May/June 2000 · Vol. 2 · No. 3

On the use of population-based registries in the clinical validation of genetic tests for disease susceptibility

Quanhe Yang, PhD[1], Muin J. Khoury, MD, PhD[2], Steven S. Coughlin, PhD[3], Fengzhu Sun, PhD[4], and W. Dana Flanders, MD, ScD[5]

Purpose: Many new genetic tests for susceptibility to adult-onset diseases are developed on the basis of selected and high-risk groups. Before such tests can be used in medical practice, however, epidemiologic studies must be conducted to evaluate their clinical sensitivity, specificity, and positive predictive value in the general population. For many common adult-onset diseases, this process may take decades of follow-up. **Method:** We illustrate how clinical validation of new predictive genetic tests can be done retrospectively using case-control studies that are derived from population-based registries of diseases. We use the examples of birth defects and cancer registries to illustrate a hypothetical process by which such tests can be clinically validated. **Results:** We demonstrate how such epidemiologic studies can be successfully used to derive measures of a test's sensitivity, specificity, positive predictive value, negative predictive value, and of the population attributable fraction of disease due to the disease-susceptibility genes. Under certain assumptions, data derived from population-based case-control studies provide adequate estimates of lifetime risks for disease (penetrance) among people with specified genotypes. **Conclusions:** With adequate protections of human subjects, studies involving population-based registries of disease will increasingly become valuable in validating the numerous genetic tests that will emerge from advances in human genetic research and the Human Genome Project. ***Genetics in Medicine*, 2000;2(3):186–192**

Key Words: *Genetic testing, registries, case-control study, epidemiologic methods*

The rapid pace of genetic discoveries has resulted in the proliferation of genetic tests for a wide variety of diseases. Currently, DNA-based tests are available for more than 600 conditions,[1] and more than 6,000 conditions have been mapped to specific chromosomes.[2] Genetic tests of susceptibility to common adult-onset disorders in the general populations have become increasing available. Examples include tests for BRCA1/BRCA2 gene mutations in relation to breast and ovarian cancer,[3–5] tests for apolipoprotein E-E4 allele in relation to Alzheimer disease,[6,7] and tests for hereditary prostate cancer 1 (HPC1) in relation to prostate cancer.[8,9]

Before such genetic tests can be used in medical practice, their clinical validity needs to be established.[10] Determining the clinical validity of the tests involves measuring their clinical sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) with respect to disease occurrence.[11] Ideally, estimates of these clinical measurements of genetic testing would come from large-scale population-based cohort studies; such studies, however, will take considerable time and resources. In this paper, we illustrate how clinical validation of new predictive genetic tests can be done retrospectively using incident case-control studies that are derived from population-based registries of diseases. We demonstrate how such epidemiologic studies can be successfully used to measure the clinical sensitivity, specificity, PPV, and NPV of such tests, as well as the population attributable fraction of disease due to the disease-susceptibility genes (i.e., the proportion of cases in the population attributable to a positive genetic test—an adjusted measure of clinical sensitivity). Under certain assumptions, data derived from population-based case-control studies provide adequate estimates of lifetime risk that people with specific genotypes have for common diseases (penetrance).

By genetic testing, we mean the analysis of a specific gene, its product, or function, or other DNA and chromosome analysis that is done to detect or exclude an alteration likely to be associated with a genetic disorder.[12] Others have provided a more detailed definition of genetic testing.[13] Here, we focus on genetic tests for susceptibility to common complex diseases in humans: the term "complex" disease implies that the inheritance of an allelic variant or even combination of alleles at multiple loci may increase risk but does not always result in the

From the [1]Birth Defects and Pediatric Genetics Branch, Centers for Disease Control and Prevention; [2]Office of Genetics and Disease Prevention, Centers for Disease Control and Prevention; [3]Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention; [4]Department of Mathematics, University of Southern California; [5]Department of Epidemiology, School of Public Health, Emory University, Atlanta, Georgia.

Dr. Quanhe Yang, Birth Defects and Pediatric Genetics Branch, Centers for Disease Control and Prevention (CDC), Mailstop F-45, 4770 Buford Hwy, Atlanta, GA 30341.

Received: October 25, 1999.

Accepted: February 19, 2000.

disease (incomplete penetrance). The primary objective of such testing is to determine the probability of developing a disease given a positive genetic test result. Such genetic tests for susceptibility differ from diagnostic genetic tests or population screening program for certain genetic diseases. The measurements used in the clinical validation of genetic testing can be applied to tests for various diseases with different ages of onset, such as birth defects, cancers, and cardiovascular disorders. In estimating lifetime risk for disease among carriers of disease-susceptibility genes, we focus on common diseases with varying age of onset.

