both >1. A possible example here is the association between lung cancer and cigarette smoking in relation to the debrisoquine hydroxylase genotype (Caporaso et al. 1989). Increasing evidence exists of biological interaction between smoking and this genetic polymorphism ($R_{ge} > 1$). Further, smokers without this genotype seem to have an increased risk for lung cancer ($R_e > 1$). However, we assume here for illustration that the susceptibility genotype alone probably does not increase the risk for lung cancer ($R_g = 1$).

Type 3 interaction.—Disease risk is increased by the genotype in the absence of the risk factor but not by the risk factor alone. The effect of the risk factor is thus restricted to persons with the susceptibility genotype. Hence, $R_e = 1$, but R_g and R_{ge} are both >1. A possible example here is the interaction between genetic forms of glucose-6-phosphate dehydrogenase deficiency (G6PD) and fava-bean ingestion in causing acute hemolytic anemia (Glader 1990). Fava beans per se do not cause hemolytic anemia ($R_e = 1$); however, certain genetic forms of G6PD deficiency can lead to hemolysis in the absence of fava-bean ingestion ($R_g > 1$) (e.g., G6PD forms affected by certain antimalarial drugs). Also, there are other causes of hemolytic anemia that are unrelated to either G6PD deficiency or fava-bean ingestion (i.e., the background risk is not zero).

Analyses

In the illustrations, we compare values of PPV and PAF for a risk factor under different scenarios: (1) marginal effects of the factor in the absence of genotypic information, (2) multiplicative effects of factor and genotype, (3) type 1 interaction, (4) type 2 interaction, and (5) type 3 interaction. In these illustrations, we vary values of prevalences of genotype and risk factor, disease lifetime risks, as well as marginal relative risks of the factor.

Results

Effects on PPV

In table 2, we show the PPV of a risk factor as a function of prevalence of the susceptibility genotype. For a disease with lifetime risk of 1%, a risk factor with 10% prevalence and relative risk of 2 will have a PPV of only 1.8%. The disease-predictive value of the factor (in the presence of the genotype) can be much higher, depending on the underlying model of interaction. For all models, however, the PPV of the risk factor/genotype decreases with increasing prevalence of the genotype. For type 1 interaction, when all the risk-factor effect is confined to individuals who carry a susceptibility genotype, the PPV is >90% for genotypes of 1% prevalence but is <5% for genotypes of 25%. For type 2 interactions, the PPV will be less than that for type 1. This can be intuitively understood because the risk factor has an effect on disease risk even for those individuals without

Table 2

Disease PPV (%) among Individuals with a Risk Factor and a Susceptibility Genotype, by Prevalence of Genotype and Type of Genotype–Risk Factor Interaction

	Type of Interaction[a]				
Prevalence of Genotype	Marginal Effect	Multiplicative	Type 1	Type 2	Type 3
1%	1.8	8.0	91.8	44.6	95.6
5%	1.8	7.1	19.1	10.0	22.3
10%	1.8	6.3	10.0	5.7	12.7
25%	1.8	4.5	4.5	3.1	6.1
50%	1.8	3.1	2.7	2.2	3.4
75%	1.8	2.3	2.1	1.9	2.3
100%	1.8	1.8	1.8	1.8	1.8

NOTE.—Assumptions: population disease risk = 1%; risk factor prevalence = 10%; and relative risk = 2.

[a] Type 1—relative risks for factor alone and genotype alone = 1; type 2—relative risk for factor alone = 1.5, for genotype alone = 1; type 3—relative risk for gentotype alone = 5, for factor alone = 1; and multiplicative—relative risk for factor alone = 2, for genotype alone = 5.

the susceptibility genotype ($R_e > 1$). For type 3 interactions, PPV values will be at least equal to those for type 1 interactions. They could be higher, depending on how much the susceptibility genotype has an effect on disease risk without the risk factor ($R_g > 1$). Generally, for multiplicative risk-factor/genotype effects, values of PPV will be higher than marginal PPV but much lower than the other special cases of interaction. For example, with multiplicative effects, the PPV associated with risk-factor/genotype combination will be 8% (for 1% prevalence of genotype), compared with 1.8% marginal PPV.

