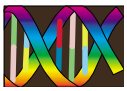
Genetic Case-Control Association Studies in Neuropsychiatry



Patrick F. Sullivan, MD, FRANZCP; Lindon J. Eaves, PhD, DSc; Kenneth S. Kendler, MD; Michael C. Neale, PhD

Case-control association studies use genetic markers as putative etiologic risk factors. The approach is controversial and has tended to produce associations in neuropsychiatry that do not stand the test of time. We studied the processes that can bias the outcomes away from a true representation of the relationship between a genetic marker and a neuropsychiatric disorder. If conducted with care and mindfulness of the potential pitfalls, case-control association studies can be an important tool for psychiatric genetic research.

Arch Gen Psychiatry. 2001;58:1015-1024

Case-control association studies are an increasingly popular approach to identifying genes that cause neuropsychiatric disorders.[1-3] Their appeal is certainly influenced by notable successes in complex disorders[4,5] (eg, Alzheimer disease[6] and type 1 diabetes mellitus[7]). The availability of the primary sequence of the human genome will hasten the already burgeoning use of genetic markers as risk factors in neuropsychiatry. These studies are often relatively easy to conduct, and genotypic data are among the most reliable and inexpensive biological markers.

Case-control association studies are often viewed as alternative or complementary to linkage studies, whose application in neuropsychiatry has not yielded the spectacular successes seen with mendelian disorders.[8,9] Briefly, linkage studies[4,5,10-12] investigate correlations between a disease and inheritance of specific chromosomal regions in families, whereas association studies focus on differences in the frequency of specific genetic markers in groups of affected and unaffected individuals.

Despite their popularity, case-control association studies remain controversial.[13-16] Moreover, they have been the source of considerable confusion, with several prominent examples of the excitement engendered by a significant association between a neuropsychiatric disorder and a genetic marker in an initial study yielding to multiple nonreplications.[8,9,17] In this nontechnical review, we hope to provide a cogent discussion of the key issues in the design, analysis, and interpretation of case-control association studies.

From the Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond.

GENETIC MARKERS AS RISK FACTORS

Excellent introductions to molecular genetics are available in print[18] and on the Internet.[19,20] Many human genes are essentially invariant, as mutations lead to death or severe dysfunction. However, there is enormous variability in the genome, and the genetic markers used as risk factors in case-control association studies capitalize on this variation. For

See also pages 995 and 1005

some markers, specific alleles (variants of a genetic marker at a specific genomic location) may be known to be associated (or not associated) with alterations in protein expression or function. In most instances, the functional consequences of alleles at a marker are unknown; resolving this knowledge gap will be a major area of inquiry in the future. It is important not to dismiss the potential importance of genetic markers with no known functional consequences. Although a marker may not itself have functional consequences, it may be located near a functional mutation. The "signal" from a causal mutation might be

		In "Truth," an Association Is	
		Absent	Present
Case-Control Study Result	Not Significant	True Negative (Correct Outcome) *Explanation:* No Etiologic Relationship	False Negative (Incorrect Outcome) *Explanations:* Experimental bias Chance (Type II Error) Inappropriate Controls (Overmatching)
	Significant	False Positive (Incorrect Outcome) *Explanations:* Experimental Bias Chance (Type I Error) Inappropriate Prior Probability Inappropriate Controls (Stratification)	True Positive (Correct Outcome) *Explanations:* Noncausal Correlation Causal Relationship

Figure 1. Outcomes of a case-control study.

detected in nearby genetic markers that of themselves have no functional significance.[2,3,11,12]

The use of genetic markers as putative risk factors[21] has many attractive qualities. The sine qua non of causation is that a cause precedes its effect in time.[22,23] It is generally and often implicitly assumed that "exposure" to a genetic risk factor is lifelong and precedes its effects. This may be a reasonable assumption, but there is a critical caveat. Because gene expression is controlled in a complex manner, a gene may be differentially expressed only in specific circumstances (eg, in a developmental period or after some exposure, life event, or injury) or at a specific anatomic location. An individual's genotype at a particular genetic marker is a static representation that may or may not accurately reflect temporal patterns of exposure. From the perspective of measurement, genotyping can be accomplished with high reliability (reproducibility rates are usually <1%) at any point in a person's life (including the prenatal and postmortem periods). Genotyping is technically accessible and possible in high volumes. A cost of $1 (or less) per genotype makes it one of the cheapest biological markers in neuropsychiatry. Technical advances make it possible to generate thousands of genotypes rapidly.[24-27] DNA of sufficient quantity and quality for thousands of genotypes is readily obtained from a variety of tissue sources (eg, blood or buccal epithelial cells).

CASE-CONTROL STUDIES

The typical approach is to ascertain cases with a trait of interest and controls without the trait, obtain DNA samples, and genotype all subjects for a genetic marker believed to be of etiologic relevance. Statistical analysis compares allele or genotype frequencies[28] in cases vs controls.[12] As with any case-control approach, there are numerous sources of bias[29]; considerable care must be taken to ensure that cases and controls are representative of the 2 arms of the population in which one is interested. Fundamentally, cases and controls should represent "identical" samples from a single population except for the diagnostic differences. The following are some issues in defining cases and controls for association studies.

- Subjects should be unrelated, in the conventional use of the term.
- Investigators should adopt a lifelong perspective. Case definitions should allow for the lifelong (rather than current) presence of a disorder and, ideally, for its lifelong absence in controls.
- Cases and controls should have passed through comparable amounts of the period of risk for the disorder under study.
- Investigators should recognize the limitations of psychiatric nosologies.[30,31] The relationship between these case definitions and genetic etiologic factors may not be precise. It may be advisable to investigate genetic influences on a more basic trait (eg, neuroticism) rather than on a disorder (eg, major depression).
- It may be beneficial to focus on the subtypes of a disorder for which genetic influences may be more likely (eg, recurrent or early-onset major depression).[1]
- Some conditions in psychiatry involve conditional disease processes, in which a preceding event must precede the subsequent trait (M.C.N., K.S.K., and Edward Harvey, PhD, unpublished data, February 2001). Examples include initiation of drug use and substance dependence or childbirth and postpartum depression. The difficulty lies in the choice of a control group, as controls should have experienced the preceding event and yet not developed the subsequent trait.

FOUR OUTCOMES

Assume that the results of the case-control study are dichotomous, ie, that there is either a significant or a nonsignificant association between a trait (present in cases, absent in controls) and a risk factor. If we assume further that we can discern the actual truth of an association, then significant or nonsignificant results can be further subdivided. **Figure 1** depicts the outcomes of a case-control study, with the 4 cells representing the combinations of the study result and omniscient knowledge of whether an association actually exists. The cells correspond to 4 outcomes: whether the study results are true negative, false negative, false positive, or true positive. We use this schema to illustrate issues in the design, analysis, and interpretation of a case-control association study.