We use the examples of population-based birth-defect and cancer registries to illustrate a hypothetical process of clinically validating maternal MTHFR C677T mutation testing as a means of estimating child's risk for neural tube defects (NTDs) and BRCA1 testing as a means of estimating a woman's risk for breast cancer.

METHODS

The clinical validity of a genetic test reflects its ability to correctly classify people who have or will develop disease as test-positive and those who will not develop disease as test-negative. Measurements of validity include sensitivity, specificity, PPV, and NPV. Here, sensitivity is defined as the probability of a positive genetic test result among people who will develop disease, specificity is the probability of a negative genetic test results among people without the disease, PPV is the probability of developing disease among people with a positive test result, and NPV is the probability of not developing disease among people with a negative test result.[14]

The parameters needed to estimate these measurements can be obtained from a population-based registry and a case-control study derived from such a registry.

Population-based registries

Population-based registries involve a continuous and systematic process of collection, analysis, interpretation, and dissemination of descriptive information for the disease concerned as illustrated by birth defects and cancer registries.[15–18] For most population-based registries, the population usually represents residents within a specified geographic area, which may be a city, region, or nation. The primary objective of a population-based registry is usually to collect information on all new cases of the disease over time within the defined geographic region and monitor changes over time. The data collected from these registries provide basic information about the descriptive epidemiology of the health problems, such as the overall rate of the disease in the population and its change over time. Such information can also be used to evaluate intervention programs and to help formulate policy decision regarding health care and resource allocation.[19] For the clinical validation of genetic testing, population-based registries can provide information about the overall disease rate in the population as well as serve as a framework for population-based case-control studies.

Population-based case-control study

Although population-based registries collect all cases of a disease with basic demographic information, they may not contain sufficient detailed information for in-depth epidemiologic studies. However, population-based registries can serve as a framework for investigators designing detailed epidemiologic research such as a population-based case-control study.

In a population-based case-control study, investigators ascertain all new cases of diseases diagnosed in a certain time interval in a defined population, and the control subjects are randomly chosen from the same underlying population.[20] The advantages of population-based case-control studies include the ability to quantify the magnitude of disease risks in the underlying population, to estimate the population attributable risk for the disease due to specified risk factors, and to estimate the absolute risk for disease.[21] For the clinical validation of genetic testing, such studies can be used to derive the sensitivity, specificity, PPV, and NPV of such tests. Under certain assumptions, one can also use data from case-control studies to estimate the life time risk (penetrance) of the common diseases with varying age of onset.

Estimations of sensitivity, specificity, PPV, and NPV

For simplicity, we assume a single disease-susceptibility gene with two alleles. The estimations also apply to multiple alleles at multiple loci. If one views the result of genetic testing of disease-susceptibility genotype as a risk factor (positive or negative on the test), then estimating the probability of developing disease among people with disease-susceptibility genotype (PPV) can be done by estimating exposure-specific incidence rates using data from a case-control study.[22,23] Table 1 shows how to estimate sensitivity, specificity, PPV, and NPV. For conditional probability, Bayes' theorem states that:

$$P_k(D/G) = \frac{P_k(G/D)P_k(D)}{[P_k(G/D)P_k(D) + P_k(G/\bar{D})(1 - P_k(D))]} \qquad (1)$$

where $P_k(D/G)$ = the annual risk of disease among people with a positive genetic test result for the disease-susceptibility gene in the *k*th age interval, which is, in other words, PPV; $P_k(G/D)$ = the probability of a positive test result for the disease-susceptibility gene among case subjects in the *k*th age interval; $P_k(G/\bar{D})$ = the probability of a positive test result among control subjects in the *k*th age category; and $P_k(D)$ = the prior probability of disease (overall rate of disease) among people in the *k*th age interval. To simplify, we assume $P_k(G/\bar{D})$ is constant across age. For most diseases, the annual risk and the incidence rate per person year will be nearly the same numerically.

Similarly, one can estimate NPV as follows:

$$P_k(\bar{D}/\bar{G}) = \frac{P_k(\bar{G}/\bar{D})(1 - P_k(D))}{P_k(\bar{G}/\bar{D})(1 - P_k(D)) + P_k(\bar{G}/D)P_k(D)} \qquad (2)$$

where $P_k(\bar{D}/\bar{G})$ = the annual risk of disease among people with a negative genetic test result for the disease-susceptibility gene in the *k*th age interval (NPV); and $P_k(\bar{G}/D)$ = probability of a

Table 1
Estimation of a genetic test's "sensitivity," "specificity," positive predictive value (PPV), and negative predictive value (NPV) on the basis of data derived from a population-based case-control study