In figure 1, we show simultaneously, for type 1 interaction, the impact of marginal relative risk, exposure, and genotype frequencies on PPV for different disease frequencies. As can be seen, the higher the marginal relative risk, the higher the PPV of the risk factor in combination with the susceptibility genotype. For a disease with lifetime risk of 1% and a genotype with 10% prevalence, the PPV is 10% when the relative risk is 2 but is 48% when the relative risk is 10.

In figure 1, we also see the impact of the prevalence of the risk factor on PPV. The higher the prevalence of the risk factor, the lower the PPV associated with the combination of risk factor/genotype at the same prevalence of the genotype. Finally, we can also see the well-known impact of the prevalence of the disease itself on PPV (Khoury et al. 1985). The higher the lifetime risk of disease, the higher the PPV associated with the combination of risk factor/genotype at the same prevalence of the genotype. For example, when the disease occurs in 1/1,000 individuals (e.g., some birth defects), the predictive value of a risk-factor/genotype combination will be generally low ($\leq$1%), but it will be much higher for diseases with lifetime risk of 10% (e.g., some cancers and arthritis), for the same levels of relative risk, exposure, and genotype frequencies.

Effects on PAF

In table 3, we show the values of PAF for a risk-factor/genotype combination under different scenarios. We use the example of disease with lifetime risk of 1%, a risk factor with 10% prevalence, and relative risk of 2. When the marginal PAF is calculated, we can see that 9.1% of the disease can be attributed to the risk factor. Under type 1 interaction, when all the risk-factor effect is confined to persons with an underlying susceptibility genotype, the PAF associated with risk-factor/genotype combinations is the same as the marginal one. Under type 2 interaction, the PAF will be less than that under type 1 interaction, since some of the risk-factor effect occurs even in the absence of the genotype ($R_e > 1$). Under type 3 interaction, the PAF will be at least equal to the marginal PAF, since there is a genetic effect even in the absence of the risk factor ($R_g > 1$). Values of PAF could be lower than the marginal PAF and those calculated under specific types of interaction. In contrast to the effects on PPV, increasing the prevalence of the susceptibility genotype tends to increase values of PAF.

Discussion

In these analyses, we have provided an epidemiological approach to assessing the impact that the use of genetic testing has on the usefulness of disease risk factors. In particular, we examined the effect of genetic testing on the disease PPV and PAF. PPV was chosen because of its clinical and preventive implications in risk communication and counseling. Given that numerous

risk factors exist for common chronic diseases, it is important to assess the disease absolute risks rather than relative risks, which are the usual estimated parameters in epidemiological studies. The second measure (PAF) was chosen because of its public health implications in terms of the reduction of disease occurrence in the absence of a risk factor (Levin 1953). Both parameters are important, and we have shown that they usually operate in opposite directions.

The major findings of these analyses are that (1) most risk factors for common chronic diseases with relative risks of, say, ~2 or ~3 have, per se, low PPV and PAF for disease; (2) the introduction of genetic testing can, under certain conditions of genotype–risk factor interactions, markedly increase the ability of the risk factor to predict disease but can decrease the ability to prevent more cases in the population—this would occur if preventive measures are applied only to individuals with susceptibility genotypes; (3) a high prevalence of a genotype can decrease markedly the ability of the genotype to predict disease among individuals with a risk factor (but can increase PAF)—this is especially true for rare diseases in the population; (4) the same can be said about risk factors with high prevalence—the higher the prevalence, the lower is their ability to predict disease (but the higher is their PAF); (5) when risk factors are associated with increased risk of disease even in the absence of the susceptibility genotype (type 2), limiting intervention to those with such genotypes will have less effect on disease risk in the population than will intervening in all those with the risk factor, thus limiting the value of this approach; and, finally, (6) while using genetic testing among individuals with disease risk factors can increase disease PPV to varying degrees, it is unlikely that it can ever fully predict disease risk. Given the complexities of the etiology and pathogenesis of common human diseases, it is likely, except under specific conditions, that one can only provide estimates—never certainty—of disease risk associated with a risk factor if an underlying susceptibility genotype is present. This finding is in line with Motulsky's (1994, p. 603) recent editorial on predictive genetic diagnosis, in which he applies the term "susceptibility diagnosis" and states that "in such conditions the results of the test do not provide an unequivocal answer; they only give information on whether an increased risk exists. The results are probabilistic."