True-Negative Result

Such findings represent the situation where the study result and "truth" are in agreement: the study outcome is correct, as no association is detected and none exists.

False-Negative Result

Experimental Bias. Many types of bias (ie, systematic effects that deflect study results away from truth) have been described in case-control studies.[23,29,32] We consider here 3 experimental biases related to genotyping. First, genotyping is a complex undertaking and a multi-

tude of technical factors can lead to erroneous data. Considerable care must be taken to ensure that genotypes are valid and reliable. Second, genotypes must be determined with blinding to case-control status to minimize the risk of a result influenced by an investigator's preconceptions, particularly as many genotyping methods require subjective judgments. Third, genotyping procedures should not differ in cases and controls. For example, bias could result if cases and controls are genotyped in separate batches months apart. These biases can—depending on the circumstances—lead to false-negative or false-positive results.

Chance (Type II Error). The statistical power of a study is its probability of detecting a true effect. Power ranges from unity to a lower limit determined by the significance criterion. Adequate power in neuropsychiatry is usually considered to be a probability of 0.80 or greater of detecting an effect of moderate size. Adequate power is no guarantee: a study with power less than 0.80 might get lucky and a study with power of 0.80 or more (but less than unity) might get unlucky. Statistical power is determined by the significance level (α), the number of cases (n_1) and controls (n_0), the prevalence of the risk factor in controls (p_0), and the relative risk conferred by the risk factor (R).[32,33] Power increases with α, sample size, and R and varies in a biphasic manner with p_0.

Power calculations for a case-control association study must take into account that humans are diploid and genotypes are composed of 2 alleles. Instead of simply using the allele frequencies in controls as p_0, interactions between the alleles (eg, dominant or recessive) must be taken into account to yield the "effective prevalence" of the genetic marker. For example, the effective prevalence of a genetic risk factor with an allele frequency of 0.20 is 0.36 (derived from calculating $[2 \times 0.20] - 0.20^2$) under a dominant model and 0.04 (which is 0.20^2) under a recessive model (assuming Hardy-Weinberg equilibrium). **Figure 2** illustrates statistical power for recessive and dominant alleles.

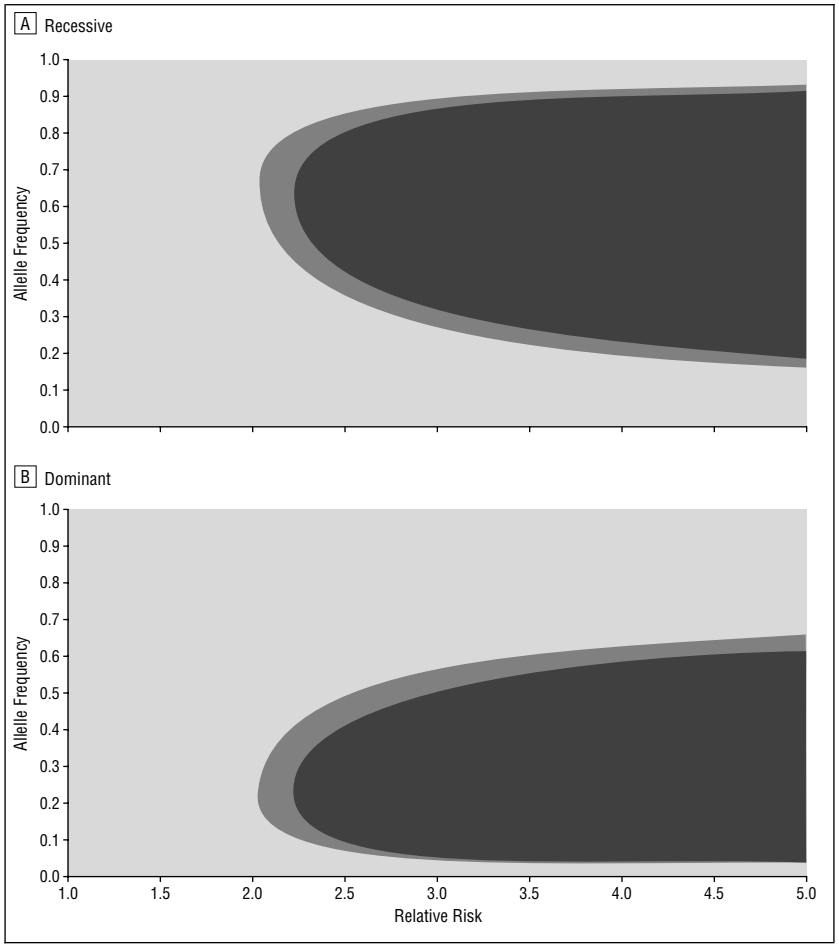


Figure 2. Contour plots showing the power of a case-control association study (analytic calculations assuming $\alpha = .05$, n = 100 per group, and Hardy-Weinberg equilibrium). A, Fully recessive model of gene action. B, Fully dominant model. The relative risk conferred by the "effective prevalence" of the increasing allele is on the x-axis and the increasing allele frequency, on the y-axis. Black indicates power of 0.80 or more; dark gray, power greater than or equal to 0.70 but less than 0.80; and light gray, power less than 0.70.

The calculations depicted in Figure 2 assume a simple and direct relationship between the genetic risk factor and the trait. In particular, the calculations assume that genetic or locus heterogeneity (cases result from different genes in different individuals) and etiologic heterogeneity (cases also result from nongenetic factors) are absent. Heterogeneity is likely for neuropsychiatric disorders, making the calculations in Figure 2 highly optimistic.

Inappropriate Control Group (Overmatching). A false-negative result can occur when cases and controls are matched for a variable correlated with the outcome.[32] For example, assume that a gene predisposes individuals to high levels of neuroticism, which increases the risk for major depression. If cases with major depression and controls without major depression are also matched for neuroticism, then statistical power to detect true differences in gene frequency predisposing to major depression will be reduced.

