

Population Stratification in Epidemiologic Studies of Common Genetic Variants and Cancer: Quantification of Bias

Sholom Wacholder, Nathaniel Rothman, Neil Caporaso

Background: Some critics argue that bias from population stratification (the mixture of individuals from heterogeneous genetic backgrounds) undermines the credibility of epidemiologic studies designed to estimate the association between a genotype and the risk of disease. We investigated the degree of bias likely from population stratification in U.S. studies of cancer among non-Hispanic Caucasians of European origin. **Methods:** An expression of the confounding risk ratio—the ratio of the effect of the genetic factor on risk of disease with and without adjustment for ethnicity—is used to measure the potential relative bias from population stratification. We first use empirical data on the frequency of the *N*-acetyltransferase (NAT2) slow acetylation genotype and incidence rates of male bladder cancer and female breast cancer in non-Hispanic U.S. Caucasians with ancestries from eight European countries to assess the bias in a hypothetical population-based U.S. study that does not take ethnicity into consideration. Then, we provide theoretical calculations of the bias over a large range of allele frequencies and disease rates. **Results:** Ignoring ethnicity leads to a bias of 1% or less in our empirical studies of NAT2. Furthermore, evaluation of a wide range of allele frequencies and representative ranges of cancer rates that exist across European populations shows that the risk ratio is biased by less than 10% in U.S. studies except under extreme conditions. We note that the bias decreases as the number of ethnic strata increases. **Conclusions:** There will be only a small bias from population stratification in a well-designed case-control study of genetic factors that ignores ethnicity among non-Hispanic U.S. Caucasians of European origin. Further work is needed to estimate the effect of population stratification within other populations. [J Natl Cancer Inst 2000;92:1151–8]

When the risk of disease varies between two ethnic groups, then any factor—whether genetic or environmental—that also varies between the groups will appear to be related to disease. A classic thought experiment is given by Lander and Schork (1): Any genotype that is more common in Americans of Chinese origin than in those of European origin will be positively associated with the “use of chopsticks” phenotype in studies that ignore ethnicity carried out in San Francisco (CA).

This phenomenon is called population stratification in epidemiologic studies investigating the effect of a genetic factor on disease. Concerns about population stratification have raised doubts about the credibility of reported findings (2) and have led to calls for routine use of related controls in case-control studies of genetic factors (3) to eliminate the possibility of population stratification. The editors of *Nature Genetics* stated recently, “Population stratification (or admixture) errors are clearly a problem for case-control studies” (4). The textbook by Strachan and Read (5) includes the statement that association studies “be-

came discredited because of statistical problems,” particularly population stratification.

In fact, population stratification is a manifestation of confounding (Fig. 1). Classically, confounding is the distortion of the relationship between the exposure of interest and disease due to the effect of a true risk factor that is related to the exposure. Similarly, population stratification is the distortion of the relationship between a genotype of interest and disease due to the effect of a true risk factor that is related to the genotype. In population stratification, ethnicity acts as a surrogate for the true risk factor, which may be environmental or genetic; as such, controlling for ethnicity can reduce the confounding bias. For example, the misleading association in the chopsticks example presented above, where both the genotype and the phenotype are distributed differently in Americans of Chinese origin than in those of European origin, would be greatly reduced by controlling for ethnicity.

The classic example of population stratification often cited in the literature is the relationship between the presence of the Gm^{3;5,13,14} haplotype and type 2 diabetes mellitus among residents of a Pima-Papago Indian reservation with varying degrees of European ancestry (6). With restriction to those who are 35 years of age or older and with zero or eight Indian great-grandparents [Table 1, based on Tables 3 and 4 of Knowler et al. (6)], the estimate of the odds ratio between haplotype status and diabetes is 0.35 without accounting for Indian heritage; on adjustment for Indian heritage, the estimate of the odds ratio rises to 0.73, substantially closer to 1 and providing much weaker evidence for an effect on risk of diabetes.

Our purpose is to assess whether population stratification should be an important consideration in the study design and in the interpretation of cohort and case-control studies with unrelated controls planned to evaluate associations between common genetic polymorphisms and cancer (7). We, therefore, use empirical data on factors that affect the bias and carry out theoretical calculations to evaluate the extent of bias from population stratification in estimates of the effect of a genotype on risk of cancer.

METHODS

A general measure of bias from confounding is the confounding risk ratio (CRR) (8), which is simply the ratio of the crude and adjusted relative risk parameters. Cornfield et al. (9) first used the idea of relative bias to refute the “constitutional hypothesis” that a genetic factor confounded and thus explained away the smoking-lung cancer association; later, Miettinen (8) showed how to quantify the bias formally in a general setting. By use of the CRR, we assess the

Affiliation of authors: Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

Correspondence to: Sholom Wacholder, Ph.D., National Institutes of Health, EPS 8046, 6120 Executive Blvd., Bethesda, MD 20892-7244 (e-mail: Wacholder@nih.gov).

See “Notes” following “References.”

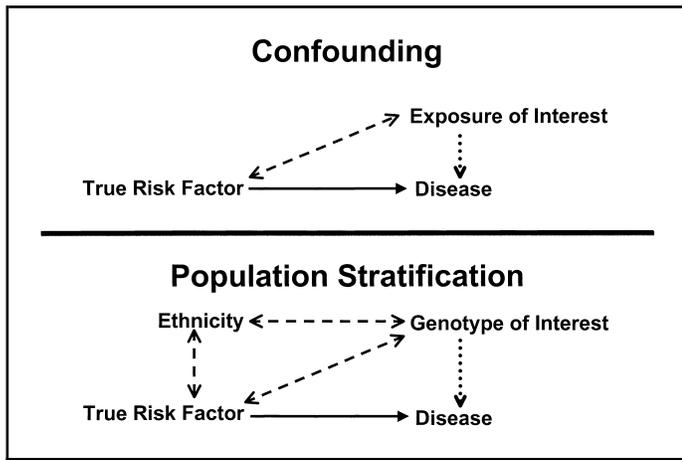


Fig. 1. Schematic drawing of classical confounding and population stratification. **Dotted unidirectional arrows** indicate confounded association. **Solid unidirectional arrows** indicate direction of causal relationship. **Broken line with bidirectional arrows** indicate correlation that is not causal. In classical confounding, an exposure of interest correlated with the true risk factor but not truly causal of disease can be misleadingly seen to be associated with risk of disease. Statistical adjustment for the true risk factor, if known and accurately measured, can eliminate the bias. In population stratification, the genotype of interest, correlated with the true risk factor because both are correlated with ethnicity, can be incorrectly seen to be associated with risk of disease. Again, statistical adjustment for the true risk factor that is known and accurately measured will eliminate bias (as noted in the “Discussion” section); if the true risk factor is unknown, adjustment for ethnicity can reduce bias from confounding to the extent that ethnicity correlates with the genotype of interest and the true risk factor.