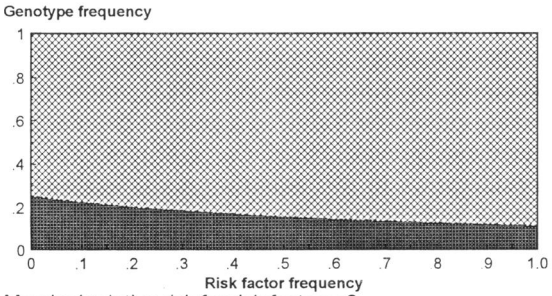
The impact of a high prevalence of the genotype on the PPV of a risk factor deserves further comments. It has long been suggested that, the more common a disease-susceptibility genotype, the more its public health impact, and the more desirable and successful a genetic screening program could be in identifying individuals at high risk (Omenn 1982). As shown earlier, for a given disease prevalence and a given disease–risk factor association, the more common the underlying interacting genotype, the lower is its disease penetrance, and the worse is its usefulness in improving the predictive value of the risk factor in question. Very frequent alleles may thus be not very suitable to improve the predictive value of risk factors, especially in the face of rare conditions.

These findings are consistent with previous analyses presented by Khoury et al. (1985), showing how the disease-predictive value of a genetic test decreases with increasing prevalence of the genotype. An important consideration in this equation is the interplay between disease and genotype frequencies. When the genotype is more common than the disease per se, such a genotype will have a more limited ability to predict disease, although it may account for more cases in the population (attributable fraction). To illustrate this point, consider the results of a recent study by Ward et al. (1993), in which a molecular variant of angiotensinogen T235 was associated with sevenfold increased risk of preeclampsia. It seems that this marker could account for a substantial proportion of cases in this study (50%–80% attributable fraction). On the other hand, two-thirds of the population may have this marker (by being either homozygous or heterozygous). Given that preeclampsia occurs in $\leq 5\%$ of pregnancies, such a genetic marker will have a low predictive value, per se, for the condition, precluding its usefulness in a clinical predictive setting. It is also unlikely to improve markedly the predictive value of any other risk factor (such as primigravidity) for preeclampsia.

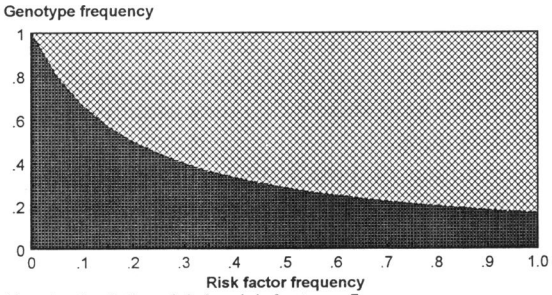
It is important to discuss the impact of the pattern of genotype–risk factor interaction on the ability to improve the disease PPV of a risk factor. Under the extreme type 1 interaction effect, in which the risk-factor effect is limited only to individuals with an underlying susceptibility genotype, the PPV of the risk factor can be tremendously increased if the underlying genotype is measured. On the other hand, with less extreme forms of interaction—in particular, when there are straightforward multiplicative joint effects of risk factor/genotype—there will be less gain in improving the PPV of a risk factor by performing genetic testing, especially in the face of declining PAF values.