False-Positive Result

Chance (Type I Error). The number of significant results expected by chance can be calculated with the binomial theorem,[34] which is applicable to statistically independent or unlinked genetic markers (more complex methods are needed for linked markers). The probability of 1 or more positive results expected by chance equals $1 - (1 - \alpha)^m$, where α is the significance level and m is the number of independent markers. The Bonferroni correction $(1 - [1 - 0.05]^{1/m})$ yields a constant false-positive probability of 0.05 no matter the size of m, but is conservative.[23]

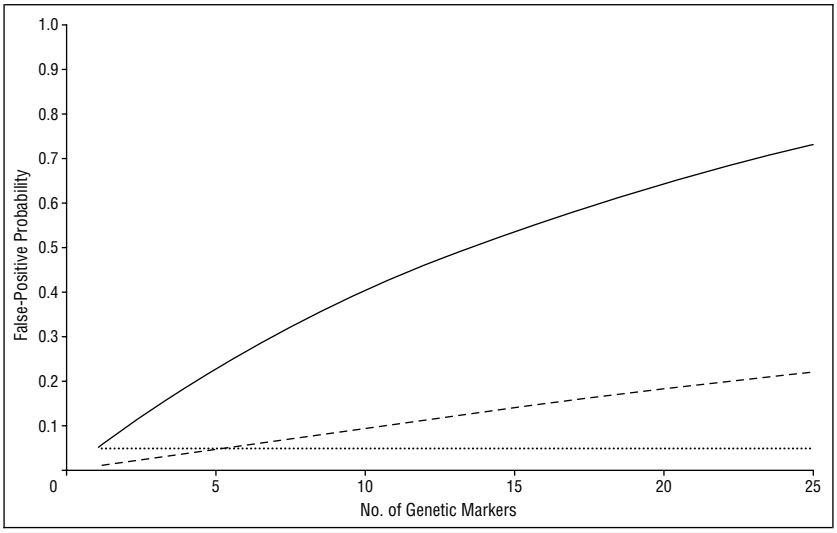


Figure 3. False-positive probability as a function of the number of independent statistical comparisons (or, in this instance, the number of unlinked genetic markers) for $\alpha = .05$ (solid line; analytic calculations). The dashed line depicts the function for $\alpha = .01$; the dotted line, for the Bonferroni correction.

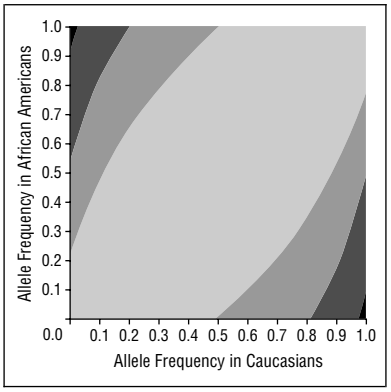


Figure 4. Contour plot of the impact of different combinations of allele frequencies in 2 groups of different ancestry (analytic results). Assumptions are as follows: independence of the allele and trait; 200 cases and 200 controls; 2 ancestral groups randomly allocated to cases and controls; 80% of the sample from the first group and 20% from the second group; and trait prevalences of 17.9% and 11.9% in the first and second groups, respectively. The contours show *P* values due to stratification and are in 4 shades: the lightest indicates $.5 < P \leq 1.0$; second lightest, $.25 < P \leq .5$; second darkest, $.1 < P \leq .25$; and darkest, $P \leq .1$.

Figure 3 illustrates the relationship between the number of unlinked genetic markers and the probability of a false-positive finding. By definition, the probability is constant at 0.05 for the Bonferroni correction. For $\alpha = .05$, the probability ranges from 0.05 for 1 marker to 0.72 for 25 markers. In the near future, it may become straightforward to generate thousands of genotypes per person. The risk of a false-positive finding due to multiple statistical tests is of particular concern in case-control association studies given the enormous number of genes in the human genome and the increasing ease of obtaining genotypes.

In addition, many investigators use multiple diagnostic schemes (eg, narrow, intermediate, and broad definitions of schizophrenia), which can increase the chance of false-positive results. As these schemes are usually not independent, adjustment for multiple comparisons is complex (akin to the situation for linked genetic markers).

Inappropriate Prior Probability. Given the vast size of the human genome, the prior probability that a randomly selected gene will be associated with a neuropsychiatric disorder is very small. Our knowledge of the etiology of these disorders is imprecise and does not strongly implicate candidate genes. Therefore, as has been noted,[14] there is a strong chance that most statistically significant candidate genes will be false positives.

Inappropriate Control Group (Population Stratification). False-positive findings resulting from inappropriate controls is a concern for many methodologic approaches; the appropriate control group for an association study is particularly controversial[13] because of population stratification (a sample composed of ≥2 subgroups with differing genetic histories) resulting from admixture. A sample drawn from many urban centers will consist of individuals from quite different genetic backgrounds (eg, sub-Saharan indigenous African, indigenous Asian, and European). A false-positive result can occur in an admixed or stratified population if a trait is more prevalent in one ethnic group and if the frequency of the genetic marker also differs by ethnicity. Of particular concern is cryptic admixture whereby a sample that appears to be homogeneous by self-reported ethnicity is nonetheless admixed.

The classic example is a hypothetical case-control association study of chopstick use and a genetic marker (eg, HLA) that differs markedly between Asians and Europeans in a city like San Francisco.[4] As the trait prevalence and marker frequencies are very different in the 2 subgroups, an association study might find that HLA "causes" chopstick use when the association is merely an artifact of population stratification.

We are aware of 3 general approaches to the problem of population stratification. First, it can be argued that differences in genotype frequency and/or disease frequency must be fairly gross for stratification to produce an artifactually significant result[35] and that simply controlling for reported ancestry will greatly reduce the risk.[35,36] Consider a case-control association study of major depression in which the sample of 200 cases and 200 controls contains European Americans (lifetime prevalence of major depression, 17.9%) and African Americans (lifetime prevalence, 11.9%)[37] in proportions similar to those found in Virginia (80% and 20%, respectively). **Figure 4** depicts the *P* values that result from a null hypothesis of no stratification effect across different combinations of allele frequencies in European and African Americans. Rejecting the null hypothesis (ie, detecting a "significant" association that results solely from stratification) is tenable only when the allele frequencies are highly discrepant in the 2 groups. In general, stratification effects become more probable with increas-

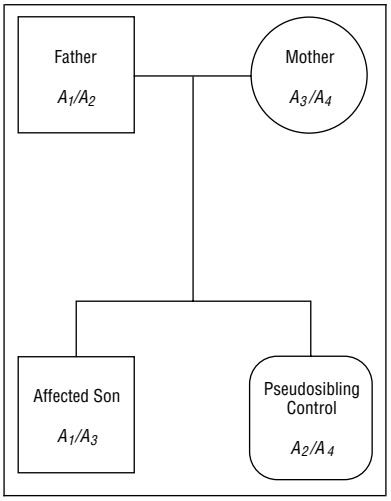


Figure 5. Diagram showing how the genotype of an affected individual (A_1/A_3) combined with parental genotypes (father A_1/A_2 and mother A_3/A_4) can be used to create a matched pseudosibling control (A_2/A_4) to eliminate the potential for a false-positive result due to population stratification.

ing sample size, with greater trait differences between groups, and as the mixture of groups approaches 50:50.

