Am. J. Hum. Genet. 66:1933–1944, 2000

The Power of Genomic Control

Silviu-Alin Bacanu,[1] B. Devlin,[1] and Kathryn Roeder[2]

[1]Department of Psychiatry, University of Pittsburgh, and [2]Department of Statistics, Carnegie Mellon University, Pittsburgh

Although association analysis is a useful tool for uncovering the genetic underpinnings of complex traits, its utility is diminished by population substructure, which can produce spurious association between phenotype and genotype within population-based samples. Because family-based designs are robust against substructure, they have risen to the fore of association analysis. Yet, if population substructure could be ignored, this robustness can come at the price of power. Unfortunately it is rarely evident when population substructure can be ignored. Devlin and Roeder recently have proposed a method, termed "genomic control" (GC), which has the robustness of family-based designs even though it uses population-based data. GC uses the genome itself to determine appropriate corrections for population-based association tests. Using the GC method, we contrast the power of two study designs, family trios (i.e., father, mother, and affected progeny) versus case-control. For analysis of trios, we use the TDT test. When population substructure is absent, we find GC is always more powerful than TDT; furthermore, contrary to previous results, we show that as a disease becomes more prevalent the discrepancy in power becomes more extreme. When population substructure is present, however, the results are more complex: TDT is more powerful when population substructure is substantial, and GC is more powerful otherwise. We also explore general issues of power and implementation of GC within the case-control setting and find that, economically, GC is at least comparable to and often less expensive than family-based methods. Therefore, GC methods should prove a useful complement to family-based methods for the genetic analysis of complex traits.

Introduction

The past decade has witnessed the rise of family-based designs as an alternative to, and then a replacement for, population-based association studies, such as the case-control design. This succession is motivated by the fact that, because of population substructure, case-control designs can produce spurious associations (Li 1972), whereas family-based designs, when teamed with appropriate test statistics, do not (e.g., the TDT test) (Spielman et al. 1993; Ewens and Spielman 1995). Indeed, some journals now hesitate to publish studies employing case-control designs.

Recent articles, however, have explored the relative merits of both family-based and case-control designs (e.g., see Morton and Collins 1998; Risch and Teng 1998). As an alternative to spurious associations that arise from population substructure, they suggest that the unreliability of association tests might be due to low power combined with the large number of tests conducted. Another source of false positives occurs when affected individuals are cryptically related, as might occur because they share a genetic disorder (Devlin and Roeder 1999). In this instance, test statistics for case-control studies are apt to be inflated, relative to expectations, under the assumption of an independent sample and no genetic association with the disease.

Under the assumption of little population stratification, there are two significant advantages to population-based designs. The samples can be easier and less expensive to ascertain. Furthermore, it is commonly believed (Morton and Collins 1998) that case-control studies are more powerful than family-based designs, although this opinion is not held universally.

The reality of population stratification, however, weighs against case-control designs. How can concerns about diminished statistical efficiency for family-based designs be weighed against concerns about stratification for population-based designs? In a recent article, Devlin and Roeder (1999) present a population-based association method termed "genomic control" (GC) that automatically accounts for nonindependence, in a case-control sample, caused by population stratification and cryptic relatedness. It comes at the price, however, of additional genotyping—specifically, genotyping multiple loci unlikely to affect liability. We call these "null loci," because they are assumed to have no effect on the disease under study.

For a case-control analysis of candidate genes, GC computes χ^2 test statistics for independence for both null

Received February 17, 2000; accepted for publication April 6, 2000; electronically published May 11, 2000.

Address for correspondence and reprints: Dr. Bernie Devlin or Dr. Silviu-Alin Bacanu, Department of Psychiatry, University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, PA 15213. E-mail: devlinbj@msx.upmc.edu; or bacanus@msx.upmc.edu

0002-9297/2000/6606-0022$02.00

and candidate loci. By means of the variability and magnitude of the test statistics observed at the null loci, which are inflated by the impact of population stratification and cryptic relatedness, a multiplier is derived to adjust the critical value for significance tests for candidate loci. In this way, GC permits analysis of stratified case-control data without an increased rate of false positives. If population stratification and cryptic relatedness are not detected from null loci, then GC is identical to a standard test of independence for a case-control design.

For case-control methods, GC is currently limited to single nucleotide polymorphisms (SNP) for statistical and population genetic reasons given in Devlin and Roeder (1999). This drawback is ameliorated by the fact that a dense set of SNP covering the genome is under development (Collins et al. 1998; Wang et al. 1998). In addition, efficient methods to assess SNP genotypes are expected in the near future. An outstanding question, however, is how the GC methodology compares to family-based designs in terms of statistical and economic efficiency, especially when the research goal is to identify genes affecting liability to complex disorders. For statistical efficiency, the answer likely depends on the population under study (Ewens and Spielman 1995). We address this issue in detail in this report. Economic efficiency is more difficult to assess because it changes over time. At the present time, GC is not prohibitive, and appears to compare very favorably to family-based designs. However, with the cost of genotyping dropping and the cost of ascertainment rising, we envision GC will be considerably more cost effective in the near future.

This report is organized into six major sections. A brief overview of GC is first presented ("GC Method"). Then the statistical efficiency of GC and TDT for simplex families (mother, father, and affected progeny) are contrasted when population substructure is absent ("No Population Stratification") and when it is present ("Population Stratification"). We then examine power for GC for selected versus unselected controls ("Power for GC"), with the premise that the former is more powerful and the latter less expensive. Finally, we discuss the choice of family-based versus case-control designs.

Throughout this report, we assume that alleles affect liability directly, rather than having an indirect effect mediated by linkage disequilibrium. For the latter case, the relative efficiency of trios versus GC is the subject of ongoing study. Moreover, we do not evaluate the impact of allelic heterogeneity; see Slager et al. (2000) for the effect of allelic heterogeneity on association tests.

GC Method

For the analysis of case-control data, Devlin and Roeder (1999) presented GC methods using Bayesian and frequentist techniques and, for two study designs, targeted candidate-gene analysis and association scans over large portions of the genome (e.g., see Risch and Merikangas 1996). Here, we review the frequentist version of GC for candidate-gene analysis.

For a case-control study and n biallelic markers, let N denote the number of subjects genotyped, where $0 < \phi < 1$ is the proportion of N that are cases. For our analyses, we assume ϕ is a constant near 0.5, which is typical for case-control data. The data for each marker are given by a 2×3 table of genotype by case and control (see table 1, which implicitly defines the remaining notation). To test for lack of independence, 1-df χ^2 statistics corresponding to dominant, recessive, and additive genetic models might be applied. For reasons discussed in Devlin and Roeder (1999), we choose the additive model and employ Armitage's (1955) trend test

$$Y^2 = \frac{N\{(r_1 + 2r_2) - \phi(n_1 + 2n_2)\}^2}{\phi(1-\phi)\{N(n_1 + 4n_2) - (n_1 + 2n_2)^2\}}, \quad (1)$$

which corresponds to the additive genetic model.