The analyses presented here are, of course, limited by the scope of the examples presented and by the underlying assumptions of interaction. Biological effects are likely to be more complex, and dose-response issues will be relevant both for risk factors and for genotypes. Modes of interaction between genotypes and exposures may be neither additive nor multiplicative, and they would be difficult to illustrate in this paper. Also, the models of interaction presented here are relatively extreme ones and depart from the usual "statistical" interpretation. Finally, multiple effects of risk factors should be considered. One could envision situations in which

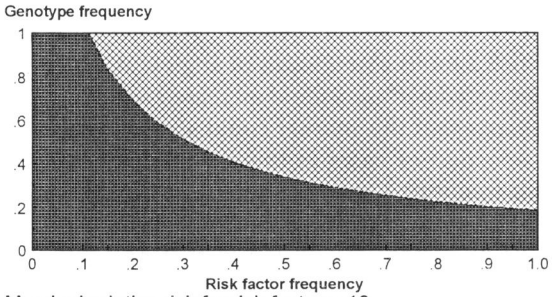
Disease frequency = .1



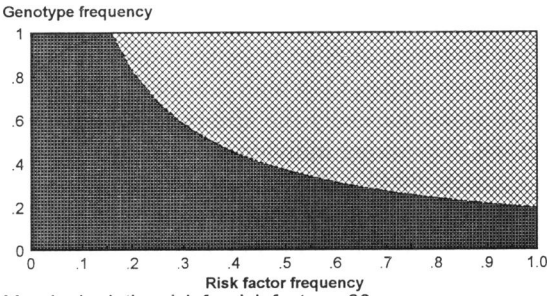
Marginal relative risk for risk factor = 2



Marginal relative risk for risk factor = 5

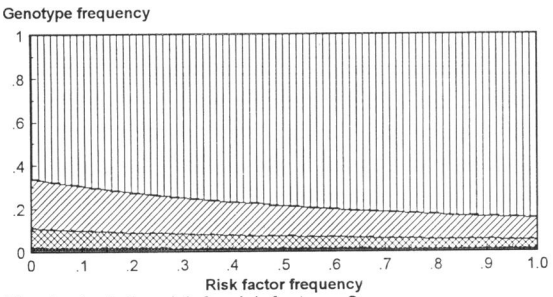


Marginal relative risk for risk factor = 10

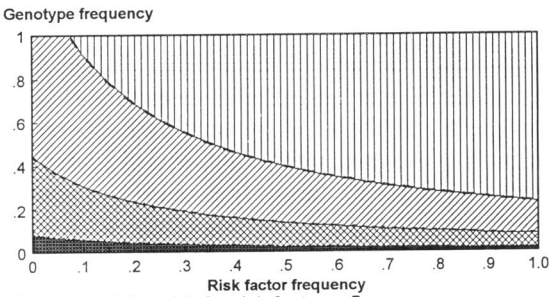


Marginal relative risk for risk factor = 20

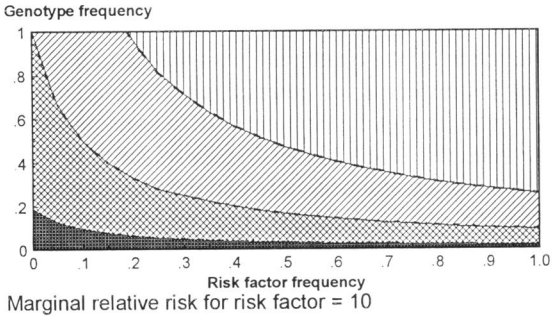
Disease frequency = .01



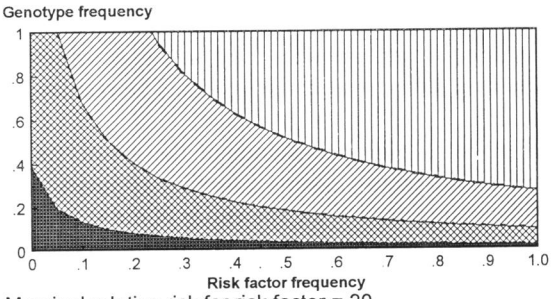
Marginal relative risk for risk factor = 2