The second approach is based on study design, with "family-based" association studies the most widely discussed approach.[12,38] In its most basic form, the design requires genotypes on affected probands (cases) and both biological parents. For a particular genetic marker, assume the following genotypes for a family triad: father A_1/A_2, mother A_3/A_4, and the affected proband A_1/A_3 (**Figure 5**). The parents transmitted to their affected son alleles A_1 (father) and A_3 (mother). In a singularly clever twist, the control for the affected son is the imaginary "pseudosibling" with the untransmitted alleles (A_2/A_4). Analysis of alleles transmitted to affected offspring and alleles transmitted to pseudosiblings yields a matched comparison.

The advantage of family-based association studies is that the control group is ideally matched and impervious to false-positive results caused by admixture. This design has been expanded in many useful ways (eg, incorporating interactions between genes and environmental risk factors).[39-41] However, relatively optimistic[42] power calculations for family-based association studies[16,43] have been shown to be problematic in the presence of heterogeneity[44] (note that heterogeneity also erodes power for classic case-control association designs). Finally, this design can be difficult procedurally, as the unit of analysis is 3 people, any one of whom may be uncooperative or unavailable. More important, this could also introduce important biases if complete triads differ systematically from the larger sample of affected individuals. For example, the resulting sample of complete families could have an earlier age at onset or lesser severity than incomplete families. Preliminary data suggest that this might be of concern for bipolar disorder and schizophrenia.[45]

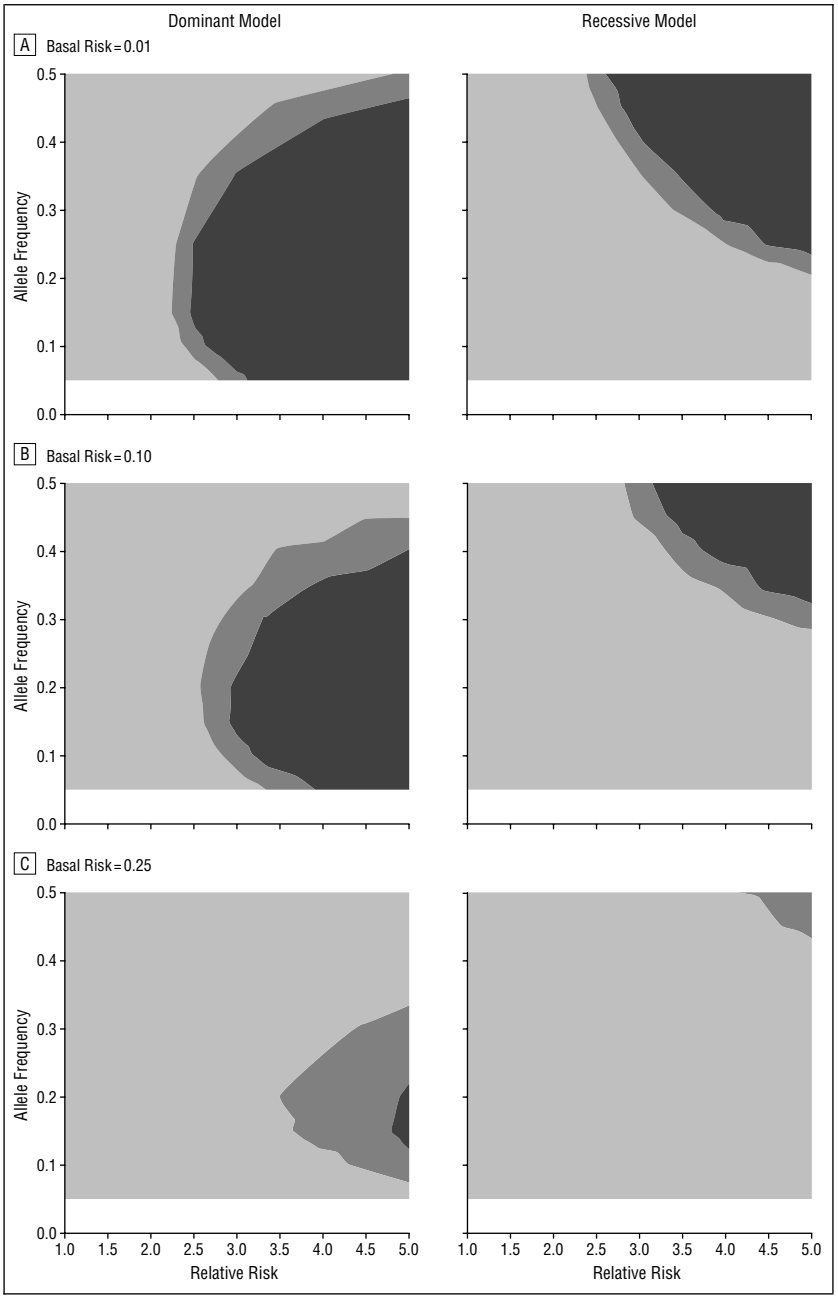


Figure 6. Contour plots demonstrating the power of a family-based association study (assuming $\alpha = .05$ and 100 triads composed of an affected offspring and both biological parents) for dominant and recessive models of gene action. Power was estimated via simulation with 1000 repetitions per parameter combination. Black indicates power of 0.80 or more; dark gray, power greater than or equal to 0.70 but less than 0.80; and light gray, power less than 0.70.

Figure 6 illustrates the statistical power of the family-based association design. As it is widely suspected that samples of individuals with identical *DSM-IV* diagnoses will

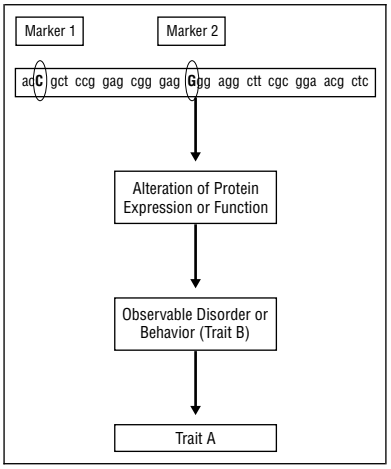


Figure 7. A highly simplified illustration of how a true-positive association might not be a causal path. A portion of the DNA sequence of the human dopamine transporter is depicted at the top of the diagram with 2 fictitious polymorphisms (markers 1 and 2). Intermarker distances are usually considerably greater than shown here. Mutation in marker 2 leads to altered protein expression or function and thence to an observable characteristic (trait B), which in turn causes trait A.

be a mixture (eg, affected because of nongenetic reasons or different genes), we added a parameter to describe the baseline risk in the absence of the effects of the gene under study to make these analyses more realistic. If this parameter is small (eg, 1%), then cases are rarely caused by any effects other than the genetic locus being studied. As this parameter becomes larger (10% or 25%), then cases are more often caused by environmental effects or by genes other than the one being studied (etiologic and genetic heterogeneity).