	Disease Status	
Test Results	Case	Control
Positive	a	b
Negative	c	d
	N_1	N_2

a = the number of people who test positive and have the disease.
b = the number of people who test positive but do not have the disease.
c = the number of people who test negative but have the disease.
d = the number of people who test negative and do not have the disease.
"Sensitivity" = $P(\bar{G}/\bar{D}) = a/N_1$: the probability that people with disease test positive. "Specificity" = $P(\bar{G}/\bar{D}) = d/N_2$: the probability that people without disease test negative.

$$PPV = P_k(D/G) = \frac{P_k(G/D)P_k(D)}{[P_k(G/D)P_k(D) + P_k(G/\bar{D})(1-P_k(D)]} \quad \text{OR} \quad \frac{\psi_k P_k(D)}{[P_k(G)(\psi_k - 1) + 1]}$$

PPV is the probability of developing disease in the kth age interval given a positive genetic test result (see text for estimation of PPV).

$$NPV = P_k(\bar{D}/\bar{G}) = \frac{P_k(\bar{G}/\bar{D})(1 - P_k(D))}{P_k(\bar{G}/\bar{D})(1 - P_k(D)) + P_k(\bar{G}/D)P_k(D)}$$

NPV is the probability of not developing disease in the kth age interval given a negative genetic test result (see text for estimation of NPV).

ψk = the risk ratio (odds ratio) of having disease-susceptibility gene in the kth age interval.

negative test result for the disease-susceptibility gene among case subjects in the *k*th age interval.

It is important to note that the probability of people with the specified genotype developing disease is age dependent. For diseases expressed at birth, such as external structure malformations, $k = 0$; for all other diseases, *k* represents age at disease onset. If there is no matching or other important biases in subjects selection, one can estimate the test (genotype) probability among case and control subjects from the case-control study as $P_k(G/D) = a/N_1$, and $P_k(G/\bar{D}) = b/N_2$ (Table 1). Thus, $P_k(G/D)$ approximates the sensitivity of a genetic test, and $P_k(\bar{G}/\bar{D})$ approximates the specificity of a genetic test in a follow-up cohort study. In this sense, we define $P_k(G/D)$ and $P_k(\bar{G}/\bar{D})$ as "sensitivity" and "specificity" of the test. For overall rate of disease in the population, $P_k(D)$, the numerator represents the number of incidence cases in a specific time interval, and the denominator is the size of the population at risk. One may get these overall disease rates from population-based registries, for example, using data from birth defects surveillance programs or from the Surveillance, Epidemiology, and End Results Program (SEER) for cancer incidence rates in the general population.

If estimates of risk ratio for disease among carriers of the disease-susceptibility gene and the prevalence of disease-susceptibility gene in the population are available, the estimate of PPV can also be obtained by:

$$P_k(D/G) = \frac{\psi_k P_k(D)}{[P_k(G)(\psi_k - 1) + 1]} \tag{3}$$

where ψ_k = risk ratio for disease among carriers of the disease-susceptibility gene in the *k*th age interval; and $P_k(G)$ = the prevalence of the disease-susceptibility gene in the *k*th age interval in a population. The estimations of "sensitivity" and "specificity" can also be expressed in terms of ψ_k, $P_k(G)$ and $P_k(D)$.[14]

In addition, using data from a population-based case-control study, one can estimate the population attributable fraction, which is the proportion of all new cases in a given period that were identified by a positive genetic test result. Assuming no confounding of genetic test-disease association, one can estimate attributable fraction using the following formula:[24]

$$AF_k = \frac{f_k(\psi_k - 1)}{[1 + f_k(\psi_k - 1)]} \tag{4a}$$

$$\psi_k = \frac{AF_k(1 - f_k) + f_k}{f_k(1 - AF_k)} \tag{4b}$$

where AF_k is the population attributable fraction for age group k, f_k is the proportion of people in age group k with the disease-susceptibility gene in the population, and ψ_k is risk of having the disease in question among those who have the disease-susceptibility gene compared with those who have no disease-susceptibility gene. Other formulas are also available for estimating the population attributable fraction.[25,26]

Estimation of lifetime risk (penetrance) of common diseases

Formulas 1 and 3 provide the point estimates of the genotype-specific incidence rate by age; these estimates can be used to approximate the penetrance of the disease-susceptibility for diseases expressed at birth, such as birth defects, where $k = 0$. For common complex diseases with varying age at onset, we are interested in the cumulative risk over certain time intervals of the disease-susceptibility gene among people with positive test results, for example, the cumulative risk of BRCA1-gene mutation carriers for breast cancer by the different age group. To calculate cumulative risk, one can apply either formula 1 or 3, depending on the availability of data, to each age group to get separate estimates of the age-genotype-specific incidence rate, and calculate the cumulative incidence (risk):