Marginal relative risk for risk factor = 5



Marginal relative risk for risk factor = 10



Marginal relative risk for risk factor = 20

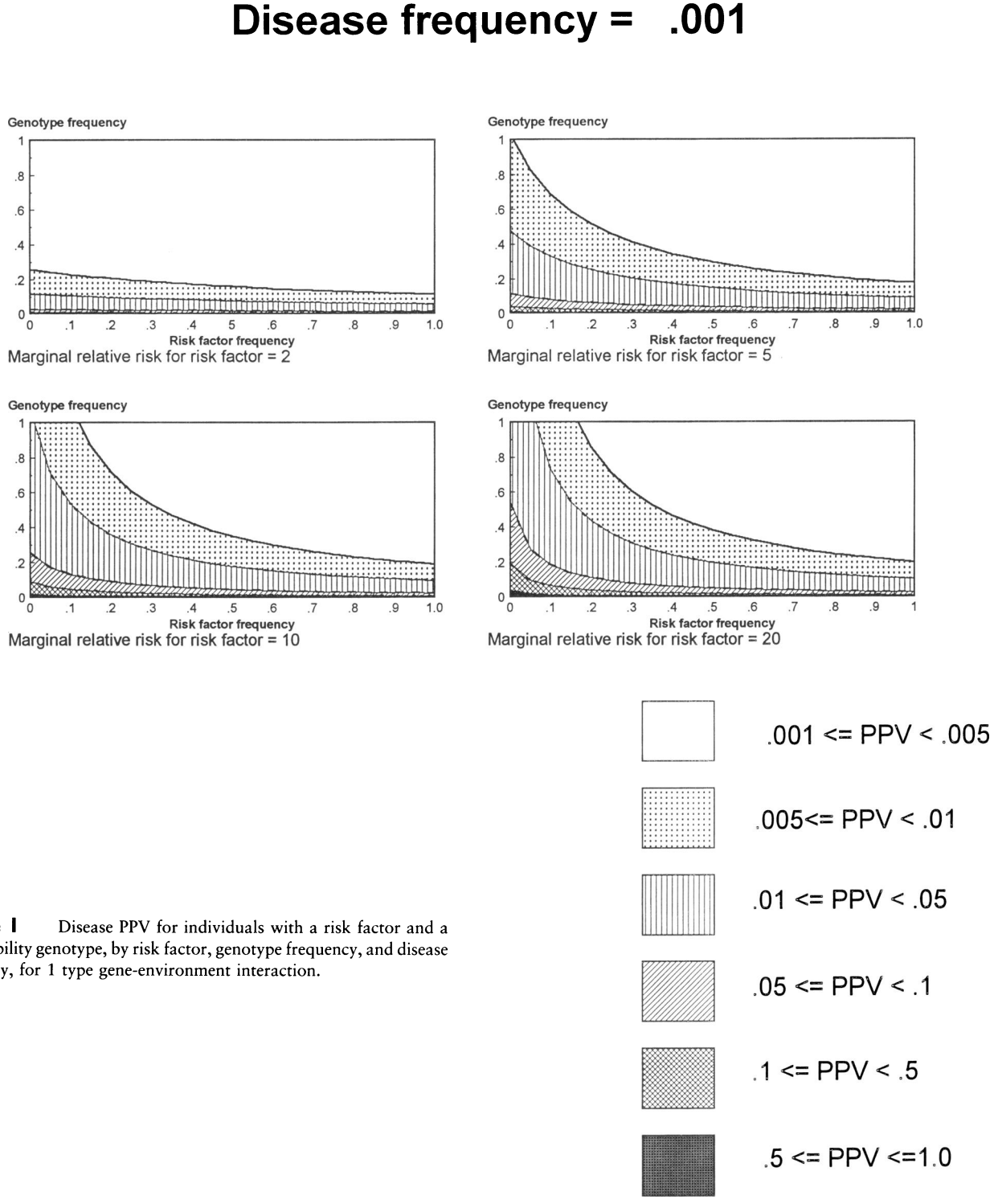


Figure 1 Disease PPV for individuals with a risk factor and a susceptibility genotype, by risk factor, genotype frequency, and disease frequency, for 1 type gene-environment interaction.

a genotype influences the risk of one disease associated with an exposure (e.g., lung cancer from smoking) but does not influence the risk of another disease associated with the same exposure (e.g., heart disease from smoking).

Two methodological issues, confounding and genotypic misclassification, must be considered in using genetic marker information to improve disease PPV of risk factors. Confounders could be other, unmeasured genetic determinants or environmental factors that could produce spurious associations with disease risk. Racial, ethnic, and other sources of population stratification are

Table 3

PAF (%) among Individuals with a Risk Factor and a Susceptibility Genotype, by Prevalence of Genotype and Type of Genotype–Risk Factor Interaction

	Type of Interaction				
Prevalence of Genotype	Marginal Effect	Multiplicative	Type 1	Type 2	Type 3
1%	9.1	.8	9.1	4.6	9.4
5%	9.1	3.3	9.1	4.8	10.5
10%	9.1	5.5	9.1	5.0	11.5
25%	9.1	9.1	9.1	5.6	12.6
50%	9.1	10.7	9.1	6.7	12.2
75%	9.1	10.3	9.1	7.9	10.9
100%	9.1	9.1	9.1	9.1	9.1

NOTE.—See footnotes to table 2.

major sources of confounding. One clear example of confounding is the association between the genetic marker $Gm^{3;5;13;14}$ and non–insulin-dependent diabetes among the Pima Indians (Knowler at al. 1988). In this cross-sectional study, individuals with this Gm haplotype had a higher prevalence of diabetes than did those without this marker (29% vs. 8%). However, this particular genetic marker turned out to be inversely correlated with White admixture. When the analysis was stratified by the degree of White admixture, the observed association all but disappeared.