Figure 6 depicts contour plots for the statistical power of allele frequency and relative risk for 6 combinations of genetic model (dominant or recessive) and baseline risk (1%, 10%, and 25%). Power declines sharply as the baseline risk increases. Many combinations of relative risk and allele frequency possess low statistical power.

The third approach for guarding against the risk of a false-positive result caused by population stratification is empirical. These methods use genotypes at locations in the genome not involved in the etiology of the trait to evaluate empirically the degree of admixture. In one method, cases and controls are genotyped for 15 to 20 additional genetic markers to construct a χ^2 test for stratification.[46] If this "diagnostic test" for stratification is positive, additional genotyping ($\geq$100 unlinked loci per subject) can be used to assign individuals to subgroups based on empirical ancestry and to test for association within subgroups.[47] Another approach is to use "genomic control"; with additional genotyping (>60 markers per subject), the degree of stratification within a sample can be computed and used to adjust the test statistic for a genetic marker.[36,48] Although these approaches appear promising, actual experience with them is limited.

True-Positive Result

We consider last the holy grail of case-control association studies, in which the study results are in accord with truth. The awkward phrase "association study" was coined with care. If a study finds a true association between a genetic marker and a neuropsychiatric disorder, it does not necessarily follow that the gene is causal, in much the same way as a correlation between 2 variables does not necessarily represent causation.

Noncausal Pathways. A true-positive result could represent a noncausal path. It is possible that a true association could capture an aspect of a causal association although it is not itself causal. **Figure 7** depicts a simplified way in which this might occur. At the top of the diagram is a DNA sequence with 2 polymorphisms (markers 1 and 2). The causal path leads from marker 2 to a manifest change that leads to an observable characteristic (trait B). Trait B is correlated with trait A, which has no direct relationship with marker 2.

For example, a study might find an association between trait B and marker 1. Because of linkage disequilibrium,[11,12,18] the "signal" from the causal mutation (marker 2) might be detectable in nearby markers. As another example, an association between marker 2 and trait A might be detected. Although marker 2 causes trait A, the association detected does not disclose the more fundamental relationship between marker 2 and trait B.

Causal Relationship. This last outcome is a goal of case-control association studies: to elucidate the specific genetic basis of critically important neuropsychiatric disorders. Genetic causes of neuropsychiatric disorders are likely to be different from the direct and strong etiologic relationships of classic epidemiology.[21] Genes causally associated with neuropsychiatric disorders might act in an "ecological" manner[49]: effects will be subtle and variable, risk will be altered in a probabilistic manner, genetic heterogeneity will be evident, and causal genes will interact with environmental factors.[5]

FALSE-POSITIVE RESULTS AND THE NOTORIETY OF CASE-CONTROL ASSOCIATION STUDIES

Case-control association studies have the perhaps justified reputation of producing false-positive results in neuropsychiatry.[13,14] However, the critical sources of false-positive findings—experimental bias, type I error, inappropriate prior probabilities, and population stratification—can, in principle, be minimized or controlled via experimental or analytic techniques. The most important of these reasons may be the use of inappropriate prior probabilities. The human genome is enormous and the prior probability of association for a randomly selected gene is very low. As Crowe[14] noted, for $\alpha = .05$ and a low prior probability, nearly all positive results (>99%) will be false positive.

We illustrate this problem in **Table 1** and suggest a tentative set of anchors for prior probability referenced by the log transformation of the prior probability ("level"). The scale ranks hypotheses from level 5 (most unlikely) to level 0 (certainty). Given current uncertainties regarding the number of genes in the human genome (see footnotes to Table 1), the anchor points may change. A level 5 hypothesis corresponds to a gene selected at random from the genome. If we assume that half of human genes are expressed in the

Table 1. Prior Probabilities in Neuropsychiatry

Level*	Prior Probability	Odds Against†	Anchors	OR (α = .01)‡
5.0	0.00001	99 999:1	Random gene in the human genome§	0
4.5	0.00003	31 622:1	Random gene expressed in the central nervous system‖	0
4.0	0.00010	9999:1	One of the "usual suspects"¶	0
3.5	0.00032	3161:1		0
3.0	0.00100	999:1	Positional candidate gene—broad region#	0
2.5	0.00316	315:1	Positional candidate gene—narrower region#	0.2
2.0	0.010	99:1	Positional candidate gene—with follow-up work	2.4
1.5	0.032	31:1		11
1.0	0.10	9:1		42
0.5	0.32	2:1		197
0.3	0.50	1:1	Even odds	. . .
0.0	1.00	0:1	Certainty	. . .

*The level equals $-\log_{10}$ (prior probability).

†The odds against the hypothesis equals (1 – probability)/probability.

‡Assume that there is a pool of *N* genes (*N* = 1/prior probability) that are tested for association. The *N* genes are assumed to be statistically independent. One gene is causal and the other *N* – 1 genes are noncausal. The probability of detecting the causal gene is 0.80. The rightmost column shows the odds ratio (OR) for detecting the true gene assuming α = .01. The OR is unfavorable for hypotheses less likely than level 2.0.

§As of this writing (January 2001), estimates of the number of genes in the human genome vary widely but most are in the level 4.5 (32 000 genes) to level 5 (100 000 genes) range.[50]

‖If half of all genes in the human genome are expressed in the central nervous system, prior probabilities range from level 4.2 (32 000 genes) to level 4.7 (100 000 genes).

¶The "usual suspects" are serotonergic, dopaminergic, and noradrenergic genes broadly implicated in neuropsychiatry. Each "system" contains proteins involved in synthesis, storage, release, receptor binding, second messenger, and bioconversion of the neurotransmitter. Assignment of level 4.0 as a naive prior probability reflects imprecise knowledge of their involvement in psychotic, affective, anxiety, eating, impulse control, and addictive disorders.