$$R_j = 1 - \exp[-\sum_{k=1}^{K} ID_{kj} * \Delta t_k] \tag{5}$$

where R_j is the cumulative risk for disease (penetrance) among carriers of the disease-susceptibility genotype by the *k*th age group. ID_{kj} is the age-genotype-specific incidence rate in the *k*th age group (i.e., j = 0 for carriers of the nonsusceptibility genotype and j = 1 for carriers of the disease-susceptibility genotype), and Δ_{tk} is the width of age interval. We estimate ID_{kj} as $P_k(D/G)$ for j = 1 and $P_k(D/\bar{G})$ for j = 0, assuming 1 year risk, $P_k(D/G)$, is rare. If we do not take into account other risk factors or competing risks, the cumulative risk for disease

(R_j) is equivalent to the penetrance of the disease-susceptibility genotype for diseases with varying ages of onset.

When the study is stratified by one or more risk factors such as smoking status, the above formulas may be applied to different strata to get stratum-genotype-specific estimates. The difference of estimates for different strata may provide some evidence of gene-environment interaction. Although Bayes' theorem is used mainly in prevalence and cumulative incidence studies, it also provides a good approximation in incidence density studies.[20]

Examples

We present two examples of how data from population-based case-control studies can be used to estimate the "sensitivity," "specificity," PPV, and NPV of genetic tests, and to estimate the lifetime risk for disease (penetrance) among people with a disease-susceptibility gene. In one example, we use data derived from a population-based case-control study of the association between homozygosity for 5,10-methylenetetrahydrofolate reductase (MTHFR) gene C677T mutation and risk for neural tube defects (NTDs). In the other example, we created a case-control data set to estimate, by age group, the penetrance of BRCA1-gene for breast cancer among carriers.

MTHFR and NTDs

Several recent studies reported an association between homozygosity for the MTHFR gene C677T mutation and risk for NTDs.[27–30] In a meta-analysis of that association, Botto and Yang[31] found a pooled odds ratio of 1.73 (95% CI 1.39–2.16) among infants who were homozygous for MTHFR C677T variation and an odds ratio of 1.15 (95% CI 0.98–1.34) among those who were heterozygous for the mutation. Shaw *et al.*[30] investigated the joint effects of having MTHFR C677T genotype and maternal use of supplements containing folic acid on an infant's risk for NTDs in a population-based case-control study conducted in California (1987–1991 birth cohorts). The study genotyped the allelic variants of MTHFR in 214 liveborn case infants with spina bifida and 503 control infants without birth defects. Of the 214 case infants, 41 had two alleles for the mutation, 100 had one allele, and 73 had none; of the 503 control infants, the corresponding figures were 72, 213, and 218 with a estimated allele frequency of 35.5% (95% CI 32.5% to 38.5%) in the population. Table 2 shows our estimates of the "sensitivity," "specificity," PPV, NPV, and population attributable fraction of infants for MTHFR gene C677T mutation for risk of NTDs. The estimated prevalence of NTDs is approximately 1/1,000 in the general population. As shown in Table 2, if homozygosity for the C677T variant is used to measure risk for NTDs, the "sensitivity" and PPV of such test are low, 19.2% and 0.14% respectively, and the NPV is high (99.9%). The estimated population attributable fraction of NTDs due to homozygosity for the MTHFR gene C677T mutation is approximately 7.9%.

BRCA1 and breast cancer

To ensure sufficient numbers of case and control subjects for the second example, we used the results of Whittemore *et al.*[32] to calculate the parameters necessary to estimate PPV, NPV, and lifetime risk for disease (penetrance). In that study, Whittemore *et al.*, using data from three U.S. population-based case-control studies, estimated that the percentages of breast cancer cases attributable to BRCA1/BRCA2 mutations were 11.2%, 10.7%, 8.6%, and 5.8% for women aged 15–29, 30–39, 40–49, and 50–59 years, respectively, and the estimated carrier prevalence of BRCA1 mutations in the U.S. population is approximately 0.0029 ($P(G/\bar{D})$). Another study suggested that approximately 60% of BRCA1/BRCA2 mutation carriers were BRCA1 carriers.[33] Therefore, we multiplied the estimates of Whitemore *et al.*[32] estimates of the percentages of breast cancer attributable to BRCA1/BRCA2 mutations by 0.6 to estimate the percentages attributable to BRCA1 only ($P_k(G/D)$). We used Equation 4b to estimate age-specific risk ratios (ψ_k) and assumed that the proportion of people with the disease-susceptibility gene in the population was constant (f_k = 0.0029 for all k). For overall disease rates $P_k(D)$, we used the SEER 1988 breast cancer incidence rate.[34] With these estimated parameters [$P_k(D)$, ψ_k, and $P_k(D)$], we then apply Equation 3 to estimate age-specific exposure rates ($P_k(D/G)$). Finally, we used Equation 5 [cumulative risk for disease (penetrance)] to estimate penetrance of breast cancer among BRCA1 mutation carriers (Table 3). Figure 1 presents our estimates of BRCA1-gene's penetrance among BRCA1 carriers and the estimates of two other studies.[35,36] These estimates were comparable, although we used estimates of the percentages of BRCA1 mutation carriers among case and control subjects from a published study.[32] This example illustrates how the population-based registries provide the estimates of absolute risk, how the population-based case-control study provides estimates of relative risk, and how one can then use these values to estimate lifetime risk (penetrance). It should be pointed out that the main pur-