Table 1. Indian heritage, Gm^{3,5,13,14} haplotype, and risk of diabetes in Pima-Papago Indians over the age of 35 years [adapted from Knowler et al. (6)]*

Indian heritage	Gm ^{3,5,13,14} haplotype	No. without diabetes	No. with diabetes
No Indian great-grandparents	Present	14	1
	Absent	3	2
Full-heritage Pima-Papago	Present	7	10
	Absent	706	1058

*Among those 35 years or older, the estimate of relative risk is 0.35 (95% confidence interval [CI] = 0.17–0.73) without adjusting, but changes to 0.73 (95% CI = 0.30–1.8) on adjustment for Indian heritage. The abstract of the study by Knowler et al. (6) presents a prevalence ratio of 0.27 (95% CI = 0.18–0.40) among 37 cases in 4920 individuals aged 5 years or above, without adjusting for age or ethnicity; the prevalence ratio adjusted for age is 0.53 (95% CI = 0.39–0.72) (6). On restricting to those with none, four, or eight Indian great-grandparents, the prevalence ratio adjusted for Indian heritage and age is 0.83 (6). Thus, the best estimate of bias from ignoring ethnicity, after controlling for age, is 0.53/0.83 = 0.64.

bias due to population stratification when studying the effect of a genotype without accounting for ethnicity.

Properties of the CRR

The CRR can be used to determine whether *E*, the genotype of interest, is related to risk of disease without bias from a confounder *C*. CRR is defined as the ratio of the risk ratio RR_p obtained from pooling over (i.e., by ignoring) the confounder *C* to the risk ratio RR obtained after adjusting for the confounder *C*. Thus, the CRR is equal to 1 when there is no bias, including specifically the situations where the genotype is unrelated to the true risk factor. A CRR of 0.9 or 1.1 means that the pooled estimate of the risk ratio from the genotype is 10% lower or 10% higher than the adjusted risk ratio, respectively. In general, the CRR can be expressed [p. 96 in (10)], as in equation 1:

$$CRR = \frac{RR_p}{RR} = \frac{\sum_k \Pr(C = k|E = 1)RR_k}{\sum_k \Pr(C = k|E = 0)RR_k}, \quad [1]$$

where *C* has *K* levels *k* = 1 through *K*, *E* of 1 or 0 indicates the presence or absence of the genotype, respectively, and RR_k is the ratio of R_k , the risk of disease in ethnic group *k*, relative to R_1 , the risk in ethnic group 1 when *E* = 0. Equation 1 defines the expression of CRR as the ratio of the crude and adjusted estimates; equation 2, below, shows that the CRR is a ratio of weighted averages of the rate ratios specific to each level of *C* (8).

Equation 2 applies generally to the bias from confounding by any factor. To estimate the confounding bias from population stratification due to ethnicity in an analytic study of a genotype and cancer, we re-express CRR in terms of data parameters that we are able to obtain, particularly, $\Pr(EIC = k)$, the proportions of population with an allele of interest, and R_k , the disease incidence rates in level *k*. In particular, we obtain equation 2 below from equation 1 by replacing RR_k with R_k/R_1 (justified by the definition of rate ratio) and $\Pr(C = k|E = i)$ (*i* can be equal to 0 or 1) with $\Pr(E = i|C = k)\Pr(C = k)/\Pr(E = i)$ (justified by Bayes theorem):

$$CRR = \frac{\sum_k \Pr(E = 1|C = k) \Pr(C = k) RR_k / \Pr(E = 1)}{\sum_k \Pr(E = 0|C = k) \Pr(C = k) RR_k / \Pr(E = 0)} \\ = \frac{\sum_k \Pr(E = 1|C = k) \Pr(C = k) R_k / \sum_k \Pr(E = 1|C = k) \Pr(C = k)}{\sum_k \Pr(E = 0|C = k) \Pr(C = k) R_k / \sum_k \Pr(E = 0|C = k) \Pr(C = k)}. \quad [2]$$

That is, the CRR is again expressed as a ratio of two weighted averages of level-specific rates. The weights in the numerator $\Pr(E = 1)$ and denominator $\Pr(E = 0)$ of CRR are complements of one another, i.e., the proportions of the population with and without the genotype of interest and in level *k* of *C*, respectively.

Equation 2 shows clearly that the extent of the bias measured by CRR depends on the variability of both the disease rates R_k and the genotype frequencies across the *K* levels of the confounder *C*. Remarkably, however, no measure of the strength of the genotype–disease relationship appears in equation 2; indeed, the CRR is not affected by whether or not the genotype itself is related to disease. Although incidence rates of a disease with a genotype that is a major determinant of risk might track with the prevalence of the genotype across levels of the confounder, the amount of confounding (and the CRR) is based on the cancer rates of those without the genotype of interest and is, therefore, unaffected by the risk associated with the genotype. For simplicity, our empirical investigation uses the readily available national cancer-incidence rates, which are weighted averages of the cancer rates in individuals with and without the genotype (11).