We assume the n loci under study consists of c biallelic polymorphisms in candidate genes and $(n - c)$ null SNP dispersed throughout the genome. For the disorder of interest, we assume the null loci have no impact on liability and that they are not in linkage disequilibrium with polymorphisms affecting liability. (As described below, these assumptions need not be met strictly.) Although the test statistic will be computed for all n loci, only the candidate gene polymorphisms will be tested for association.

For each marker locus l we obtain a statistic Y_l^2 using the trend test, $l = 1, \ldots, n$. When the marker is in linkage equilibrium with the disorder and there is no population substructure or cryptic relatedness, Y_l^2 is distributed as $\mathcal{X}_1^2(0)$. The GC model allows for extra variance by assuming that the test statistic is inflated by a factor λ; consequently, $Y_l^2/\lambda \sim \mathcal{X}_1^2(0)$. The GC-model approach is based on the assumption that the variance inflation factor λ is approximately constant across the genome for all loci that are not associated with the disorder. In Devlin and Roeder (1999), we verify that this assump-

Table 1

Genotype Frequencies for a Case-Control Design

	A Alleles			
Subjects	0	1	2	Total
Case	r_0	r_1	r_2	ϕN
Control	s_0	s_1	s_2	$(1-\phi)N$
Total	n_0	n_1	n_2	N

tion holds whenever F_{st} is approximately equal across biallelic loci.

Consequently, λ can be estimated. From the properties of the gamma distribution, a robust estimator of λ is derived (Rice 1988, p. 336), as follows: $\hat{\lambda} = \{\text{median}(Y^2_{c+1}, Y^2_{c+2}, \ldots, Y^2_n)/0.456\}$, where the loci are ordered so that $l = 1, \ldots, c$ indexes the candidate genes). More efficient estimators exist for λ, but we favor the median, because it provides a consistent estimate of the inflation factor even if a small fraction of the null loci actually affect liability to the disease or are linked to the gene under study. When n and N are large, $Y^2/\hat{\lambda}$ is approximately distributed χ^2_1, under the null hypothesis. When c candidate genes are examined, a Bonferroni correction provides the critical value for the test: $\chi^2_1(\alpha/c)$. When λ is constant across the genome, this simple adjustment will result in a test statistic with type I error rate equal to the nominal level. When λ follows a distribution across the genome, with standard deviation of the same order as the mean, then this adjustment will result in a test statistic with type I error rate roughly equivalent to the nominal level. Conditions ensuring that variability in λ is small are delineated in Devlin and Roeder (1999).

Devlin and Roeder (1999) discuss frequentist and Bayesian procedures for performing GC for a genome scan. Several frequentist outlier procedures are applicable to the situation (see Barnett and Lewis 1995, chapter 6). One simple frequentist procedure is similar to the one discussed above, except that all markers are tested for association. The modified estimate for λ is $\hat{\lambda} = \{\text{median}(Y^2_1, Y^2_2, \ldots, Y^2_n)/0.456\}$ and the critical value is $\chi^2_1(\alpha/n)$. The effect of treating outliers as null loci in the estimation of $\hat{\lambda}$ is a slight positive bias in the estimator, which has the effect of decreasing both the power and the size of the test somewhat. Because λ is robust to outliers, the impact of this on the test will be negligible. We favor this simple method over more efficient, classical methods for outlier detection, because the method is fairly robust to swamping and masking and is applicable even if the markers are spaced densely enough to possess significant spatial correlation. The Bayesian procedure presented by Devlin and Roeder (1999) models outliers in a more comprehensive manner.

For each marker, the data in table 1 also can be summarized via a 2 × 2 allelic table (table 2). Sasieni (1997) discusses the pitfalls of the use of allelic, rather than genotypic, analyses. Here we note that, in the special case of no population stratification, the two tests are essentially identical (Devlin and Roeder 1999). Consequently, the GC approach reduces to the allelic test under this condition.

Table 2

Allele Distribution for a Case-Control Design

Subjects	A_1	A_2	Total
Case	$r_1 + 2r_2$	$r_1 + 2r_0$	$2\phi N$
Control	$s_1 + 2s_2$	$s_1 + 2s_0$	$2(1 - \phi)N$
Total	$n_1 + 2n_2$	$n_1 + 2n_0$	$2N$

No Population Stratification

In this section, we assume no population stratification, and, hence, from a mathematical point of view, a comparison between the TDT and GC reduces to a comparison between the TDT and the simple allelic test for association (Sasieni 1997; Devlin and Roeder 1999). For the family-based sample, we evaluate simplex families of mother, father, and affected child (henceforth referred to as "trios") using the TDT test.

Properties of the Case-Control Sample

Assume there exists a single susceptibility locus with alleles A and a, which occur with frequencies p and $q = 1 - p$, respectively. A increases susceptibility for a disease of population prevalence K. We denote affected individuals by $\mathcal{D}$ and unaffected individuals by $\mathcal{N}$. For the moment, we follow Risch and Merikangas (1996) by considering a multiplicative model with penetrance functions $P(\mathcal{D}|aa) = f_0$, $P(\mathcal{D}|Aa) = f_1 = \gamma f_0$ and $P(\mathcal{D}|AA) = f_2 = \gamma^2 f_0$. Under the assumption that the population is in Hardy-Weinberg equilibrium, $K = q^2 f_0 + 2pqf_1 + p^2 f_2 = (p\gamma + q)^2 f_0$.

In the allele table for the case-control sample, each individual has two entries reflecting the alleles comprising their genotype (table 2). With this rule in mind, the joint probabilities $p_{A\mathcal{D}}, p_{A\mathcal{N}}, p_{a\mathcal{D}}, p_{a\mathcal{N}}$ can be calculated directly. The expected values for the allele table are given below.

$$
\begin{aligned}
p_{A\mathcal{D}} &= \frac{\phi p\gamma f_0(p\gamma + q)}{K} \\
p_{A\mathcal{N}} &= \frac{(1 - \phi)p[1 - \gamma f_0(p\gamma + q)]}{1 - K} \\
p_{a\mathcal{D}} &= \frac{\phi q f_0(p\gamma + q)}{K} \\
p_{a\mathcal{N}} &= \frac{(1 - \phi)q[1 - f_0(p\gamma + q)]}{1 - K}
\end{aligned} \tag{2}
$$

The odds ratio associated with the allelic table is:

$$\theta = \gamma\left[\frac{1 - f_0(p\gamma + q)}{1 - \gamma f_0(p\gamma + q)}\right] .$$