Another methodological concern is genotypic misclassification. Indirect methods are often used to assign individuals' genotypes, which can lead to genotypic misclassification. When genotypes are measured at the DNA level, misclassification can occur because of linkage disequilibrium. Under ideal conditions, if the gene of interest has been completely sequenced, the presence and location of one or more mutations within the gene could be correlated with an altered gene product and then with disease risk in epidemiological studies. However, more often than not, we only have markers in either the region of the candidate gene or a nonexpressed portion of the gene. Unless the actual site(s) of a deleterious mutation involved in the disease is targeted, it may well be that an association between a marker allele and the disease could reflect linkage disequilibrium between the measured allele and the actual susceptibility allele(s) associated with the disease (Khoury and Beaty 1994). The angiotensinogen T235 variant alluded to earlier in relation to the risk of preeclampsia may be such a marker. This would dilute the magnitude of the estimated association (relative risk) between the marker allele and the disease, toward the null, and would underestimate the effect of the locus in predicting disease risk. With increasing ability to directly identify disease-related genes and mutations, it will be possible to use more "refined" genetic testing as part of this approach to disease-predictive value. Finally, it is important to keep in mind complications resulting from genetic heterogeneity at the molecular level. If a genetic test does not detect all susceptibility alleles, the advantages of this approach will be reduced.