#Assume that a linkage study implicates a genomic region and thus "positional" candidate genes. Assume further that the sex-averaged map length of the human genome is 3488 centimorgans(cM)[51] and that there are 80 000 genes in the genome. If the linkage peak is quite broad (40 cM), an average of about 920 genes would lie in the support interval (level 3). If the linkage peak is narrower (15 cM, eg, for genes of strong effect or large sample sizes[52]) an average of about 340 genes would be in the support interval (level 2.5). The level could also be decreased with follow-up work (eg, via fine-mapping and statistical analyses[53]). However, individual regions can contain many more or fewer genes than average: a linkage peak in a gene-poor region may have a considerably higher prior probability.

central nervous system and that this is characteristic of genes that influence neuropsychiatric disorders, then level 4.5 corresponds to a random gene expressed in the central nervous system. The level 4 anchor is for one of a set of several dozen genes involved in monoamine neurotransmission broadly implicated in a range of neuropsychiatric disorders ("round up the usual suspects"[54]). If prior knowledge suggested that it was highly probable that these genes were causal (eg, the HLA region in type 1 diabetes mellitus), then a lower level might be justified. As current knowledge of the etiopathology of most neuropsychiatric disorders is imprecise, we believe that only a modestly decreased level is warranted. Finally, a level 3 prior probability may be warranted for a candidate gene in a genomic region strongly implicated by a linkage study.

As shown in the rightmost column of Table 1, the odds ratio for detecting a true gene effect against the background of chance false-positive results is not favorable unless the hypothesis is at least level 2. Conversely, the risk of a false-positive result is quite high when the prior probability is in the level 2.5 to 5 range. The use of inappropriately optimistic prior probabilities could be the single most important reason for the propensity of case-control association studies to produce apparently false-positive results in neuropsychiatry.

There are basically 2 options for the researcher interested in controlling the risk of a false-positive result. First, a more stringent α level could be used. The difficulty here is that smaller α levels also mean markedly lower statistical power in a context in which many studies are already likely to be underpowered. Second, researchers could increase the prior probabilities of the genes they study. The choice of candidate genes is usually dictated by prior knowledge of the pathophysiology of a disorder. For example, for many complex traits of public health importance (eg, asthma, hypertension, and diabetes mellitus), there is a considerable body of knowledge about the anatomic and metabolic correlates that suggests a logical set of candidate genes. For most neuropsychiatric disorders, knowledge of pathophysiology is considerably more limited (in fact, elucidation of cryptic causes is usually an explicit aim of case-control association studies of these disorders). Thus, the most important way to increase prior probabilities is via replication and meta-analysis.

ESTABLISHING BELIEF IN AN ASSOCIATION

At this point, we drop the fictions underlying Figure 1. Researchers do not know the truth of an association but merely a quantitative *P* value (which may be above or below some significance level) that summarizes the results of an effortful association study. The next question is, how do we come to believe whether significant results are true or false positives or whether nonsignificant results are true or false negatives?

It is axiomatic that true-positive or true-negative findings should be replicable. It is less widely appreciated that there are valid reasons

Table 2. Recommendations for Association Studies in Neuropsychiatry

Topic	Recommendation
Phenotype	Cognizance of limitations of *DSM-IV*/*ICD-10* definitions
	Clear and explicit definitions
	Use of reliable diagnostic procedures
	Lifelong perspective
	Determination blinded to genotype
	Collect covariates (ancestry, sex, age at onset, comorbidity, etc)
Genotype	Rigorous error control and checking procedures
	Determination blind to phenotype
	Similar procedures for cases and controls
Sampling	Appropriate selection of cases and controls
False-negative results (type II error)	Power calculations that account for heterogeneous and subtle genetic effects dictate very large sample sizes
False-positive results (type I error)	Address threat of population stratification empirically or by design
	Explicit statement of number of markers genotyped or planned to be genotyped
	Use realistic prior probabilities
Interpretation	Highly cautious interpretation of results from single studies
Replication	Greatly facilitated by collaboration
	Routine use of meta-analysis
Publishing	Journals develop a format for convenient publishing of negative results

why a study result might not be replicable. A significant finding might not be replicable because of chance. Assume that 5 investigators conduct independent case-control association studies of the same genetic marker and neuropsychiatric disorder and that each study has a power of 0.80. The probability[34] of all 5 studies detecting the association is only 0.33 (assuming the association is true and the effect sizes are comparable). Furthermore, neuropsychiatric disorders are complex traits.[4,5] If these disorders result from different types of causes (genetic and etiologic heterogeneity) whose mixture varies across samples, nonreplication might occur. For example, it may be true that variation in the aldehyde dehydrogenase gene[55] markedly alters the risk of alcoholism in indigenous Asians and true that the polymorphism is rare in a second sample: nonreplication reflects true differences in etiology across samples.

SUMMARIZING THE LITERATURE

An essential part of nearly all scientific reports is a review of the previous literature. The simplest approach is to tally the positive and negative studies. This "box score" or "vote-counting" method is problematic because equal weights are given to each study. Because of equal weighting, small studies count as much as very large studies. More important, the focus is nearly entirely on whether a study produces a *P* value that is or is not below some significance level; the effect sizes are disregarded.[56] For example, the box score approach would weight studies with $P = .04$ similarly, even if the effect sizes were highly variable. The more worrisome scenario is when the studies in the literature are nonsignificant (eg, $P = .10$) yet the individual studies all possess low statistical power. In such an instance, the literature could actually support a cumulative effect that the box score approach would miss.

With additional work,[57,58] a meta-analysis[59] can be conducted. Meta-analysis improves on the box score approach by weighting studies for sample size and effect size and is an increasingly standard approach in the genetic association literature in neuropsychiatry.[60-62] Bayesian approaches may be particularly useful.[63,64] Meta-analysis assumes that the primary studies were conducted in a sufficiently comparable manner and that studies available for review are representative of the studies actually conducted. The validity of a meta-analysis may be low if there are a substantial number of unpublished negative results (the "file drawer problem"). By combining individual studies that may be underpowered, meta-analysis has the potential to bring clarity to a body of research.

CONCLUSIONS

In this brief review, we attempted to explicate the critical issues in the design, analysis, and interpretation of genetic case-control association studies. We did not cover a number of important topics (eg, the use of haplotypes in association studies and a more detailed treatment of the similarities and differences between linkage and association[5]), and we made simplifying assumptions (eg, considering only biallelic markers). However, we attempted to provide references for the interested reader.