Table 2

Estimates of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and population attributable fraction of infant for MTHFR gene C677T mutation and risk for NTDs using data from a population-based case-control study, California

MTHFR C677T Alleles	NTDs	No NTDs	OR (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV %	NPV %	Pop. attrib. fraction %
Homozygous	41	72	1.7 (1.0–2.8)	19.2 (14.1–25.1)	85.7 (82.3–88.6)	0.14	99.9	7.9
Heterozygous	100	213	1.4 (0.97–2.0)	46.7 (40.0–53.7)	43.3 (39.0–47.8)	0.11	99.9	13.4
Noncarrier	73	218	Referent	–	–	–	–	–

Table 3
Estimated lifetime risk (penetrance) for breast cancer among BRCA1 mutation carriers

Age	Population Incidence* (case/10^5 yr) P_K (D)	Case P_K (G/D)	Case − P_k ($\bar{G}$/D)	Control % BRCA1 + P_k (G/$\bar{D}$)	Control Carrier − P_k ($\bar{G}$/$\bar{D}$)	Risk Ratio ψ_k	Age-specific exposure rate (cases/10^5 yr) P_k (D/G)	Cumulative risk (penetrance) R_j
25–29	7.5	6.7	93.3	0.29	99.7	25.8	180.3	1.0
30–34	25.2	6.4	93.6	0.29	99.7	24.6	579.9	3.7
35–39	63.8	6.4	93.6	0.29	99.7	24.6	1467.7	10.5
40–44	125.4	5.2	94.8	0.29	99.7	19.9	2367.4	20.5
45–49	197.8	5.2	94.8	0.29	99.7	19.9	3734.3	34.1
50–54	232.7	3.5	96.5	0.29	99.7	13.5	3033.1	43.4

*SEER 1988 breast cancer incidence rates.[34]
The proportion of case and control subjects who were BRCA1-gene mutation carriers was derived from the results of Whittemore *et al.*[32]

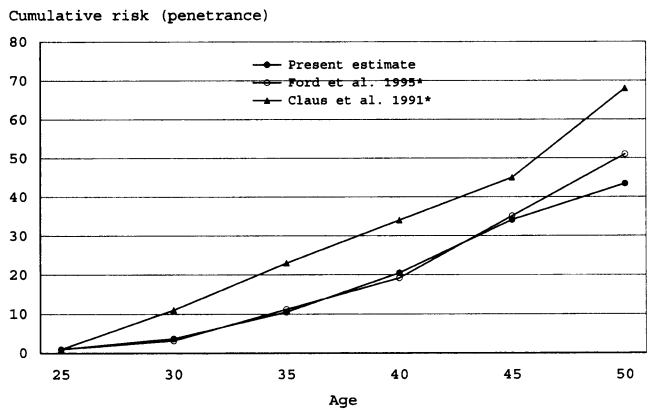

Fig. 1 Estimated cumulative risk (penetrance) for breast cancer among BRCA1-gene carriers by age group. *We linearly interpolated missing values for some age groups used in studies by Claus *et al.*[35] and Ford *et al.*[36]

pose of our estimations is to illustrate how the clinical validity of genetic testing can be calculated using data from population-based case-control studies, not the precision of such estimates.

DISCUSSION

We illustrated how to use data derived from population-based case-control studies in the clinical validation of genetic testing. We presented a simple yet valid method of estimating lifetime risks for disease among carriers of disease-susceptibility genes for common diseases with varying ages at onset. The rapid development of molecular genetics and the Human Genome Project will make genetic testing for various diseases increasingly feasible. Clinical validation of genetic testing is an important component to consider before such tests can be used in medical practice. Ideally, the clinical validation of such tests and the lifetime risk (penetrance) estimates of disease susceptibility will come from large-scale population-based cohort studies. However, in the absence of such studies, the simple approaches presented here may prove to be helpful in clinical validation of genetic tests and in determining the lifetime risk for disease (penetrance) among carriers of disease-susceptibility genes.