Let us address our original question: With advancing genetic technology, will it be possible to refine the risk-factor approach to target preventive measures to individuals with risk factors who also carry disease-susceptibility allele(s)? The answer to the question will depend on numerous factors in addition to the epidemiological considerations presented earlier. Criteria to be discussed before implementing this approach include the validity and reliability of the available genetic test, the public health impact of the disease, the magnitude of the association between the genotype and the disease—and of the interaction with known and modifiable environmental risk factors, the availability of safe and effective intervention/prevention strategies, as well as ethical considerations needed to insure individuals' autonomy and confidentiality (Khoury et al. 1993; authors' unpublished data).

There are mounting concerns regarding the misuse of genetic information in society (Holtzman 1989; Suzuki and Knudston 1989; Hubbard and Wald 1993; Garver and Garver 1994). One potential drawback of using genetics to refine the predictive value of risk factors is to shift attention from the risk factor for the disease, to stigmatize and discriminate against individuals' susceptibility genotypes (Holtzman 1989). Examples here include possible genetic susceptibility to the effects of cigarette smoking and occupational hazardous exposures (Omenn 1982). On the other hand, some relevant prevention examples are currently discussed in medical and public health settings: (i) Given the discussion as to whether mammography is cost-effective at a relatively

young age (40–50 years), will it be possible to use genetics to find out who are the women at highest risk of breast cancer, so that early intervention and detection can be done (King et al. 1993; NACHGR 1994)? This discussion is compounded by the concerns that low-dose radiation from mammography may be more harmful to individuals with genetic susceptibility and that the sensitivity of mammography in young women may be reduced because of breast tissue density. (ii) Given the discussion about dietary habits and the difficulties in changing people's diets, will it be possible to use genetics to find out who are the individuals who will lower their risk of coronary heart disease by aggressive dietary intervention (Allred 1993; Simopoulus et al. 1993)? This issue is compounded by our lack of knowledge as to whether individuals at highest risk are willing, as a result of genetic testing, to change dietary habits. (iii) Given the current debate on food fortification with folic acid to prevent neural tube defects, a debate driven by possible adverse effects of masking B_{12}-deficiency pernicious anemia, will it be possible to use genetics to find out who are the women at risk of neural tube defects and who will thus benefit from periconceptional supplementation with folic acid (CDC 1992)?

These are just some examples of the potential usefulness of using genetics in refining the risk-factor approach in disease prevention. As society deals with ethical issues related to predicting disease risks according to genetic background, ethical safeguards should include protection of individual autonomy and the right to decide on the basis of a proper informed-consent process, preservation of confidentiality of results of genetic testing, limiting genetic testing in the workplace and by insurance companies, scientific evaluation of both the ability of genetic tests to measure the underlying susceptible genotype and the evidence for increased disease risks from certain risk factors, and proper education of the medical profession and the general public regarding the importance of genetics in the practice of clinical and preventive medicine (Holtzman 1989).

In conclusion, the use of genetic tests is likely to improve the disease-predictive value of risk factors. Therefore, a new paradigm in the primary prevention of many chronic diseases could be the identification and interruption of environmental cofactors that lead to clinical disease among persons with susceptibility genotypes. For most diseases and genetic risk factors, such cofactors are still poorly understood, and a lot of work, particularly population-based epidemiological studies, is needed before the results of basic genetic research are translated into population-based interventions. With continued advances in genetic technology, the medical community, public health organizations, and society should collectively consider the applicability of this new paradigm in disease prevention.

Acknowledgement

The authors thank the members of the Genetics Working Group at the Centers for Disease Control and Prevention for helpful comments.