Under the null hypothesis H_0 of no association, $p_{A|\mathcal{D}} - p_{A|\mathcal{N}} = 0$, and, under the alternative H_a, $p_{A|\mathcal{D}} - p_{A|\mathcal{N}} \neq 0$. Under H_0 the common probability of the A allele is $p_{A\cdot} = \phi p_{A\mathcal{D}} + (1 - \phi)p_{A\mathcal{N}}$. Let $z_{\alpha/2}$ and $z_{1-\beta}$ be the $(1 - \alpha/2)$ and β percentiles of the standard normal, respectively. The total number of subjects (both cases and controls) needed for a size α test that has power $1 - \beta$ is

$$N_{GC} = \frac{\left[z^{\alpha}_{\frac{2}{2}}\sqrt{p_{A\cdot}(1 - p_{A\cdot})(\frac{1}{\phi} + \frac{1}{1-\phi})} + z_{1-\beta}\sqrt{p_{A\mathcal{D}}(1 - p_{A\mathcal{D}})\frac{1}{\phi} + p_{A\mathcal{N}}(1 - p_{A\mathcal{N}})\frac{1}{1-\phi}}\right]^2}{(p_{A\mathcal{D}} - p_{A\mathcal{N}})^2} .$$

After substituting the table probabilities (2) and simplifying the resulting expression, we find:

$$N_{GC} = \frac{(U + V)^2}{2\ q\ p\ (\gamma - 1)^2(1 - \phi)\ \phi} ,$$

where

$$U = z^{\alpha}_{\frac{2}{2}}\ \sqrt{[1 - K + p\ (1 - \phi)\ (\gamma - 1)][1 - \phi + p\ (1 - \phi)\ (\gamma - 1) - K\ \gamma + \phi\ \gamma]} ,$$

and

$$V = z_{1-\beta}\ \sqrt{(1 - K)^2\ \gamma - \phi\ (\gamma - 1)\ [q\ (1 - K - p) + (K - p)\ p\ \gamma]} .$$

Efficiency of GC versus TDT

Using the results of Knapp (1999), we obtain the number of subjects N_{TDT} (not trios) needed to obtain 80% power for the multiplicative model. The relative efficiency, defined as the ratio N_{TDT}/N_{GC}, is complex, but the limit for alternatives close to the null simplifies to a convenient form. The asymptotic efficiency (as $\gamma \to 1$) is

$$r(\phi) = lim_{\gamma \to 1}\frac{N_{TDT}}{N_{GC}} = \frac{6\ (1 - \phi)\ \phi}{(1 - K)^2} . \qquad (3)$$

With equal sampling of cases and controls, this ratio is $3/[2\ (1 - K)^2]$. It is clear from this expression why it is advantageous to sample an equal number of cases and controls: when $\phi = 0.5$, the relative power is optimized. Case-control studies are known to have reduced power when ϕ approaches 0 or 1, which affects relative efficiency.

Similar issues about efficiency were explored in the epidemiological literature of the late 1960s, in the context of matched pairs versus unpaired, case-control designs. Chase (1968) shows the Pitman-like efficiency is equal for the two procedures. That is to say, for alternatives very near the null, the two procedures have equal power. This result supports Morton and Collins (1998), who claim that TDT is two-thirds as efficient as case-control tests when trios are measured, because three genotypes are required to obtain two transmissions for the TDT test. When the prevalence is small, our result approximately matches the theoretical results obtained by Chase (1968) and the predictions made by Morton and Collins (1998). For common disorders, however, the comparison favors GC over TDT even more strongly than is predicted by the literature. Our results differ from those of Chase because he computed the efficiency under the assumption of a fixed number of pairs with nonmatching exposure levels in the matched case-control sample. This assumption does not apply to the genetic setting, because the number of heterozygous parents is a random quantity that cannot be controlled.

The asymptotic relative efficiency only addresses a small portion of the space of alternatives. Thus, we explore relative efficiency as a function of prevalence K, risk γ, risk allele frequency p, and for power of 80%. Exploring four levels of prevalence K (0.002, 0.02, 0.1, 0.2), the relative efficiency only changes substantially for larger values of K. For $K = 0.2$, the relative efficiency varies between 2.04 and 2.52 (fig. 1). For $K = 0.002$, on the other hand, relative efficiency ranges from 1.46 to 1.54, with the minimum at small and large p and rel-

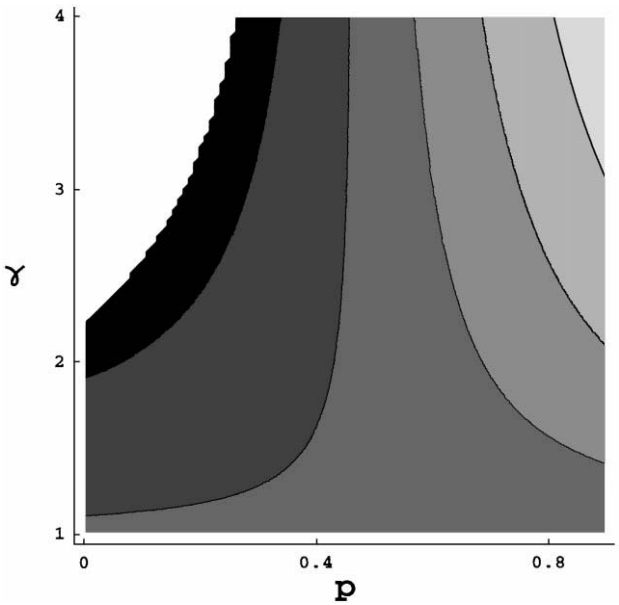

Figure 1 Relative efficiency of TDT versus case control for the multiplicative model and disease prevalence $K = 0.2$. The axes represent the disease-allele frequency p, the relative-risk parameter for the multiplicative model γ, and the ratio of the number of subjects necessary to obtain 80% power, N_{TDT}/N_{CC}. The discontinuity in the surface in the upper left-hand corner represents the region where no genetic models exist.

atively large γ, and the maximum at intermediate values of p and relatively large γ. For $K = 0.02$, the range is, roughly, 1.45–1.6; for $K = 0.1$ the range is, roughly, 1.7–1.95. It is apparent from this analysis that GC tests are considerably more powerful than the TDT test for multiplicative models and no population substructure. Notably, for common diseases, GC sometimes requires only 40% as many subjects to obtain the same power as the TDT test (fig. 1).

Results for General Models

To investigate the asymptotic relative efficiency for models other than the multiplicative model, we first generalize our notation. Let f_0, f_1 and f_2 be the true model penetrances for 0, 1, and 2 liability alleles (labeled A) at the locus, and let $f_j = f_{j-1} + \delta_j$, $j = 1,2$. In the above parameterization, $K = f_0 + (2pq + p^2)\delta_1 + p^2\delta_2$. As given by Suarez et al. (1978), $V_A = 2pq(q\delta_1 + p\delta_2)^2$ is the additive variance and $V_D = p^2q^2(\delta_2 - \delta_1)^2$ is the dominance variance for the locus. The heritability of the locus is given by

$$H = \frac{V_A + V_D}{K(1-K)} = \frac{2pq(q\delta_1 + p\delta_2)^2 + p^2q^2(\delta_2 - \delta_1)^2}{K(1-K)} .$$

The genotypic probabilities for the true model is given in table 3. Taking the limit as $(\delta_1 \to 0, \delta_2 \to 0)$, the asymptotic relative efficiency in the general model matches that given for the multiplicative model in (3); see Appendix A for proof.