Other researchers who examined the relationship between genetic testing parameters (sensitivity, specificity, and PPV) and the frequency of disease-susceptibility genes, the frequency of the disease concerned, and the relative risk for the disease among carriers of disease-susceptibility genes concluded that the suitability of testing for a particular disease-susceptibility gene depends on the objectives of the testing.[10,13] For most of the common complex diseases, any single disease-susceptibility gene may increase a carrier's risk for disease only moderately, and the genotype frequency in the population may be relatively high.[37] Under these circumstances, the sensitivity and PPV of genetic tests will tend to be low for rare diseases such as birth defects and cancers.

We presented our examples using data derived from population-based birth defects and cancer registries, which are well known disease registries. Other established population-based registries include registries for diabetes, stroke, and communicable diseases.[38–42] Population-based registries for other common complex diseases, such as cardiovascular and mental disorders, are still relatively rare, which limits the application of the present method to other nonregistered diseases. In addition, some genes may increase carriers' risks for multiple health problems; for example, BRCA1 mutations are associated with breast and ovarian cancers, and p53 gene variants are associated with multiple primary cancers. Population-based registries may be available for some, but not all of these diseases.

As with any case-control design, the definition and selection of case and control subjects are crucially important to the validity of the study findings. Particularly, the effects of population stratification may invalidate attempts to clinically validate a genetic test on the basis of data from a case-control study. For example, if a disease-susceptibility gene occurs more often in a particular ethnic population, and this population also has relatively high risk for the disease, the association of disease with genetic markers may be overestimated. To reduce such a spu-

rious association between disease and genetic factors in a case-control study, the investigators should select the appropriate control subjects, possibly by matching control subjects with case subjects within major racial and ethnic subgroups; they should collect other cultural and anthropologic information and environmental factors for refined matching of case and control subjects; they should consider using DNA markers that characterize the genetic differences in subgroups of populations as the biological markers for population admixture.[21]

In estimating lifetime disease risk (penetrance), our approach did not take into account other risk factors or competing risks. Feuer *et al.*[43] presented an extended version of Equation 5 that accounts for competing risks. Although model-based estimates of exposure-specific incidence rate and cumulative risk estimates developed for other studies [44–47] can be used to make inferences about the general population, they are most useful when applied to assess individualized risk.[43,44] A number of genetic analysis models have also been used to estimate the penetrance of disease-susceptibility genes.[34,35] The calculations used in these approaches, however, are more complicated and require specialized genetic analysis programs.

Clinical validity is only one of the criteria used to develop safe and effective genetic testing. Others include analytical validity, which measures the ability of a test to predict correctly the underlying genotype, and clinical utility, which measures the benefits and risks to people with positive and negative testing results.[11] Other important issues in the development of safe and effective genetic testing include the quality of laboratories performing genetic tests; patients' psychological responses to genetic testing; and the ethical, legal, and social implications of genetic testing, which are discussed by many other studies.[10–13,48–55]

The rapid advances in human molecular genetics and then Human Genome Project seen in the past few years indicate that within the next decade genetic testing will be increasingly widely used in population screening for susceptibility of various common diseases, and in the diagnosis of diseases, and management of patients. Data derived from population-based case-control studies can help researchers assess the clinical validity of such genetic testing and to estimate the lifetime risk for disease (penetrance) among carriers of disease-susceptibility genotypes.

Acknowledgments

The authors thank Drs J. David Erickson and Adolfo Correa for their helpful comments.