Empirical Study of *N*-Acetyltransferase 2 Gene Slow Acetylators and Male Bladder Cancer and Female Breast Cancer

We use the CRR to estimate the bias from population stratification in a study of a genotype and cancer in the United States that did not account for ethnicity at all. With the use of equation 2, we estimate the CRR, the ratio of the rate ratios obtained by dividing the rate ratio for cancer obtained from a hypothetical analytic study that pooled immigrants from different countries and their descendants by the rate ratio adjusted for the country of ancestry. We make two simplifying and conservative assumptions to quantify the bias from population stratification in studies in the United States: 1) that there is no intermarriage across ethnic groups among Americans of European descent and 2) that the genotype frequencies and the cancer rates among these groups are the same as in their country of ancestry.

For our empirical study, we focus on the effect of being a slow acetylator due to mutations in the NAT2 gene on cancer risk. The NAT2 genotype is characterized phenotypically (by measuring the extent of acetylation of a drug by this enzyme) or by direct analysis of DNA. Below, we refer to “NAT2 slow acetylators,” regardless of how the determination was made. We chose to study NAT2 acetylation because there is some evidence that slow acetylators may be at elevated risk of developing cancer (12) and because of the availability in the literature of extensive international data on the frequency of this genotype. We chose to study bladder cancer incidence in men and breast cancer incidence in women; NAT2 slow acetylators are believed to be at increased risk of bladder cancer (12) but not of breast cancer (13). For determining the proportion of the population with confounder level *k*, i.e., $\Pr(C = k)$, we use information on “first ancestry” based on the first ethnicity reported to the U.S. census (14) (also available from URL <http://venus.census.gov/cdrom/lookup/922371159>) to determine the numbers of American descendants of immigrants from the eight Eu-

ropean countries. These ancestries were chosen because they represent the top eight fractions of non-Hispanics of European origin in the United States. We obtained cancer incidence rates, R_k , from a registry for each country included in *Cancer in Five Continents (11)* and frequency of NAT2 slow acetylators from a published study from each of these countries: Germany (15), Ireland (16), England (17), Italy (18), France (19), Scotland (20), Poland (21), and Sweden (22). We used published data on frequency of slow acetylators in the eight countries to determine $\Pr(E = 1|C = k)$ for our calculations (Table 2), with C being the confounder with eight levels corresponding to $K = 8$ countries and $E = 1$ indicating an NAT2 slow acetylator.

Recognizing that these results are limited to a study of a single genotype, we perform sensitivity analyses to assess whether results based on actual slow acetylation frequencies and numbers of descendants from the different countries as shown in Table 2 are applicable more generally. In addition to evaluating CRR in a population where the number of subjects from each ancestry are proportional to those reported in the census data (Table 2), we also consider a population in which the numbers of descendants from each of the K countries were equal to one another; i.e., $\Pr(C = k) = 1/K$ for all K levels. Table 3 presents the results obtained from these analyses (no census weights applied). To mimic a genotype with a much broader range, we calculate the CRR based on modified frequencies labeled as “stretched frequencies” of NAT2 slow acetylators. In generating these stretched frequencies, we preserved the ordering of the original frequencies but stretched the spacing between them: We add a factor of displacement, i.e., a product of the logarithm of 2 and $-3.5, -2.5, -1.5, -0.5, 0.5, 1.5, 2.5,$ and $3.5,$ respectively, to the logit of the genotype frequencies ordered from the lowest to the highest so that the stretched frequencies ranged from 8% to 95% (see Table 2). We also subtract a constant number, 45%, from the actual frequencies, giving a range of 6%–20%, to see the impact of bias when the genotype frequencies were lower. We also calculate the CRR when the relationship between slow acetylation frequency and the cancer rate is monotonic (i.e., for the sake of analysis, we hypothetically paired the lowest genotype frequency with the lowest cancer rate, etc., see in Fig. 3) and compared it with the CRRs over all $8! = 40\,320$ possible pairings of frequency of slow acetylation and cancer rate.

Finally, we assess the effect of the number of different ethnicities in the population. We report the maximum and minimum CRR from every one of the $8!/k!(8-k)!$ possible subsets of size $k = 2, \dots, 8$ ethnic groups (28, 56, 70, 56, 28, 8, and 1 possible comparisons for a given subset size, respectively).

Theoretical Study of Bias From Population Stratification

We also performed some theoretic calculations using equation 2 to estimate the bias from population stratification when a genotype is unrelated to cancer risk. We investigate the CRR over a wide variety of ranges of genotype frequencies and of rates of disease. We first focus on ranges that are representative of those found in European populations from which most non-Hispanic Americans descend and then expand to values that represent populations that are even more diverse. For this analysis, we assume that there are eight strata of equal size in the population, with a specified range of allele frequencies and relative rates

Table 3. Estimated bias from pooling eight ancestry groups in the study of *N*-acetyltransferase (NAT2) slow acetylators and male bladder cancer and female breast cancer

	Confounding risk ratio (bias) in estimating effect of NAT2 slow acetylation*			
	Actual frequencies†		Stretched frequencies‡	
	Male bladder cancer	Female breast cancer	Male bladder cancer	Female breast cancer
Weights				
Census‡	0.992	1.010	0.946	1.058
None	0.985	1.009	0.915	1.017

*Bias factors for each combination of NAT2 slow acetylation frequency (actual or stretched), cancer site (male bladder or female breast), and population (mixture of ethnicities in the same proportion as in the census data (Table 2) or assuming mixture of ethnicities in equal proportions).

†Actual and stretched frequencies of NAT2 slow acetylation phenotype/genotype as shown in Table 2.

‡Census weight is proportional to the numbers of descendants in the United States (Table 2).

of disease, spaced to be equidistant on a logistic and logarithmic scale, respectively.

Of the five genotype frequency ranges (GFRs) we studied, four (0.001–0.10, 0.01–0.20, 0.10–0.40, and 0.30–0.60) are broadly representative of the range of frequencies of different alleles for a large number of genes reported by Cavalli-Sforza et al. (23) in Europe (excluding ethnicities not found in the United States in substantial numbers, such as Lapps and Basques). The other GFR (0.01–0.99) is more extreme. Of the range of six rate ratio ranges (RRRs) that we studied, RRRs of 1–1.5, 1–2, and 1–3 represent realistic values for the range of cancer rates among European countries (11); a relative rate of 3 is probably a very conservative upper bound in any two European-American ethnic groups present in the United States for several generations. To demonstrate what can happen in an extreme case, perhaps for an outcome other than cancer, we also considered RRRs of 1 : 10 and 1 : 100.