References

Allred JB (1993) Lowering serum cholesterol: who benefits? J Nutr 123:1453–1459

Badimon JJ (1993) Coronary atherosclerosis: a multifactorial disease. Circulation 87 Suppl 2:3–16

Balant LP, Balant-Gorgia AE (1992) Clinical relevance of pharmacogenetics rediscovered. Clin Ther 14:328–330

Baron J (1989) The clinical utility of risk factor data. J Clin Epidemiol 42:1013–1020

Caporaso N, Hayes RB, Dosemeci M, Hoover R, Ayesh R, Hetzel M, Idle J (1989) Lung cancer risk, occupational exposure and the debrisoquine metabolic phenotype. Cancer Res 49:3675–3679

CDC (Centers for Disease Control and Prevention) (1992) Recommendations for the use of folic acid to reduce the number of cases of spina bifida and other neural tube defects. MMWR Morb Mortal Wkly Rep 41 Recommendations 14:1–7

Garver KL, Garver B (1994) The Human Genome Project and eugenic concerns. Am J Hum Genet 54:148–158

Glader BE (1990) Hereditary red blood cell disorders (excluding hemoglobinopathies and thalassemias). In: Emery AEH, Rimoin DL (eds) Principles and practice of medical genetics, 2d ed. Vol 2. Churchill-Livingstone, Edinburgh, pp 1343–1369

Hamsten A (1993) Lipids as a coronary risk factor: analysis of hyperlipidaemias. Postgrad Med J 69 Suppl 1:S8–S11

Hegele RA (1992) Gene-environment interactions in atherosclerosis. Mol Cell Biochem 113:177–186

Hoffman EP (1994) The evolving genome project: current and future impact. Am J Hum Genet 54:129–136

Holtzman NA (1989) Proceed with caution: predicting genetic risks in the recombinant DNA era. Johns Hopkins University Press, Baltimore

——— (1992) The diffusion of new genetic tests for predicting disease. FASEB J 6:2806–2812

Hubbard R, Wald E (1993) Exploding the gene myth. Beacon, Boston

Khoury MJ, Adams MJ Jr, Flanders WD (1988) An epidemiologic approach to ecogenetics. Am J Hum Genet 42:89–95

Khoury MJ, Beaty TH (1994) Application of the case-control method in genetic epidemiology. Epidemiol Rev 16:134–150

Khoury MJ, Beaty TH, Cohen BH (1993) Fundamentals of genetic epidemiology. Oxford University Press, New York

Khoury MJ, Newill NA, Chase GA (1985) Epidemiologic evaluation of screening for risk factors: application to genetic screening. Am J Publ Health 75:1204–1208

King MC, Rowell S, Love SM (1993) Inherited breast and ovarian cancer: what are the risks? what are the choices? JAMA 269:1975–1980

Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988)

$Gm^{3;5;13;14}$ and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 43:520–526

LaRosa JC (1992) Cholesterol and cardiovascular disease: how strong is the evidence? Clin Cardiol 15 Suppl 3:2–7

Levin ML (1953) The occurrence of lung cancer in man. Acta Union Int Contra Cancer 9:531–541

Motulsky AG (1994) Predictive genetic diagnosis. Am J Hum Genet 55:603–605

NACHGR (National Advisory Council for Human Genome Research) (1994) Statement on use of DNA testing for presymptomatic identification of cancer risk. JAMA 271:785

Omenn GS (1982) Predictive identification of hypersusceptible individuals. J Med 24:369–374

Scriver CR, Kaufman S, Woo SLC (1989) The hyperphenylalaninemias. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic basis for inherited disease, 6th ed. Vol 1. McGraw-Hill Information Services, New York, pp 495–546

Simopoulos AP, Herbert V, Jacobson B (1993) Genetic nutrition: designing a diet based on your family medical history. MacMillan, New York

Stehbens WE (1992) Causality in medical science with particular reference to heart disease and atherosclerosis. Perspect Biol Med 36:97–119

Suzuki D, Knudston P (1989) Genethics: the clash between the new genetics and human values. Harvard University Press, Cambridge, MA

Ward K, Hata A, Jeunemaitre X, Helin C, Nelson L, Namikawa C, Farrington PF, et al (1993) A molecular variant of angiotensinogen associated with preeclampsia. Nat Genet 4:59–61

Woodard DA (1993) The impact of diet on coronary heart disease. Med Clin North Am 77:849–862