Using the methods and parameters given in the previous section, we also have explored additive, dominant, and recessive models for the full parameter space. In each instance, there is no region in which TDT is more efficient than GC (data not shown).

Population Stratification

Clearly, populations are stratified, some more than others. Thus, the power of population-based and family-based designs should be contrasted under more realistic scenarios of population substructure. GC is critical in this instance, because standard tests of independence for the case-control design will yield a false-positive rate higher than the nominal value, inappropriately raising the power of the test. In this section, we evaluate efficiency by simulations, because substructure is too complex to be analytically tractable, except in special cases.

Probability Model for Generating Multilocus Systems

This section lays out a convenient model for multiple liability loci, the *competing-risk* model from reliability theory (Leemis 1995). The advantage of this model is that it is relatively easy to ensure that parameter constraints are met even in the multilocus setting. For this model, a subject becomes affected by the disease if at least one of the loci transmits a disease signal. Let X_l, $l = 1, \ldots, c$ be the binary random variables associated with each locus; $X_l = 1$ means that locus l transmits a disease signal. If at least one of the X_l has value 1 (i.e., transmits the disease signal), then the subject is affected.

Provided $P(X_l = 1)$ is small, the competing-risks model is roughly equivalent to a model that behaves additively over loci, because rarely will more than one locus send a signal. Otherwise, the effect of the loci on liability is subadditive, to account for the intersection of signals from liability loci. Let $f'_{l,j}$ be the penetrance of the genotype with j liability alleles at locus l, under the assumption that all the other liability loci are "turned off." As in the one-locus case, we parameterize as follows: $f'_{l,j} = f'_{l,j-1} + \delta'_{l,j}$, with the restriction $\delta'_{l,j} \geqslant 0$, $j = 1,2$. Provided $P(X_l = 1)$ is small, the resulting $\delta'_{l,j}$ is approximately $\delta_{l,j}$, and so the results are not affected by the use of $f'_{l,j}$ instead of $f_{l,j}$. Notice $f'_{l,j}$ is not the true (i.e., marginal) penetrance when all loci contributing to the disease are present.

Simulations

Using the competing-risk model, we generate $c = 5$ disease loci that contribute equally to the heritability. Define per locus heritability as H_l, $l = 1, \ldots, c$; likewise, let $K_l = 1 - (1 - K)^{\frac{1}{c}}$. The liability allele for the lth locus has frequency in the entire population of $p_l = 1 - q_l \in$

Table 3

Genotype Probabilities

	A Alleles			
Subjects	0	1	2	Total
Case	$\frac{\phi q^2 f_0}{K}$	$\frac{\phi pq f_1}{K}$	$\frac{\phi p^2 f_2}{K}$	ϕ
Control	$\frac{(1-\phi)q^2(1-f_0)}{1-K}$	$\frac{(1-\phi)pq(1-f_1)}{1-K}$	$\frac{(1-\phi)p^2(1-f_2)}{1-K}$	$1-\phi$

Note.—In this notation, the prevalence is K, the frequency of the disease allele A is $p = 1 - q$, the penetrances are $f_j, j = 1, \ldots, 3$, and the fraction of the sample that are cases is ϕ.

Table 4

Median Inflation Factors, λ

Type and No. of Strata	λ at F_{st} .001	.003	.01	.03
Common:				
2	1.04	1.03	1.05	1.09
3	1.04	1.03	1.06	1.08
4	1.03	1.03	1.04	1.08
6	1.03	1.03	1.04	1.08
10	1.03	1.03	1.04	1.07
Disjoint:				
2	1.43	2.16	4.74	11.94
3	1.22	1.60	2.91	6.63
4	1.16	1.41	2.25	4.84
6	1.11	1.31	1.97	3.87
10	1.13	1.26	1.78	3.32

[0.01,0.99], but its frequency varies among subpopulations. Let F_{st} denote Wright's standardized measure of variation among subpopulations. It is assumed that p_l follows a Beta $[\frac{(1-F_{st})}{F_{st}}p_l, \frac{(1-F_{st})}{F_{st}}(1-p_l)]$ distribution.

The following equalities and constraints apply to our model:

$$K_l = f'_{l,0} + (2p_lq_l + p_l^2)\delta_{l,1} + p_l^2\delta_{l,2} \quad (4)$$

$$H_l = \frac{2p_lq_l(p_l\delta_{l,2} + q_l\delta_{l,1})^2 + p_l^2q_l^2(\delta_{l,2} - \delta_{l,1})^2}{K(1-K)} \quad (5)$$

with the restrictions:

$$f'_{l,0} \geq 0,\ 0 \leq \delta_{l,j} \leq 1\ ,\ j = 1,2\ ,\ f'_{l,2} = f'_{l,0} + \delta_{l,1} + \delta_{l,2} \leq 1 \quad (6)$$

To compare the power of GC and TDT, we simulated data under a broad range of conditions with 50 randomly selected genetic models drawn for each condition:

H_l in {0.002,0.01,0.025,0.05,0.075}
K in {0.0005,0.002,0.01,0.02,0.05,0.1,0.1,0.2}
F_{st} in {0.001,0.003,0.01,0.03}
n_{strata} in {2,4,6,8,10} when cases and controls were sampled from different subpopulations and {2,3,4,6,10} when cases and controls were sampled from the same subpopulations.

This range of single-locus heritabilities is reasonable for a complex disorder; for example, in terms of Risch's (1990) λ_S, the risk ratio for siblings, the heritabilities correspond to λ_S from 1.125 to 4.75 for $K = 0.01$ and purely additive effects. For a particular set of conditions, if it was possible to find a genetic model that meets the constraints (given $p_l \in [0.01,0.99]$), then data sets were generated with $c = 5$ candidate genes, $n - c = 45$ null loci, sample size $N = 360$ and $\phi = 0.5$. (For details on data generation, see Appendix B.)

Using the simulated data, we first investigated the size of the GC test to ensure that it was not adversely affected by population stratification. For $\alpha = 0.05$, we found the size of the test was slightly conservative (0.048) by integrating over the entire model space. The size of the test was very close to the nominal value for each value of F_{st}.