For each hypothesized pair of values of GFR and RRR, we calculate the lowest and the highest possible CRR using any two genotype frequencies and any two rate ratios from the set of eight. These represent the worst-case scenarios for a study of two ethnic groups, with the most extreme differences in genotype frequencies and disease rates from each set of GFRs and RRRs. We also consider the $8! = 40\,320$ CRRs from each possible permutation of the genotype frequencies in the eight groups. We calculate the geometric mean and the largest and smallest of the 40 320 CRRs as well as the 10th, 50th, and 90th percentiles. (See Table 4 for results.)

Table 2. Ancestry of non-Hispanic European-Americans, incidence rates for male bladder cancer and female breast cancer, and *N*-acetyltransferase 2 (NAT2)—slow acetylation genotype—frequencies

Country (reference No.)	No. of descendants in America, millions	NAT2 slow acetylators, frequency, %		Male bladder cancer rate, per 100 000/y‡	Female breast cancer rate, per 100 000/y‡
		Actual*	Stretched†		
England (17)	22	60	68	20	69
France (19)	9	64	91	16	86
Germany (15)	46	51	8	23	62
Ireland (16)	23	52	16	12	64
Italy (18)	11	53	29	35	74
Poland (21)	7	60	81	16	33
Scotland (20)	8	65	95	23	73
Sweden (22)	3	60	51	17	73

*The actual frequencies are based on data from the country in the cited reference.

†The stretched frequencies are generated by mathematical manipulation (see text) and maintain the order of the actual frequencies but provide an artificially increased range for empirical analysis as described in the text.

‡Cancer incidence rates were obtained from (11).

RESULTS

The numbers of American descendants, proportions of slow acetylators, and male bladder and female breast cancer rates are shown in Table 2 and displayed in Fig. 2, A. Note that there is little evidence of a linear relationship for either cancer site. The bias factors presented below changed only slightly when we evaluated the CRR based on the rates in the fast acetylators (those with at least one functional NAT2 allele) calculated by the method of Axelson (24) from national rates, assuming that the bladder cancer rate in slow acetylators was 1.4 times greater than that in fast acetylators (12).

CRRs From Empirical Study

We present the CRRs calculated for a study of NAT2 slow acetylation phenotype/genotype and male bladder cancer or female breast cancer in Table 3. The qualitative impressions of bias are almost the same for male bladder cancer or female breast cancer throughout. When the actual acetylation frequencies are used, the CRRs (ratios of the rate ratios from a study that ignored ethnicity to one that adjusted for ethnicity under our assumptions) are very close to unity (Table 3). When the range of the acetylation frequency is stretched or expanded artificially (Fig. 2, B), the CRRs are slightly higher but still too small to affect the qualitative conclusions of a study in a major way. The CRR was also close to 1 (0.978) when allele frequencies calculated

by subtracting 45% from the actual ones were used (resulting in a range of 6%–20%; see Table 2). None of the 40 320 possible permutations, or pairings of acetylation frequency and cancer rates, lead to appreciable bias; with census weights, the range of the CRR is 0.95–1.06 for male bladder cancer and 0.98–1.02 for female breast cancer (data not shown). When no weighting is used, the bias is most extreme (data not shown) for the permutations where the relationship between acetylation frequency and cancer rate has the most extreme rank associations (Fig. 3).

Effect of Number of Ethnicities

Fig. 4 shows the maximum and minimum bias, as measured by the CRR, over all subsets of sizes $k = 1-8$ among the $K = 8$ ethnicities by use of census weights. Note that the most extreme bias is for $k = 2$ and $k = 3$ and that the bias becomes closer to 1 as the number of ethnic groups in the populations increases.

Theoretical Study of Bias From Population Stratification

Table 4 shows the effect of the variability of genotype frequencies and cancer rates over a wider range of scenarios than the empirical example. As before, we consider the range of possible CRRs when a population is a mixture of eight distinct populations; we also assume that the cancer rates and genotype

Fig. 2. Frequencies of *N*-acetyltransferase 2 (NAT2) slow acetylators and female breast cancer (circles) and male bladder cancer (triangles) incidences in eight European countries used in the examples. The actual NAT2 slow acetylator frequencies are graphed in panel A; the stretched frequencies are graphed in panel B. See the "Methods" section for details about generating "stretched frequencies." The numbers next to the data points on the graph represent the country of ancestry based on the orderly arrangement of the frequency of NAT2 slow acetylators. 1 = Germany; 2 = Ireland; 3 = Italy; 4 = Sweden; 5 = England; 6 = Poland; 7 = France; and 8 = Scotland.