References

1. McKusick VA. Mendelian inheritance in man: Catalogs of human genes and genetic disorders. 12th ed. Baltimore: Johns Hopkins University Press, 1998.
2. Center for Medical Genetics, Johns Hopkins University, and National Center for Biotechnology Information. OMIM: Online Mendelian Inheritance in Man. 1999. Available at: http://www.ncbi.nlm.nih.gov/omim/.
3. Ponder B. Genetic testing for cancer risk. *Science* 1997;278:1050–1054.
4. Shattuck-Eidens D, McClure M, Simard J, Labrie F, Narod S, Couch F, Hoskins K, Weber B, Castilla L, Erdos M, Brody L, Friedman L, Ostermeyer E, Szabo C, King MC, Jhanwar S, Offit K, Norton L, Gilewski T, Lubin M, Osborne M, Black D, Boyd M, Steel M, Ingles S, Haile R, Lindblom A, Olsson H, Borg A, Bishop DT, Solomon E, Radice P, Spatti G, Gayther S, Ponder B, Warren W, Stratton M, Liu QY, Fujimura F, Lewis C, Skolnick MH, Goldgar DE. A collaborative survey of 80 mutations in the BRCA1/BRCA2 breast and ovarian cancer susceptibility gene: Implications for presymptomatic testing and screening. *JAMA* 1995;273:535–541.
5. Hearly B. BRCA genes-booking, fortune telling and medical care. *N Engl J Med* 1997;336:1448–1449.
6. Post SG, Whitehouse PJ, Binstock RH, Bird TD, Eckert SK, Farrer LA, Fleck LM, Gaines AD, Juengst ET, Karlinsky H, Miles S, Murray TH, Quaid KA, Relkin NR, Roses AD, St George-Hyslop PH, Sachs GA, Steinbock B, Truschke EF, Zinn AB. The clinical introduction of genetic testing for Alzheimer diseases: An ethical perspective. *JAMA* 1997;277:832–836.
7. Morrion-Bogorad M, Phelps C, Buckholtz N. Alzheimer disease research comes of age: The pace accelerates. *JAMA* 1997;277:837–840.
8. Smith JR, Freije D, Carpten JD, Gronberg H, Xu J, Isaacs SD, Brownstein MJ, Bova GS, Guo H, Bujnovszky P, Nusskern DR, Damber JE, Bergh A, Emanuelsson M, Kallioniemi OP, Walker-Daniels J, Bailey-Wilson JE, Beaty TH, Meyers DA, Walsh PC, Collins FS, Trent JM, Isaacs WB. Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. *Science* 1996;274:1371–1374.
9. Gronberg H, Isaacs S, Smith JR, Carpten JD, Bova GS, Freije D, Xu J, Meyers DA, Collins FS, Trent JM, Walsh PC, Isaacs WB. Characteristics of prostate cancer in families potentially linked to the hereditary prostate cancer 1 (HPC1) locus. *JAMA* 1997;278:1251–1255.
10. Holtzman NA, Murphy PD, Watson MS, Barr PA. Predictive genetic testing: From basic research to clinical practice. *Science* 1997;278:602–605.
11. Task Force on Genetic Testing. Promoting safe and effective genetic testing in the United States: Final report. Baltimore: Johns Hopkins University Press [In press]. Available at: http://www.nhgri.nih.gov/ELSI/TFGT_final/.
12. Harper PS. What do we mean by genetic testing? *J Med Genet* 1997;34:749–752.
13. Holtzman NA, Shapiro D. The new genetics: Genetic testing and public health. *BMJ* 1998;316:852–856.
14. Khoury MJ, Newill CA, Chase GA. Epidemiologic evaluation of screening for risk factors: Application to genetic screening. *Am J Public Health* 1985;75:1204–1208.
15. Edmonts LD, Layde PM, James LM, Flynt JW Jr, Erickson JD, Oakley GP Jr. Congenital malformation surveillance: Two American systems. *Int J Epidemiol* 1981;10:247–252.
16. International Clearinghouse for Birth Defects Monitoring Systems. Congenital malformations worldwide. Amsterdam: Elsevier Science Publishers BV, 1991.
17. Horm JW, Asire AJ, Young JLJ. Surveillance, Epidemiology, and End Results Program: Cancer incidence and mortality in the United States, 1973–91. Bethesda, Md: National Cancer Institute, 1994.
18. Storm HH, Michelsen EV, Clemmensen IH, Pihl J. The Danish Cancer Registry: History, content, quality and use. *Dan Med Bull* 1997;44:535–539.
19. Buehler JW. Surveillance. In: Rothman KJ, Greenland S, editors. Modern epidemiology. Philadelphia: Lippincott-Raven Publishers, 1998;435–457.
20. Rothman KL, Greenland S. Modern epidemiology. Philadelphia: Lippincott-Raven Publishers, 1998.
21. Khoury MJ, Yang QH. The future of genetic studies of complex human diseases: An epidemiologic perspective. *Epidemiology* 1998;9:350–354.
22. Neutra RR, Drolette ME. Estimating exposure-specific disease rates from case-control studies using Bayes theorem. *Am J Epidemiol* 1978;108:214–222.
23. Greenland S. Multivariate estimation of exposure-specific incidence from case-control studies. *J Chronic Dis* 1981;34:445–453.
24. Levin ML. The occurrence of lung cancer in man. *Acta Un Intern Cancer* 1953;9:531–541.
25. Khoury MJ, Beaty TH, Cohen BH. Fundamentals of genetic epidemiology. Oxford: Oxford University Press, 1993.
26. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fraction. *Am J Public Health* 1998;88:15–19.
27. van der Put NM, Steegers-Theunissen RP, Frosst P, Trijbels FJ, Eskes TK, van den Heuvel LP, Mariman EC, den Heyer M, Rozen R, Blom HJ. Mutated methylenetetrahydrofolate reductase as a risk factor for spina bifida. *Lancet* 1995;346:1070–1071.
28. Whitehead AS, Gallagher P, Mills JL, Kirke PN, Burke H, Molloy AM, Weir DG, Shields DC, Scott JM. A genetic defect in 5,10 methylenetetrahydrofolate reductase in neural tube defects. *QJM* 1995;88:763–766.
29. Ou CY, Stevenson RE, Brown VK, Schwartz CE, Allen WP, Khoury MJ, Rozen R, Oakley GP Jr, Adams MJ Jr. 5,10 methylenetetrahydrofolate reductase genetic polymorphism as a risk factor for neural tube defects. *Am J Med Genet* 1996;63:610–614.
30. Shaw GM, Rozen R, Finnell RH, Wasserman CR, Lammer EJ. Maternal vitamin use, genetic variation of infant methylenetetrahydrofolate reductase, and risk for spina bifida. *Am J Epidemiol* 1998;148:30–37.
31. Botto LD, Yang QH. MTHFR and birth defects. *Am J Epidemiol* 2000;151:862–877.
32. Whittemore AS, Gong G, Itnyre J. Prevalence and contribution of BRCA1/BRCA2 mutation in breast cancer and ovarian cancer: Results from three U.S. population-based case-control studies of ovarian cancer. *Am J Hum Genet* 1997;60:496–504.
33. Breast Cancer Linkage Consortium. Pathology of familial breast cancer: Differences