Because the true value of the inflation factor λ is difficult to compute analytically and is unknown in practice, we then used the simulations to determine the number of null loci required to yield good estimates of λ. The experiment was conducted at four levels of F_{st} (0.03, 0.01, 0.003, 0.001) (table 4); results for the two lowest levels of substructure were indistinguishable. For each experiment, $\hat{\lambda}$ was estimated as described in "GC Method." For small n, this estimator of λ was biased upward, because $\hat{\lambda}$ was forced to be >1 for all experiments. We fitted a quadratic to the resulting estimates to determine at what value of n a reliable estimate of λ was obtained (fig. 2). The fitted values clearly reach an asymptote by $n = 70$, but the test performed reliably even

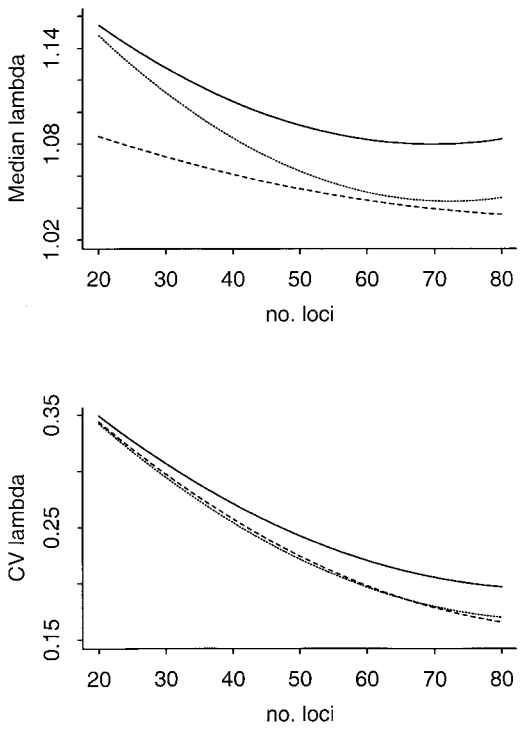

Figure 2 Estimating λ with varying numbers of loci. A smooth curve is fit to the median value of λ (*top panel*) and the coefficient of variation (*bottom panel*) obtained under varying conditions; the solid, dotted, and dashed lines are for F_{st} equal to 0.03, 0.01, and 0.003/0.001 combined, respectively.

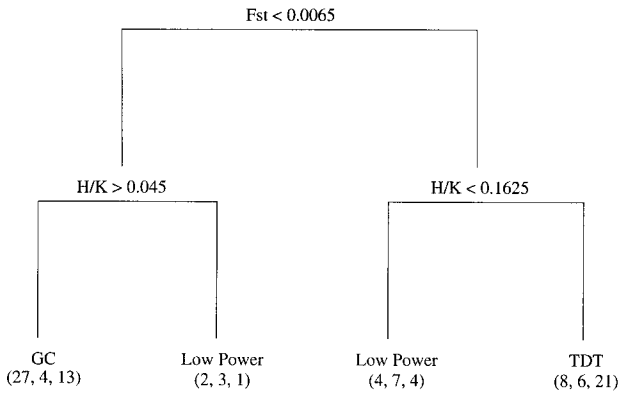

Figure 3 Classification and regression tree comparing the power of TDT and GC when cases and controls were sampled from disjoint subpopulations. Power is compared across 24,000 conditions that varied by heritability (H), F_{st}, number of strata (n_{strata}), prevalence (K) and genetic model. Each leaf indicates which test is most powerful for the majority of models described by the branch. Splits are three-way: if both models have power <20%, then the model is declared to be of "low power"; otherwise, we compare the average size of the GC and TDT tests across the five loci. The trio of numbers at each leaf are the percentage of conditions for which GC was most powerful, TDT was most powerful, or both procedures were of low power, respectively.

when only 20 null loci are used (fig. 2, and unpublished data).

To investigate the power over tens of thousands of model configurations, we approximated power for any particular model. For each configuration of parameters, 50 data sets were generated and analyzed with both the GC and TDT methods. The sample TDT and GC χ^2s were computed for each of the loci. By performing a total of 24,000 experiments, we explored a very large portion of the model space and inevitably also sampled models that occur with high probability. Per-experiment power was computed as the number of tests out of five rejected at the 0.05 level. If a test does not reject for any of the five candidates, then we conclude that the test has power <20%. If both TDT and GC had power <20% for a particular configuration, we declared the configuration as "low power." When there were ties but not low power, we computed the average of the c test statistics and declared the test with the higher average as the one having higher power.

A classification and regression tree (CART) was constructed to interpret these results on power (figs. 3 and 4). The response variable was trivariate: either GC or TDT was more powerful, or the configuration was of low power. CART seeks to find covariates that group models with similar performance. As CART naturally groups similar models, it smooths our approximations to the power of each model.

Figure 3 displays results of a simulation designed to illustrate the most extreme form of population stratification. The population is constructed from n_{strata} distinct subpopulations. All of the cases are drawn from half of the subpopulations and all of the controls from the remaining half of the subpopulations. Out of 24,000 experiments, GC and TDT are more powerful on 41% and 39%, respectively. In the remaining 20% of the experiments, the methods have low power and comparison is not fruitful. The central two branches of the tree describe a region where both methods tend to have low power for $N = 360$. This lack of power is not surprising because $\lambda_S < 1.02$ with modest levels of substructure for one branch and $\lambda_S < 1.08$ with high levels of substructure for the other branch. The outer two branches delineate the interesting region of the model space. It is apparent that GC obtains its best power when F_{st} is <0.01, whereas TDT is advantageous when subpopulations are few and highly differentiated. Not surprisingly the TDT method is most strongly favored when fewer strata ($n_{strata} \leqslant 4$) are present or when there is substantial population substructure, $F_{st} > 0.02$ (unpublished data).

Figure 4 is constructed under more reasonable conditions. Both cases and controls are sampled from the same strata. However, a natural association between strata and cases/controls is induced by the sampling process: those strata with larger p will produce cases more often than strata with smaller p. Hence, we oversample strata associated with larger p, to obtain cases, and undersample the same strata, to obtain the controls; see Appendix B for the algorithm. For these simulations, out of 24,000 experiments, GC and TDT are more powerful on 71.5% and 22%, respectively. In the remaining 6.5% of the experiments, the methods have power <20%. There is only one small portion of the tree (fig. 4) for which the methods are nearly equal in power. It occurs in that portion of the model space for which both methods have little power ($H/K < 0.045$, $\lambda_S < 1.02$). In fact, the same effect is evident throughout the tree. As

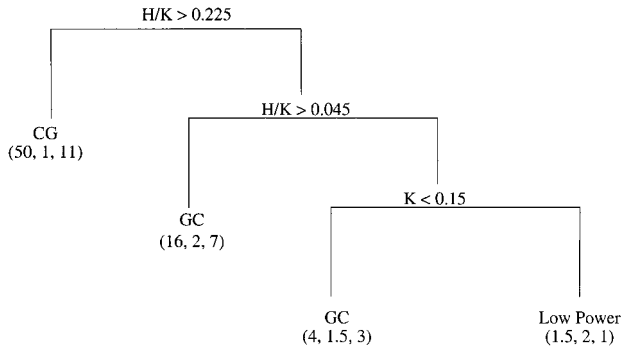

Figure 4 Classification and regression tree comparing the power of TDT and GC when cases and controls were sampled from common subpopulations. See figure 2 for details.

power diminishes from left to right in the tree (fig. 4), the performance of the two methods becomes closer to equality.