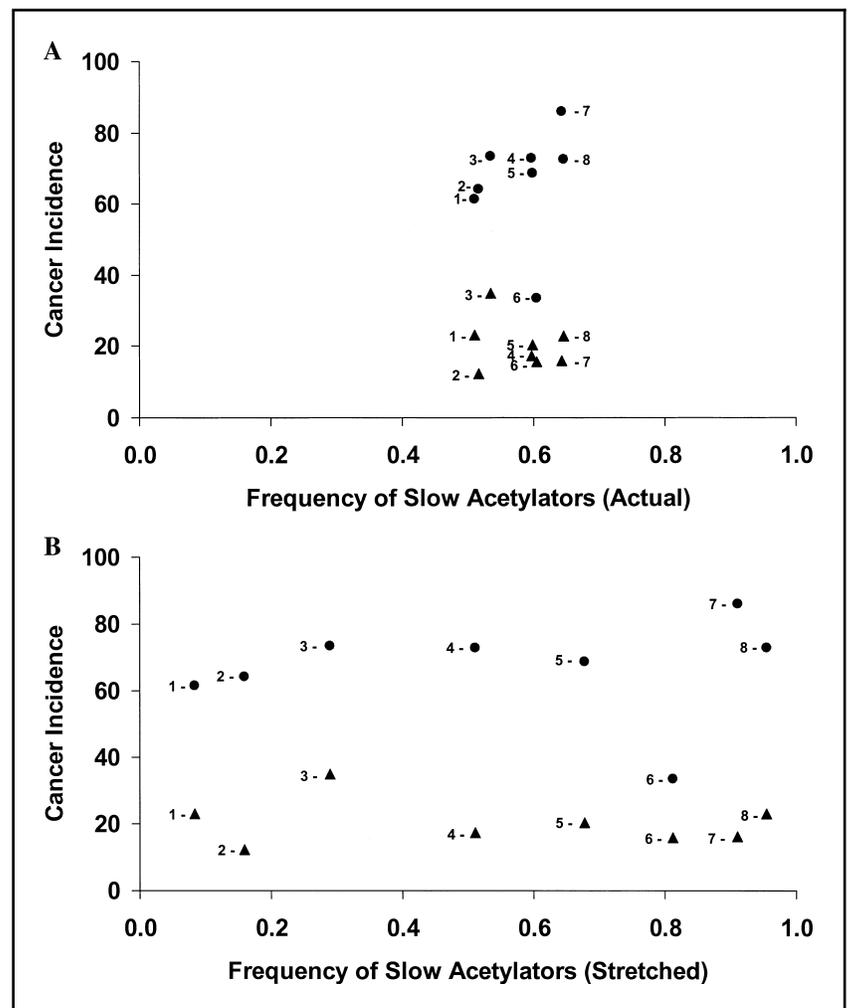


Fig. 3. Hypothetic monotonic pairing of frequencies of *N*-acetyltransferase 2 (NAT2) slow acetylator frequencies and female breast cancer (**circles**) and male bladder cancer (**triangles**) incidences in eight European countries used in the examples. This hypothetic scenario leads to the most extreme bias. The numeric values of genotype frequencies are paired in an orderly fashion; i.e., the smallest allele frequency with the smallest cancer rate and the largest genotype frequency with the largest cancer rate, etc. The numbers of genotype frequencies and cancer rates are taken from the numbers given in Table 2, but the pairs are hypothetically constructed.

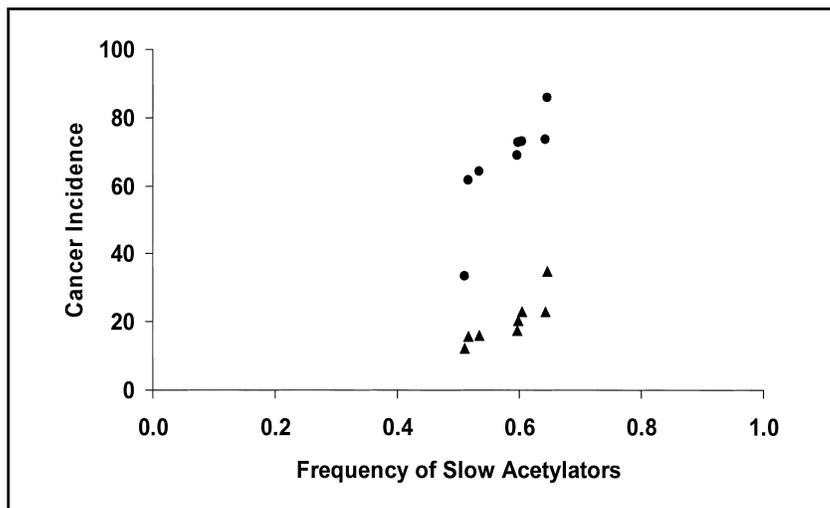
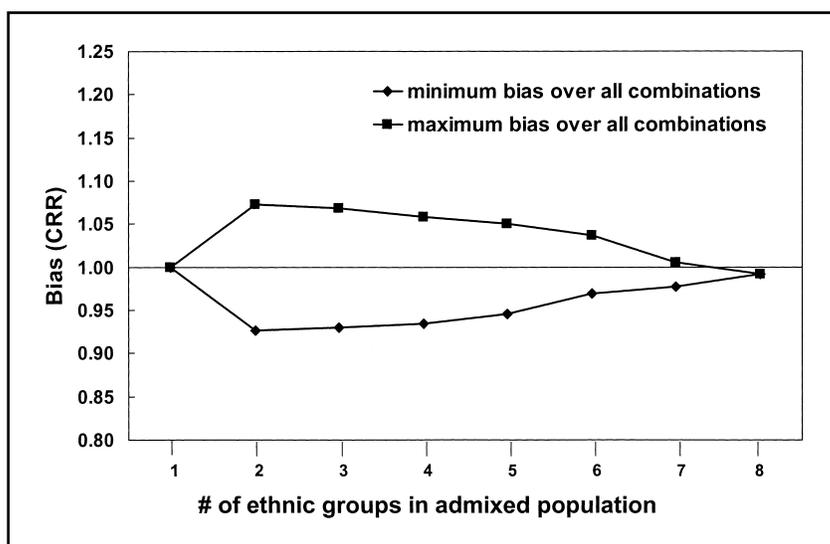


Fig. 4. Minimum and maximum confounding risk ratios over all combinations of ethnic groups for male bladder cancer in subsets of size two through eight of the eight ethnic groups. Patterns for female breast cancer are not shown. CRR = confounding risk ratio.