We suggest that case-control association studies are a useful methodologic approach that—like experimental methods—possess potential pitfalls. If conducted with care, association studies should be an important tool in genetic and clinical neuropsychiatry. Whether or not one agrees with this perspective, the reality is that association studies are likely to be widely used in neuropsychiatry in the years to come. Therefore, we summarize in **Table 2** a set of tentative recommendations that may be of use to investigators using this scientific approach.

Many of these recommendations are self-explanatory and have analogs in other experimental designs. As has become standard in the literature, phenotypic classification must be done with care and rigor. Genotyping contains numerous potential sources of error, but, in principle, these are amenable to careful and appropriate experimental design. We discuss the crucial recommendations below.

Sampling

Inappropriate sampling can yield groups of cases and/or controls that lead to false-positive or false-negative results. This topic has been extensively discussed.[23,29,32,65]

False-Negative Results

It is relatively easy to generate optimistic power calculations for case-control association studies under the assumptions of homogeneity and large genetic effect sizes. However,

studies guided by such power calculations are not likely to possess sufficient statistical power in actuality. Investigators should design studies based on realistic power calculations, which may dictate sample sizes as much as 10 to 20 times larger than those typically reported.

False-Positive Results

Historically, the main criticisms of case-control association studies have focused on their potential to generate false-positive results. This concern is warranted and poses a major threat. However, the general focus in the literature has been on the problem of population stratification, which may be less of a threat than false-positive results due to chance and inappropriate prior probabilities. Although population stratification can yield false-positive results, a number of measures can decrease or eliminate the threat by design[12,39,66] or empirically.[36,46-48] In addition, the threat may not be large for a well-designed study.[35] In an era of increasingly high-throughput genotyping, a greater threat is false-positive results due to chance. For example, if an investigator genotypes 40 markers and publishes only the best results, the net result is misleading. Therefore, it is important that reports include an explicit statement of the numbers of markers genotyped or planned to be genotyped so that reviewers and readers can evaluate the impact of chance. The greatest threat arises from inappropriately optimistic prior probabilities (Table 1) and is perhaps an important reason underlying the patterns of nonreplication in psychiatric genetics.[14] We suggest that meta-analytic approaches are particularly valuable in this context.

True-Positive Results

Given that genetic effects appear to be important for many neuropsychiatric disorders,[1,67-70] identification of the relevant genes could have dramatic effects on clinical research and treatment development and perhaps even yield rational primary and secondary preventive strategies. At the very least, these efforts will generate new and perhaps surprising hypotheses.

This is an obvious and usually elusive goal in neuropsychiatry. Clearly, the results of single studies must be interpreted with caution. The model adopted by many in psychiatric genetics is one of collaboration and gradual progress rather than one where a single study could break open a field and of itself solve a problem. To facilitate this end, given that researchers and editors alike appear to be increasingly uninterested in publishing negative results, it is crucial that the major neuropsychiatric journals develop a means by which negative results can be published in a brief format for inclusion in meta-analytic studies.

Accepted for publication March 23, 2001.

This study was supported by grants MH59160 from the National Institute of Mental Health, AI38429 from the National Institute of Allergic and Infectious Diseases, and CA85739 from the National Cancer Institute, National Institutes of Health, Bethesda, Md.

We thank the participants in our "hard problems" seminar, an anonymous reviewer who once termed our hypothesis in an association study "weak", and the anonymous reviewers from the ARCHIVES. *For critical reviews, we thank Wade Berrettini, MD (University of Pennsylvania, Philadelphia); Cynthia Bulik, PhD (Virginia Commonwealth University, Richmond); Nancy Cox, PhD (University of Chicago, Chicago, Ill); Ian Jones, MD (University of Birmingham, Birmingham, England), Caryn Lerman, PhD (University of Pennsylvania); Charles MacLean, PhD (Virginia Commonwealth University), and Alessandro Rotondo, MD (Università di Pisa, Pisa, Italy).*

Corresponding author and reprints: Patrick F. Sullivan, MD, FRANZCP, Virginia Institute for Psychiatric and Behavioral Genetics, PO Box 980126, Richmond, VA 23298-0126 (e-mail: sullivan@psycho.psi.vcu.edu).

REFERENCES

1. NIMH Genetics Workgroup. *Genetics and Mental Disorders*. Bethesda, Md: National Institute of Mental Health; 1998.
2. Risch NJ. Searching for genetic determinants in the new millennium. *Nature*. 2000;405:847-856.
3. Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet*. 2001;2:91-99.
4. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994;265:2037-2048.
5. Terwilliger JD, Goring HH. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol*. 2000;72:63-132.
6. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM, for the APOE and Alzheimer Disease Meta Analysis Consortium. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. *JAMA*. 1997;278:1349-1356.
7. Cudworth AG, Wolf E. The genetic susceptibility to type I (insulin-dependent) diabetes mellitus. *Clin Endocrinol Metab*. 1982;11:389-396.
8. Risch N, Botstein D. A manic depressive history. *Nat Genet*. 1996;12:351-353.
9. Moldin SO. The maddening hunt for madness genes. *Nat Genet*. 1997;17:127-129.
10. Craddock N, Owen MJ. Modern molecular genetic approaches to psychiatric disease. *Br Med Bull*. 1996;52:434-452.
11. Ott J. *Analysis of Human Genetic Linkage*. Baltimore, Md: Johns Hopkins University Press; 1991.
12. Sham PC. *Statistics in Human Genetics*. New York, NY: John Wiley & Sons; 1998.
13. Kidd KK. Associations of disease with genetic markers: déjà vu all over again. *Am J Med Genet Neuropsychiatr Genet*. 1993;48:71-73.
14. Crowe RR. Candidate genes in psychiatry: an epidemiological perspective. *Am J Med Genet*. 1993;48:74-77.
15. Gambaro G, Anglani F, D'Angelo A. Association studies of genetic polymorphisms and complex disease. *Lancet*. 2000;355:308-311.
16. Risch N, Merikangas KR. The future of genetic studies of complex human diseases [published correction appears in *Science*. 1997;275:1329-1330]. *Science*. 1996;273:1516-1517.
17. Stoltenberg SF, Burmeister M. Recent progress in psychiatric genetics: some hope but no hype. *Hum Mol Genet*. 2000;9:927-935.
18. Strachen T, Read AP. *Human Molecular Genetics*. 2nd ed. New York, NY: John Wiley & Sons; 1999.
19. US Department of Energy. Primer on molecular genetics. Available at: http://www.ornl.gov/TechResources/Human_Genome/publicat/primer/intro.html. Accessed June 1, 2001.
20. Online Mendelian Inheritance in Man (OMIM™). Available at: http//www3.ncbi.nlm.nih.gov/omim/. Accessed June 1, 2001.
21. Kraemer HC, Kazdin AE, Offord DR, Kessler RC, Jensen PS, Kupfer DJ. Coming to terms with the terms of risk. *Arch Gen Psychiatry*. 1997;54:337-343.
22. Hill AB. The environment and disease: association or causation. *Proc R Soc Med*. 1965;58:295-300.
23. Rothman KJ. *Modern Epidemiology*. Boston, Mass: Little Brown & Co; 1986.
24. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES. Large-scale identification, mapping, and geno-