between breast cancer in carriers of BRCA1 or BRCA2 mutations and sporadic cases. *Lancet* 1997;349:1505–1510.
34. Kosary CL, Ries LAG, Miller BA, Hankey BF, Harras A, Edwards BK, editors. SEER cancer statistics review, 1973–1992: Tables and graphs. Bethesda, Md: National Cancer Institute, 1995. NIH Publication 96–2789.
35. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. *Am J Hum Genet* 1991;48:232–242.
36. Ford D, Easton DF, Peto J. Estimates of the gene frequency of BRCA1/BRCA2 and its contribution to breast and ovarian cancer incidence. *Am J Hum Genet* 1995;57: 1457–1462.
37. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516–1517.
38. Diabetes Epidemiology Research International Group. Geographic patterns of childhood insulin-dependent diabetes mellitus. *Diabetes* 1988;37:1113–1119.
39. LaPorte RE, Tajima N, Akerblom HK, Berlin N, Brosseau J, Christy M, Drash AL, Fishbein H, Green A, Hamman R, Harris M, King H, Laron Z, Neil A. Geographic differences in the risk of insulin-dependent diabetes mellitus: The importance of registries. *Diabetes Care* 1985;8(Suppl 1):101–107.
40. Mayo NE, Chockalingam A, Reeder BA, Phillips S. Surveillance for stroke in Canada. *Health Reports* 1994;6:62–72.
41. Declich S, Carter AO. Public health surveillance: Historical origins, methods and evaluation. *Bull World Health Organ* 1994;72:285–304.
42. Halperin W, Baker EL, editors. Public health surveillance. New York: Van Nostrand Reinhold, 1992.
43. Feuer EJ, Wun LM, Boring CC, Flanders WD, Timmel MJ, Tong T. The lifetime risk of developing breast cancer. *J Natl Can Inst* 1993;85:892–897.
44. Greenland S. Multivariate estimation of exposure-specific incidence from case-control studies. *J Chron Dis* 1981;34:445–453.
45. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–1886.
46. Benichou J, Wacholder S. A comparison of three approaches to estimate exposure-specific incidence rates from population-based case-control data. *Stat Med* 1994;13: 651–661.
47. McTiernan A, Gilligan MA, Redmond C. Assessing individual risk for breast cancer: Risky business. *J Clin Epidemiol* 1997;50:547–556.
48. Kinmonth AL, Reinhard J, Bobrow M, Pauker S. The new genetics: Implications for clinical services in Britain and the United States. *BMJ* 1998;316:767–770.
49. Marteau TM, Croyle RT. The new genetics: Psychological responses to genetic testing. *BMJ* 1998;316:693–696.
50. Harper PS. Genetic testing, common diseases, and health service provision. *Lancet* 1995;346:1645–1646.
51. Bell J. The new genetics: The new genetics in clinical practice. *BMJ* 1998;316:618–620.
52. Ponder B. Genetic testing for cancer risk. *Science* 1997;278:1050–1054.
53. Grady C. Ethics and genetic testing. *Adv Intern Med* 1999;44:389–411.
54. Barber JCK. Code of practice and guidance on human genetic testing services supplied direct to the public: Advisory Committee on Genetic Testing. *J Med Genet* 1998;35:443–445.
55. King PA. The past as prologue: Race, class, and gene discrimination. In: Annas GJ, Sherman E, editors. Gene mapping: Using law and ethics as guides. New York: Oxford University Press, 1992:94–111.