frequencies are spaced equidistantly within a specified range on the logarithmic and logistic scales. For example, in Table 4, the CRR ranges between 0.89 and 1.12, where the ratio of the highest to the lowest rate of disease is 2 and the genotype frequencies range from 0.1 to 0.4. Table 4 demonstrates that, when cancer rates are the same across ethnic groups (i.e., $RRR = 1.0-1.0$), the CRR is 1.0 and there is no bias, regardless of the difference in genotype frequencies. As observed previously, the worst bias for two-ethnic group comparisons is always greater than when eight ethnic groups are compared. The CRRs for the GFRs and RRRs that might exist in a population of non-Hispanic American Caucasians descended from different European countries is almost always usually under 10%. The most extreme CRRs that are realistic for European-Americans are only 0.78 and 1.22 (see Table 4), corresponding to RRR of 1.00–3.00 and GFR of 0.001–0.10. Furthermore, these extremes always occur when there was maximum rank correlation (monotonicity) between genotype frequency and the *RR*. Even this is tempered by the observation that the CRR is 1.00 when the genotype frequencies and relative rates are independent and that the median and geometric mean of the CRR over all possible orderings are always 1.00. While the extreme values of CRR are somewhat larger for GFRs and RRRs that are unrealistic for Americans of European

origin, the geometric means and medians remain at 1.00. Furthermore, the vast majority of the simulations show CRRs to be much closer to 1.00 than the extreme values. Changing the spacing of genotype frequencies to a linear scale does not alter the impression of Table 4 (data not shown).

DISCUSSION

This analysis shows that very little bias would ensue from a population-based study of common polymorphisms and cancer that ignored ethnicity in a study base of multiethnic non-Hispanic European-Americans similar to current residents of the United States. Indeed, bias is remarkably small even when the range of genotype frequencies is extremely large or when there is a strong association between genotype frequency and cancer rate.

The results presented here indicate that the following set of conditions must be met to have important bias from population stratification when there is truly no effect of the allele. 1) The frequency of the genotype of interest varies substantially by ethnicity. 2) The cancer rate varies substantially by ethnicity. 3) The cancer rates and genotype frequencies vary together; this can only happen in a study of multiple ethnic groups when the

Table 4. Confounding rate ratios (CRRs) calculated from different hypothetical cancer incidence rate ratios and allele frequencies*

Rate ratio, range†	Genotype frequency, range‡	CRRs among all possible subsets of size 2 from the eight groups		CRRs among all 40 320 orderings of eight groups				
		Most extreme negative association, minimum	Most extreme positive association, maximum	Most extreme negative association, minimum§	10th percentile	No association, 50th percentile (median and geometric mean)	90th percentile	Most extreme positive association, maximum¶
1.00–1.00	0.0010–0.1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00–1.00	0.0100–0.2000	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00–1.00	0.1000–0.4000	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00–1.00	0.3000–0.6000	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00–1.00	0.0100–0.9900	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00–1.50	0.0010–0.1000	0.80	1.21	0.91	0.96	1.00	1.04	1.09
1.00–1.50	0.0100–0.2000	0.80	1.21	0.91	0.96	1.00	1.04	1.09
1.00–1.50	0.1000–0.4000	0.85	1.17	0.93	0.97	1.00	1.03	1.07
1.00–1.50	0.3000–0.6000	0.89	1.13	0.95	0.97	1.00	1.03	1.05
1.00–1.50	0.0100–0.9900	0.67	1.49	0.85	0.92	1.00	1.09	1.18
1.00–2.00	0.0010–0.1000	0.66	1.35	0.85	0.93	1.00	1.07	1.15
1.00–2.00	0.0100–0.2000	0.67	1.35	0.86	0.93	1.00	1.07	1.15
1.00–2.00	0.1000–0.4000	0.75	1.29	0.89	0.94	1.00	1.06	1.12
1.00–2.00	0.3000–0.6000	0.81	1.22	0.92	0.96	1.00	1.04	1.09
1.00–2.00	0.0100–0.9900	0.51	1.97	0.75	0.87	1.00	1.15	1.33
1.00–3.00	0.0010–0.1000	0.50	1.53	0.78	0.89	1.00	1.11	1.22
1.00–3.00	0.0100–0.2000	0.52	1.53	0.79	0.89	1.00	1.11	1.22
1.00–3.00	0.1000–0.4000	0.64	1.44	0.84	0.92	1.00	1.09	1.18
1.00–3.00	0.3000–0.6000	0.73	1.35	0.88	0.94	1.00	1.07	1.14
1.00–3.00	0.0100–0.9900	0.34	2.92	0.65	0.81	1.00	1.23	1.53
1.00–10.00	0.0010–0.1000	0.19	1.88	0.64	0.82	1.00	1.18	1.37
1.00–10.00	0.0100–0.2000	0.24	1.91	0.66	0.83	1.00	1.18	1.37
1.00–10.00	0.1000–0.4000	0.44	1.78	0.74	0.86	1.00	1.15	1.30
1.00–10.00	0.3000–0.6000	0.59	1.64	0.81	0.90	1.00	1.11	1.23
1.00–10.00	0.0100–0.9900	0.11	9.09	0.49	0.71	1.00	1.41	2.05
1.00–100.00	0.0010–0.1000	0.04	2.07	0.58	0.79	1.00	1.22	1.44
1.00–100.00	0.0100–0.2000	0.10	2.11	0.59	0.79	1.00	1.22	1.44
1.00–100.00	0.1000–0.4000	0.34	1.98	0.69	0.84	1.00	1.18	1.37
1.00–100.00	0.3000–0.6000	0.53	1.81	0.77	0.88	1.00	1.14	1.29
1.00–100.00	0.0100–0.9900	0.02	49.75	0.42	0.66	1.00	1.52	2.40

*Figure in **bold** are scenarios we consider to be realistic for Americans of European ancestry.

†The individual risk ratio ranges for the eight groups are spaced equidistantly on a logistic scale.

‡The genotype frequencies in the eight groups are spaced equidistantly on a logistic scale.

§The strongest negative association between the genotype frequency and cancer rates due to noncausal association between the genotype frequency and the unmeasured determinant of cancer.

||No rank association between the genotype frequency and cancer rates.