typing of single-nucleotide polymorphisms in the human genome. *Science*. 1998;280:1077-1082.
25. Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. Making and reading microarrays. *Nat Genet*. 1999;21(1, suppl):15-19.
26. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet*. 1999;22:139-144.
27. Gray IC, Campbell DA, Spurr NK. Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet*. 2000;9:2403-2408.
28. Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics*. 1997;53:1253-1261.
29. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32:51-63.
30. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Association; 1994.
31. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. Geneva, Switzerland: World Health Organization; 1993.
32. Schlesselman JJ. *Case-Control Studies: Design, Conduct, Analysis*. New York, NY: Oxford University Press; 1982.
33. Fleiss J. *Statistical Methods for Rates and Proportions*. 2nd ed. New York, NY: John Wiley & Sons; 1981.
34. DeGroot MH. *Probability and Statistics*. New York, NY: Addison Wesley Longman Inc; 1986.
35. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst*. 2000;92:1151-1158.
36. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55:997-1004.
37. Blazer DG, Kessler RC, McGonagle KA, Swartz MS. The prevalence and distribution of major depression in a national community sample: the National Comorbidity Study. *Am J Psychiatry*. 1994; 151:979-986.
38. Schaid DJ, Sommer SS. Comparison of statistics for candidate-gene association studies using cases and parents. *Am J Hum Genet*. 1994; 55:402-409.
39. Schaid DJ. Transmission disequilibrium, family controls, and great expectations. *Am J Hum Genet*. 1998;63:935-941.
40. Fulker DW, Cherny SS, Sham PC, Hewitt JK. Combined linkage and association sib-pair analysis for quantitative traits. *Am J Hum Genet*. 1999;64: 259-267.
41. Eaves LJ, Sullivan PF. Genotype-environment interaction in transmission disequilibrium tests. In: Rao DC, Province MA, eds. *Genetic Dissection of Complex Traits*. San Diego, Calif: Academic Press; 2001:223-240.
42. Knapp M. A note on power approximations for the transmission/disequilibrium test. *Am J Hum Genet*. 1999;64:1177-1185.
43. Camp NJ. Genomewide transmission/disequilibrium testing [published correction appears in *Am J Hum Genet*. 1999;64:1485-1487]. *Am J Hum Genet*. 1997;61:1424-1430.
44. Slager SL, Huang J, Vieland VJ. Effect of allelic heterogeneity on the power of the transmission disequilibrium test. *Genet Epidemiol*. 2000;18: 143-156.
45. Brunn TG, Ewald H. Selection bias of susceptibility genes possible when using parent-offspring trios in genetic association studies. *Mol Psychiatry*. 1999;4:415-416.
46. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet*. 1999; 65:220-228.
47. Pritchard JK, Stephens M, Rosenberg NA, Donnelley P. Association mapping in structured populations. *Am J Hum Genet*. 2000;67:170-181.
48. Bacanu S-A, Devlin B, Roeder K. The power of genomic control. *Am J Hum Genet*. 2000;66:1933-1944.
49. Kaufman JS, Poole C. Looking back on "Causal Thinking in the Health Sciences." *Annu Rev Public Health*. 2000;21:101-119.
50. Genesweep. Available at: http://www.ensembl.org/genesweep.html. Accessed June 1, 2001.
51. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet*. 1998;63:861-869.
52. Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS. Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Genet*. 1999;65: 876-884.
53. Cox NJ, Frigge M, Nicolae DL, Concannon P, Hanis CL, Bell GI, Kong A. Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. *Nat Genet*. 1999;21:213-215.
54. Burnett M, Alison J. *Casablanca* [motion picture]. Los Angeles, Calif: Warner Brothers; 1942.
55. Li TK. Pharmacogenetics of responses to alcohol and genes that influence alcohol drinking. *J Stud Alcohol*. 2000;61:5-12.
56. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Assoc; 1988.
57. Normand S-LT. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med*. 1999; 18:321-359.
58. Petitti DB. *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis*. New York, NY: Oxford University Press; 1994.
59. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol*. 1977;32: 752-760.
60. Williams J, Spurlock G, Holmans P, Mant R, Murphy K, Jones L, Cardno A, Asherson P, Blackwood D, Muir W, Meszaros K, Aschauer H, Mallet J, Laurent C, Pekkarinen P, Seppala J, Stefanis CN, Papadimitriou GN, Macciardi F, Verga M, Pato C, Azevedo H, Crocq MA, Gurling H, Owen MJ. A meta-analysis and transmission disequilibrium study of association between the dopamine D3 receptor gene and schizophrenia [published correction appears in *Mol Psychiatry*. 1998;3:458]. *Mol Psychiatry*. 1998;3:141-149.
61. Arranz MJ, Munro J, Sham P, Kirov G, Murray RM, Collier DA, Kerwin RW. Meta-analysis of studies on genetic variation in 5-HT2A receptors and clozapine response. *Schizophr Res*. 1998;32:93-99.
62. Turecki G, Rouleau GA, Mari J, Joober R, Morgan K. Lack of association between bipolar disorder and tyrosine hydroxylase: a meta-analysis. *Am J Med Genet*. 1997;74:348-352.
63. Goodman SN. Toward evidence-based medical statistics, 1: the P value fallacy. *Ann Intern Med*. 1999; 130:995-1004.
64. Goodman SN. Toward evidence-based medical statistics, 2: the Bayes factor. *Ann Intern Med*. 1999; 130:1005-1013.
65. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bull*. 1946;2:47-53.
66. Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. *J Natl Cancer Inst Monogr*. 1999;26:31-37.
67. Kendler KS. Overview: a current perspective on twin studies of schizophrenia. *Am J Psychiatry*. 1983;140:1413-1425.
68. Sullivan PF, Kendler KS. The genetic epidemiology of "neurotic" disorders. *Curr Opin Psychiatry*. 1998;11:143-147.
69. Sullivan PF, Kendler KS. The genetic epidemiology of smoking. *Nicotine Tob Res*. 1999;1(suppl 2):549-555.
70. Sullivan PF, Neale MC, Kendler KS. The genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000;157:1552-1562.