Power for GC

To this point, we have considered only the relative power of GC for a case-control sample versus TDT for family trios. The direct power of GC for any particular study can be estimated given certain assumptions about the population. For example, consider planning an association study for bipolar disorder and assume that 400 white cases and controls will be recruited. (For simplicity, even though individuals from other self-identified groups may be recruited, we focus on the sample of whites.) Let $K = 0.01$, an accepted value for bipolar disorder, the disease alleles behave additively, with frequency ranging from 0.005 to 0.05 and the heritability attributable to the locus ranging from 0.5% to 2%. To assess power in this case, we take $F_{st} = 0.003$ (see Discussion) and assume cases and controls are drawn from the same subpopulations.

For these power calculations, we also assume that the polymorphism/mutation tested has a direct impact on liability and that allelic heterogeneity is absent. As discussed elsewhere (e.g., see Slager et al. 2000), these assumptions are optimistic. Under these assumptions, for $\alpha = 0.0001$, GC has excellent power even for the smallest locus-specific heritabilities (table 5).

Notably, suppose the controls were not actually screened for bipolar disorder; instead, these "unselected" controls were simply a random sample from the population (i.e., containing 1% frequency of bipolar cases). Power is not reduced substantially in this instance (table 5). In fact, for the conditions explored in this report, the power for "selected" versus "unselected" controls is very similar unless K is substantial, namely $K > 0.10$ (unpublished data). A program available from the authors can be used to compute power under a variety of conditions.

Table 5

Power to Detect a Locus Accounting for .5%–2% of the Heritability of a Trait by Means of GC and a Case-Control Design of 400 Cases and 400 Controls

Heritability for Locus and Frequency of Liability Allele	Power with Use of Controls	
	Selected	Unselected
.005:		
.005	.700	.727
.01	.895	.892
.02	.976	.976
.05	.997	.996
.01:		
.005	.939	.939
.01	.994	.992
.02	1.00	1.00
.05	1.00	1.00
.015:		
.005	.982	.984
.01	.999	.999
.02	1.00	1.00
.05	1.00	1.00
.02:		
.005	.993	.995
.01	1.00	.999
.02	1.00	1.00
.05	1.00	1.00

Note.—In this power analysis, which mimics evaluation of a candidate gene acting additively to affect liability to bipolar disorder, two kinds of controls were evaluated: selected controls, which were screened to exclude individuals affected by bipolar disorder, and unselected controls, which were not screened. Other attributes of the sample are described in the text.

Discussion

Perhaps the most noteworthy methodological development during the past decade, in terms of genetic association analysis, has been the family-based design and complementary significance tests. These tests are remarkably robust to population substructure and can have excellent power to detect polymorphisms of relatively small impact on liability. Herein, we have examined one such family design, consisting of two parents and an affected child (trios), comparing its efficiency to that of the case-control design. The arithmetic of a trio required to assess the Mendelian transmission of two alleles makes family-based tests likely to be inefficient, relative to case-control tests. In this study, we investigate efficiency in detail in two settings: when population substructure is absent and when it is present.

For stratified populations, it makes little sense to compare the power of standard tests of independence for the case-control and family-based designs, because the former has an inflated false-positive rate. Tests of case-control data must be penalized for this false increase in power. Devlin and Roeder (1999) propose a method, GC, to determine this penalty—that is, the degree to which χ^2 statistics are inflated due to population substructure and a related phenomena, cryptic relatedness—and to correct for it. In its purest form, GC requires evaluation of loci unrelated to disease for both cases and controls (null loci). The degree to which tests of allelic independence at null loci deviate from the standard χ^2 distribution, measured by λ, determines the required degree of correction.

When population substructure is absent, our results show that GC is substantially more efficient than TDT (fig. 1). As the effect on liability for any locus approaches zero, the efficiency approaches a constant, regardless of the frequency of the liability allele. The constant depends on the prevalence K of the disease,

approaching three halves as K also approaches zero, but it becomes more extreme as K increases.

Even when population substructure is present, our simulations agree with the theoretical calculations in Devlin and Roeder (1999) and suggest that, for a well-designed study, the effect of population substructure is not sizable (table 4). Only when the cases and controls are sampled from nearly distinct strata, and F_{st} is large, does the inflation factor rise much above 1.0. Provided cases and controls are not sampled from disjoint strata, half of the values of λ are 1.1 or less, even in the extreme case. To compute the impact of such an inflation, recall that the GC test is not significant unless the trend test statistic (1) is greater than $\lambda\ \chi^2_1$. For $\lambda = 1.1$, the net effect is to require significance at about the $\alpha/2$ level for α in the range (0.01,0.00001). Thus it is not surprising the power for GC is greater than TDT under our simulated conditions, which we believe are similar to those encountered in practice.

The GC approach is better suited to some populations than it is for others. Because the African American population results, in part, from white/black admixture, family-based methods may be the ideal method for association analysis (Ewens and Spielman 1995). Alternatively, white populations have very low levels of F_{st} and thus are ideal for GC. Morton (1992) reports values on the order of 0.0006–0.002, with a European average of 0.001. Using a sample of 122 classical genetics markers, Devlin and Roeder (1999) estimate $F_{st} = 0.0006$ across major European populations, with a standard deviation of 0.0012. The National Health Examination Survey (1980) has examined substructure within and across regions of the United States (Northwest, Northeast, South, and West) using a stratified, random sample of two major ethnic groups, whites and blacks. The variation among regions is quite small for eight loci, with the estimated $F_{st} = 0.0002$ for whites (Chakraborty 1993). As expected, the estimate is larger for blacks, $F_{st} = 0.006$ (Chakraborty 1993).

The success of the GC approach rests, in part, on limited variability of λ (and, thus, F_{st}) across the genome. For neutral alleles and equal mutation rates across loci, theory suggests F_{st} is constant for alleles both within and between loci regardless of their frequencies (e.g., see Wright 1969). In practice, F_{st} does vary, but the variation depends on the populations examined. It also can vary as a function of allele frequencies, but, again, the relationship depends on the populations examined. For example, fitting a line to the data on F_{st} and allele frequencies (p) for worldwide populations reported in Cavalli-Sforza et al. (1994), we find a significant relationship (fig. 5). However, the greatest change occurs for $p < 0.1$, and F_{st} appears not to vary with p thereafter. Worldwide populations show the most extreme substructure (Devlin et al. 1993). It is reasonable, therefore,