¶The strongest positive association between the genotype frequency and cancer rates due to noncausal association between the genotype frequency and the unmeasured determinant of cancer.

genotype is related to a true risk factor with attributable risk so high as to explain much of the international variability. When there is little or no association, the bias is clearly small on average (Table 4). 4) The risk factors truly responsible for the variation in cancer rates across ethnicity are unknown or information about it is unavailable; if data on these risk factors were available, they could be used in the analysis, thereby reducing or eliminating the bias. 5) The investigators are unable to distinguish among ethnic groups with different genotype frequencies and cancer rates, even crudely. Even if all of these conditions are met in a single study and, in fact, produce substantial bias, the results will probably not be replicated in a distinctly different study population because the same conditions are unlikely to exist.

A genotype may not cause cancer, yet it may still be strongly related to cancer rate across ethnicities (point 3 above) if the genotype frequency varies with a true risk factor across populations. For example, an allele with a strong North–South frequency gradient in Europe would be associated with drinking

beer instead of wine or with the amount of olive oil consumed; if one of these is an important risk factor for the cancer under study, there could be some bias, even if the allele is not causally related to beer or olive oil consumption. This is the scenario, as shown in Fig. 1, of a risk factor for disease noncausally related to the allele frequency, leading to bias in assessment of the allele–cancer relationship. The bias in these examples can be reduced, however, by the standard epidemiologic tools for controlling confounding, such as adjustment for wine or olive oil consumption.

In our quantification of bias, the true impact of population stratification is exaggerated rather than underestimated by the conservative assumptions we made. We assumed that the American descendants of Europeans have the same cancer rates as current residents of their country of ancestry. In fact, cancer rates of the descendants of migrants approach those of their new country of residence in subsequent generations as their lifestyles become more similar to those of their neighbors (25). For ex-

ample, some of the difference in bladder cancer rates among European countries could be because of smoking black versus blond tobacco. Since there is almost no smoking of black tobacco in the United States, it could not be an important cause of interethnic differences in the United States. In addition, we did not allow for any adjustment of international differences in cancer rates by known risk factors, which would be controlled for in a well-conducted epidemiologic study. Finally, we assumed no intermarriage across descendants from different European ethnic groups. The cancer rates and the allele frequencies in those of mixed ancestry will lie between those of their parents, leading to less variability and, therefore, less bias.

One important and previously unrecognized conclusion from our study is that ethnic diversity actually reduces the bias from population stratification. With homogeneity ($K = 1$ ethnic group), there is no opportunity for confounding; $K = 2$ and $K = 3$ are the worst situations, and then with $K = 4-8$, the CRR moves closer to 1 as the number of ethnic groups increases further. Why is this so? We know that the direction for confounding bias from any group is positive when the genotype frequency and cancer rates are both above or both below the overall frequencies and rates; it is negative when one is above and the other is below. When there are only two ethnic groups, there is only one direction of bias. As the number of groups increases, however, the relationship between genotype frequency and cancer rate will not be monotonic or even be associated in many situations; in these instances, confounding from each group will tend to cancel as some groups contribute positive confounding and others contribute negative confounding; thus, overall, the effect of the additional ethnic groups is likely to dampen the bias. Since equation 2 applies within each ethnicity as well as among ethnicities, it is unlikely that microstratification, or residual confounding from further stratification within ethnic groups, will have a major effect.

We have focused primarily on bias from population stratification in studies of descendants of Europeans who immigrated to the United States. In studies that include, but fail to distinguish among, descendants of immigrants from Africa, Asia, and Europe, we would expect larger differences in allele frequency but a dampening of differences in cancer rates from the country of origin to prevailing U.S. rates. As Table 4 shows, even with a threefold difference in cancer rates across eight groups, the amount of bias would be modest in most circumstances, even for a range of genotype frequency of 1%–99%. Even a crude adjustment for region of origin would reduce bias further. When there is admixture within these groups, the amount of bias is likely to be even less; however, the precise quantification of the bias is difficult because of the paucity of data.

If bias from population stratification is so small, why has the issue been so controversial? Concern for population stratification seems to have originated in part with the Pima Indian example (6). However, there the frequency of the genetic variant is 75% and the disease rate 15% in those of pure tribal descent in contrast to 1% and 60% in those with no tribal ancestry (Table 1); these are extreme differences in disease rate and in allele frequency. Also, only two ethnic groups are considered. Even so, the bias factor from population stratification is only 0.64. In any case, if the rate of disease varies strongly by ethnicity, standard practice in epidemiology would be to focus not on the crude association but instead on the estimate of effect adjusted for ethnicity. Furthermore, there has also been

speculation that the inconsistent association between the D_2 dopamine receptor locus DRD2 and alcoholism (26), which has rates of disease that may vary more than cancer by ethnic group, is a consequence of population stratification (27); however, a design flaw in the original study or chance is probably a more likely explanation than population stratification for the failure of subsequent studies to confirm the association (27).

In addition, theoretical arguments have overstated the evidence for important bias from population stratification. One attempt to assess bias considered scenarios with mixtures of only two groups and with differences in genotype frequencies and disease rates that are unlikely to be applicable to studies of common variants in the cancer setting (3). Furthermore, it is tempting to infer that any bias that exists in a study of two ethnic groups will be amplified when additional distinct groups are included; we show that, in fact, the bias tends to become smaller. Also, some argue that ethnic groups can be difficult to distinguish; we think that major bias from population stratification is unlikely when this is the case because only ethnic groups that maintain their individual identities are likely to remain endogamous and retain any important differences in genotype frequencies and cancer rates. Finally, Ewens and Spielman (28) assert that the usual chi-square association statistic is “not an appropriate test” of the hypothesis of no association in subdivided populations. But, in fact, the bias in the test statistic will not be large when the CRR is small. Just as the success of epidemiology in identifying environmental causes of cancer has been achieved without perfectly valid models, so too can epidemiologic studies that incorporate genetic as well as environmental factors tolerate some slightly biased test statistics.

Some (3) have raised population stratification as a threat to the study of gene–environment interaction. The interaction between a genotype and a dichotomous factor on a multiplicative scale is the ratio of the odds ratios in the levels. If the CRRs in the strata are all near 1 or all of similar magnitudes, the distortion of the estimate of interaction will be small. When the magnitudes of the bias are large and in opposite directions for some strata, the distortion in the estimate of the interaction may be larger; it is difficult, however, to conceive a realistic situation for this scenario.

We have focused on the bias in estimation, not on hypothesis testing. The CRR is a way to evaluate theoretically the possible direction and extent of bias in a given study of an allele and disease; for example, it can be used to argue that population stratification should not be a concern when the magnitude of the allele–disease effect is greater than the worst-case CRR. The probability of a resulting false-positive finding (i.e., type I error) does depend on the magnitude of the bias, the prevalence of the genotype of interest, and the sample size. With a large enough sample size, any positive bias will lead to statistical significance. However, assessment of causality requires consideration of the estimate of effect as well as significance.

Our results suggest that bias from population stratification is unlikely to be substantial when an epidemiologic study of common genetic polymorphisms and cancer is properly designed, conducted, and analyzed. In general, cancer researchers choosing the best setting and most appropriate design for a study of the effect of a common specific genotype on risk of cancer should be more concerned about ascertainment of case patients, selection of control subjects, response rates, power, cost, and statistical efficiency than about population stratification.

REFERENCES

- (1) Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994; 265:2037–48.
- (2) Altshuler D, Kruglyak L, Lander E. Genetic polymorphisms and disease [letter]. *N Engl J Med* 1998;338:1626.
- (3) Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. *Am J Epidemiol* 1999;149:693–705.
- (4) Anonymous. Freely associating. *Nat Genet* 1999;22:1–2.
- (5) Strachan T, Read AP. Human molecular genetics. Oxford (U.K.): BIOS Scientific; 1996.
- (6) Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* 1988;43:520–6.
- (7) Caporaso N, Rothman N, Wacholder S. Case-control studies of common alleles and environmental factors. *Monogr Natl Cancer Inst* 1999;26:25–30.
- (8) Miettinen OS. Components of the crude risk ratio. *Am J Epidemiol* 1972; 96:168–72.
- (9) Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst* 1959;22:173–203.
- (10) Breslow NE, Day NE. Statistical methods in cancer research. Lyon (France): International Agency for Research on Cancer; 1980.
- (11) Parkin DM, Whelan SL, Ferlay J, Raymond L, Young J. Cancer incidence in five continents. Vol. VII. Lyon (France): IARC Press; 1997.
- (12) Marcus PM, Vineis P, Rothman N. NAT2 slow acetylation and bladder cancer risk: a meta-analysis of 22 case-control studies conducted in the general population. *Pharmacogenetics* 2000;10:115–22.
- (13) Millikan RC, Pittman GS, Newman B, Tse CK, Selmin O, Rockhill B, et al. Cigarette smoking, *N*-acetyltransferases 1 and 2, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 1998;7:371–8.
- (14) U.S. Bureau of the Census. 1990 census of population. Vol. 2. Social and economic characteristics. United States. Washington (DC): US Govt Print Off; 1993.
- (15) Roots I, Drakoulis N, Brockmoller J, Janicke I, Cuprunov M, Ritter J. Hydroxylation and acetylation phenotypes as genetic risk factors in certain malignancies. In: Kato R, Estabrook RW, Cayen MN, editors. *Xenobiotic metabolism and disposition*. London (U.K.): Taylor & Francis; 1989. p. 499–506.
- (16) Holland P, Barry M, Feely J. Pharmacogenetics and drug metabolism: an Irish perspective. *Ir J Med Sci* 1991;160:54–6.
- (17) Evans DA, Eze LC, Whibley EJ. The association of the slow acetylator phenotype with bladder cancer. *J Med Genet* 1983;20:330–3.
- (18) Fantoli U, Cattaneo C, De Ritis GC, Belasio L. Metabolismo dell'idrazide dell'acido isonicotinico. *Ann Inst Forl* 1963;23:115–42.
- (19) Kergueris MF, Bourin M, Larousse C. Pharmacokinetics of isoniazid: influence of age. *Eur J Clin Pharmacol* 1986;30:335–40.
- (20) Smith CA, Wadelius M, Gough AC, Harrison DJ, Wolf CR, Rane A. A simplified assay for the arylamine *N*-acetyltransferase 2 polymorphism validated by phenotyping with isoniazid. *J Med Genet* 1997;34:758–60.
- (21) Drozd M, Gierek T, Jendryczko A, Pilch J, Piekarska J. *N*-Acetyltransferase phenotype of patients with cancer of the larynx. *Neoplasma* 1987;34:481–4.
- (22) Paulsen O, Nilsson LG. Distribution of acetylator phenotype in relation to age and sex in Swedish patients. A retrospective study. *Eur J Clin Pharmacol* 1985;28:311–5.
- (23) Cavalli-Sforza LL, Menozzi P, Piazza A. The history and geography of human genes. Princeton (NJ): Princeton University Press; 1994.
- (24) Axelson O, Steenland K. Indirect methods of assessing the effects of tobacco use in occupational studies. *Am J Ind Med* 1988;13:105–18.
- (25) Thomas DB, Karagas MR. Migrant Studies. In: Schottenfeld D, Fraumeni JF Jr, editors. *Cancer epidemiology and prevention*. 2nd ed. New York (NY): Oxford University Press; 1996. p. 236–54.
- (26) Blum K, Noble EP, Sheridan PJ, Montgomery A, Ritchie T, Jagadeeswaren P, et al. Allelic association of human dopamine D2 receptor gene in alcoholism. *JAMA* 1990;263:2055–60.
- (27) Gelernter J, Goldman D, Risch N. The A1 allele at the D2 dopamine receptor gene and alcoholism. A reappraisal. *JAMA* 1993;269:1673–7.
- (28) Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* 1995;57:455–64.

NOTES

We thank Yukchung Kam, Walt Whitman High School, Bethesda, MD, and Carnegie-Mellon University, Pittsburgh, PA, for his help with the programming and Ms. Rebecca Albert for her help in the preparation of this manuscript.

Manuscript received October 15, 1999; revised May 12, 2000; accepted May 16, 2000.