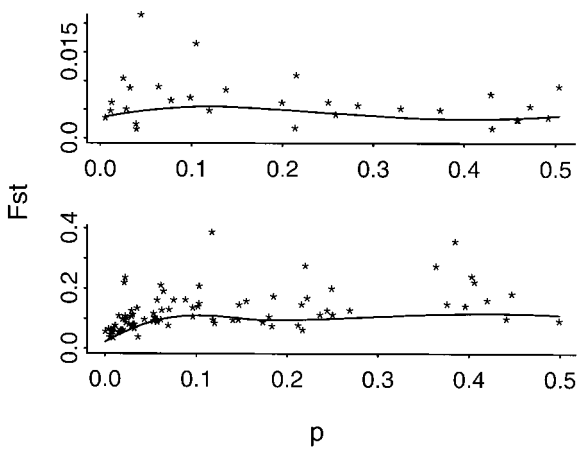

Figure 5 F_{st} versus mean allele frequency (p) for the worldwide population (*bottom*) and selected European populations (*top*). The fitted curves were obtained via a smoothed local regression by use of the S-plus loess function with the default parameters (Chambers et al. 1991). *Bottom,* data were taken from table 2.12.1 of Cavalli-Sforza et al. (1994). We deleted heavy-chain immunoglobulin alleles (IGHG1G3 alleles) and the Duffy blood group (FY*O), which show extreme variability among populations and appear to be under selection. *Top,* F_{st} and mean p were calculated from Roychoudhury and Nei (1988), by use of all loci with at least 7 populations reported (out of 11 cited in the text). For biallelic loci, the smallest p was used; for multiallelic loci, the first $P < .5$ was used. F_st was calculated by use of the shortcut formula by Weir (1990, p. 155).

to examine how F_{st} varies in less substructured populations. For this evaluation we choose a subset of European populations that roughly represent the ethnic composition of white Americans. We derived allele frequencies for 11 populations (Ashkenazi Jews, French, Germans, Greeks, Italians, Dutch, Norwegians, Spanish, UK-English, and western Russians) for loci reported in Roychoudhury and Nei (1988). Unlike in the worldwide view, there is no apparent relationship between F_{st} and p (fig. 5). Moreover, the variability of F_{st} is small.

Several conclusions follow from these observations for GC designs. First, experiments should be designed to minimize population substructure. Second, when highly substructured populations are analyzed, every effort should be made to account for the substructure in the statistical analysis, as this will increase the power of the test. Third, if highly substructured populations must be analyzed and the nature of the substructure is unknown, then one should select null loci having relatively common alleles, making GC conservative. In this instance, the relative efficiency of GC versus TDT will change, and it is likely TDT will be more efficient (fig. 3).

It is important to remember that, on average, cryptic relatedness also tends to inflate test statistics in a population-based study (Devlin and Roeder 1999). In small,

isolated populations, notable levels of kinship can be present amongst what appear to be unrelated individuals. GC adjusts for this inflation automatically, but at a substantial cost in power, because λ can become quite large. In such populations, family-based methods are likely to be more powerful.

When N family trios are recruited, GC typically requires recruitment of $<3N$ unrelated individuals to achieve equal power, at least for a well-designed study, with the fraction determined by the prevalence of the disease and the degree of population substructure. When substructure is small, $<\frac{3}{2}N$ individuals are required. Yet relative power is only one part of the equation. The other is the relative costs of the competing study designs, which involve recruitment and genotyping. In terms of genotyping costs, GC is more expensive than family-based designs, because of the cost of genotyping null loci. Even at current prices, however, the cost of genotyping 20–60 null loci for cases and controls is not prohibitive. By contrast, recruitment costs are usually lower for the case-control study, relative to a similarly powered family-based study—sometimes substantially so. Thus, on balance, GC implemented by a case-control study should be at least competitive with, and probably less expensive than, family-based designs.

As envisioned by Devlin and Roeder (1999), GC uses the genome itself to determine population history and to produce appropriate corrections for population-based association tests. GC does not depend on the case-control design. In fact, it has its roots in the linkage analyses of Puffenberger et al. (1994) and Houwen et al. (1994). When interest lies in the association of selected SNPs and a quantitative trait, GC methods can be implemented with only a population-based sample. In addition, it is possible to evaluate association by contrasting regions of the genome suspected of harboring liability loci to regions thought not to contain such loci. Devlin et al. (in press), for example, construct such tests for excess haplotype-sharing using affected individuals only. Ideally, such haplotype analyses can use the spatial correlations among dense markers to efficiently detect association. Although GC will not be ideal for all settings, we believe it is an important tool for the genetic dissection of complex traits.

Acknowledgments

This research was supported by National Institute of Health grants MH57881 and MH56193 and National Science Foundation grant DMS-9803433. We are grateful to Daniel Weeks for editorial comments.

Appendix A

Asymptotic Relative Efficiency: The General Case

Using the notation and equations given in our section "No Population Stratification," and letting $w = (q\delta_1 + p\delta_2)$, we obtain the following equations for the expected sample size:

$$N_{CC} = \frac{(\sqrt{U}z_\alpha + \sqrt{V}z_\beta)^2 K^2 (1-K)^2}{2qp(1-\phi)\phi w^2} ,$$

where

$$U = \frac{[K^2 - K(1+pw) + pw\phi][K^2 - K(1-qw) - qw\phi]}{(1-K)^2 K^2 (1-\phi)\phi}$$

and

$$V = \frac{[K-1+qw][K-1-pw]}{(1-K)^2(1-\phi)} + \frac{[K-pw][K+qw]}{K^2\phi} .$$

Using Knapp's (1999) formula for singletons, we require $\mu = (e_1 - e_2)/(\sqrt{e_1 + e_2})$ where $e_1 - e_2 = (2pqw)/K$ and $e_1 + e_2 = (2pq[2K + (1-2p)w])/K$. μ simplifies to $\sqrt{2pqw}/\sqrt{K[2K + (1-2p)w]}$. Notice that

$$\lim_{(\delta_1,\delta_2)\to(0,0)} \frac{\mu}{w} = \frac{\sqrt{pq}}{K} .$$

Define N_{TDT} as the total number of subjects (not families) required:

$$N_{TDT} = \frac{3(z_\alpha \sigma_{H_0} + z_\beta \sigma_{H_a})^2}{\mu^2} ,$$

where $\sigma_{H_0} = 1$ and

$$\lim_{(\delta_1,\delta_2)\to(0,0)} \sigma_{H_a} = \sigma_{H_0} = 1 .$$

Both N_{CC} and N_{TDT} have w^2 in the denominator. Set $0 < \phi < 1$ and $0 < p < 1$. The fact that the nonzero limits of the following continuous functions exist at $(\delta_1,\delta_2) = (0,0)$,

$$\lim_{(\delta_1,\delta_2)\to(0,0)} N_{CC} w^2 = \frac{(1-K)^2 K^2 (z_\alpha + z_\beta)^2}{2pq\phi(1-\phi)} ,$$

$$\lim_{(\delta_1,\delta_2)\to(0,0)} N_{TDT} w^2 = \frac{3(z_\alpha + z_\beta \lim\limits_{(\delta_1,\delta_2)\to(0,0)} \sigma_{H_a})^2}{\lim\limits_{(\delta_1,\delta_2)\to(0,0)} \frac{\mu^2}{w^2}}$$

$$= \frac{3(z_\alpha + z_\beta)^2}{\frac{pq}{K^2}} = \frac{6K^2(z_\alpha + z_\beta)^2}{2pq}$$

implies that

$$\lim_{(\delta_1,\delta_2)\to(0,0)} \frac{N_{TDT}}{N_{CC}} = \lim_{(\delta_1,\delta_2)\to(0,0)} \frac{N_{TDT} w^2}{N_{CC} w^2} = \frac{6\phi(1-\phi)}{(1-K)^2} \quad .$$

If the number of cases equals the number of controls ($\phi = \frac{1}{2}$), then the limit becomes

$$\lim_{(\delta_1,\delta_2)\to(0,0)} \frac{N_{TDT}}{N_{CC}} = \frac{3}{2(1-K)^2} \quad .$$

Appendix B

Outline for Simulations

For simulation data, five liability or candidate loci were produced ($c = 5$), each with heritability equal to H. Thus, if each locus had $H = 0.05$, the total heritability for the five loci was 25%. To hasten computation, while still meeting model constraints (as described above in "Population Stratification: Simulations"), the realized heritability for each locus was allowed to vary by <1% of its target value. Monotonicity of the penetrances was strict, with $f_0' < f_1' < f_2'$; otherwise the choice of genetic model was unrestricted as long as it met parameter constraints.

For a given run, the following parameters (H, K, F_{st}, n_{strata}, N, c, n) were fixed and the entire process was repeated until $n_{sim} = 50$ simulation results were obtained. Only a subset of the potential model configurations resulted in models that met the constraints. As a prelude to a simulation for a given set of parameters, the following process was performed:

Generate random $p \in (0.01,0.99)$.
Solve equalities for $\delta_j, j = 1,2$ (4) and (5).
Check all constraints (6); if met, declare model feasible.
If constraints are unmet after 200,000 attempts, assume no model exists for (H,K).

If the model was declared feasible, then the following steps were undertaken. For each locus, we randomly select models until we find one that meets the constraints.

LOOP over diseased loci $l = 1, \ldots, c$ to produce c liability loci that meet constraints.
Generate p_l for each locus.
Generate random $\delta_{l,j}, j = 1,2$.
Check constraints; if met, keep $p_l, \delta_{l,1}, \delta_{l,2}$.
Generate p_{ls}, $s = 1, n_{\text{strata}}$, given F_{st} and p_l.
LOOP over other loci $l = c + 1, \ldots, n$, to produce $n - c$ null loci. Generate p_{ls}, $s = 1, \ldots, n_{\text{strata}}$, given F_{st} and p_l.

At this point, each of the $l = 1, \ldots, n$ loci is associated with a mean allele frequency p_l and the n_{strata} subpopulations have allele frequencies p_{sl} and their variation about p_l is scaled by F_{st}. Each of the $l = 1,c$ liability loci is associated with a genetic model (δ_{l1}, δ_{l2}) that satisfies the constraints. From these populations, we now randomly draw nuclear families with one child, until we obtain N subjects who satisfy the design constraints for either TDT or GC.

LOOP from 1 to M to produce N subjects (for TDT $M = N/3$ and all children must be affected; for GC, run this twice to get $N/2$ affected and $N/2$ unaffected children).
DO UNTIL family has either an affected or an unaffected child, depending on design setting.
Draw a subpopulation at random.
Draw parents at random.
Obtain child.
Output the appropriate family data.
Compute the sample TDT / GC χ^2 for each of c loci; compute number of tests rejected at the 0.05 level, and the power of the test. If a test does not reject for any of the c candidate loci, then we conclude that the test has power <20%. Record the average of the c test statistics.

References

Armitage P (1955) Tests for linear trends in proportions and frequencies. Biometrics 11:375–386

Barnett V, Lewis T (1978) Outliers in statistical data. Wiley, New York

Chase GR (1968) On the efficiency of matched pairs in Bernoulli trials. Biometrika 55:365–369

Chakraborty R (1993) Analysis of genetic structure of populations: meaning, methods, and implications. In: Majumder PP (ed) Human population genetics, Plenum Press, NY, pp 189–206

Cavalli-Sforza L, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ

Chambers JM, Hastie TJ (1991) Statistical models in S. Chapman and Hall, London

Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L (1998) New goals for the U.S. Human Genome Project: 1998–2003. Science 282:682–689

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Devlin B, Roeder K, Wasserman L. Genomic control for association studies: a semi-parametric test to detect excess haplotype-sharing. Biostatistics (in press)

Devlin B, Risch N, Roeder K (1993) The statistical evaluation of DNA fingerprinting: critique of the NRC report. Science 259:748–749, 837

Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision and admixture. Am J Hum Genet 57:455–464

Houwen RHJ, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, Sandkuijl LA, Freimer NB (1994) Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. Nature Genet 8:380–386

Knapp M (1999) A note on power approximations for the transmission/disequilibrium test. Am J Hum Genet 64: 1177–1185

Leemis LM (1995) Reliability: probabilistic models and statistical methods. Prentice Hall, Englewood Cliffs, NJ

Li CC (1972) Population subdivision with respect to multiple alleles. Ann Hum Genet 33:23–29

Morton NE (1992) Genetic structure of forensic populations. Proc Natl Acad Sci USA 89:2556–2560

Morton NE, Collins A (1998) Tests and estimates of allelic association in complex inheritance. Proc Natl Acad Sci USA 95:11389–11393

National Health Examination Survey (1980) Selected genetic markers of blood and secretions. U.S. Dept. of Health, Education and Welfare, Publ. No. (PHS) 80-1664, Washington, DC

Roychoudhury AK and Nei M (1988) Human polymorphic genes. Oxford University Press, NY

Puffenberger EG, Kauffman ER, Bolk S, Matise TC, Washington SS, Angrist M, Weissenback J, et al (1994) Identity by descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. Hum Mol Genet 3:1217–1225

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222–228

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Risch N, Teng J (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human disease. I. DNA pooling. Genome Res 8:1273–1288

Rice JA (1988) Mathematical statistics and data analysis. Wadsworth & Brooks/Cole, Pacific Grove, CA

Sasieni PD (1997) From genotypes to genes: doubling the sample size. Biometrics 53:1253–1261

Slager S, Huang J, Vieland VJ (2000) Effect of allelic heterogeneity on the power of the transmission disequilibrium test. Genet Epidemiol 18:143–156

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

Suarez B, Rice J, Reich T (1978) The generalized sib pair IBD distribution: its use in detecting linkage. Ann Hum Genet 42:87–94

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, et al (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082

Weir B (1990) Genetic data analysis. Sinauer, Sunderland, MA

Wright S (1969) Evolution and the genetics of populations. Vol 2: The theory of gene frequencies. University of Chicago Press, Chicago