European Journal of Human Genetics (2010) 18, 1032–1038

www.nature.com/ejhg

ARTICLE

The power of the Transmission Disequilibrium Test in the presence of population stratification

Ronnie Sebro*[,1,2] and John J Rogus[3]

The Transmission Disequilibrium Test (TDT) is a family-based test for association based on the rate of transmission of alleles from heterozygous parents to affected offspring, and has gained popularity as this test preserves the Type I error rate. Population stratification results in a decreased number of heterozygous parents compared to that expected assuming Hardy–Weinberg Equilibrium (*Wahlund Effect*). We show that population stratification changes the relative proportion of the informative mating types. The decrease in the number of heterozygous parents and the change in the relative proportion of the informative mating types result in significant changes to the sample sizes required to achieve the power desired. We show examples of the changes in sample sizes, and provide an easy method for estimating TDT sample sizes in the presence of population stratification. This method potentially aids in reducing the number of false-negative association studies.
European Journal of Human Genetics (2010) **18,** 1032–1038; doi:10.1038/ejhg.2010.57; published online 5 May 2010

Keywords: Transmission Disequilibrium Test; population stratification; power

INTRODUCTION

Case–control analysis using singletons has been shown to be generally more powerful than any other design per genotyped individual for detecting genes associated with disease.[1] However, this design is vulnerable to excessive false-positive findings in the presence of population stratification.[2–7] If a study population comprises two subpopulations, and the disease prevalence is greater in one subpopulation compared to the other, then cases will typically be over-sampled from the subpopulation with higher disease prevalence. Spurious associations will be observed between the disease and any genetic marker where the allele frequencies differ in both subpopulations.

The Transmission Disequilibrium Test (TDT) proposed by Spielman *et al*[8] compares the rate of transmission of each allele from a heterozygous parent to an affected offspring. The TDT maintains the desired Type I error rate in the presence of population stratification, as non-transmitted parental alleles from heterozygous parents serve, in effect, as the control population. Therefore, the power of the TDT is heavily dependent on the number of heterozygotes (informative parents), who may or may not transmit the allele of interest. It is well appreciated that population stratification results in a loss of heterozygosity compared to that expected assuming Hardy–Weinberg Equilibrium (HWE). This is known as the *Wahlund Effect*, which results in less informative parents for the TDT. However, no guidelines currently exist for sample size calculations for the TDT in the presence of population stratification.

In this paper, we lay out a method of estimating the two parent genotype patterns (mating types) seen in the presence of population stratification using the parental allele frequencies and Wright's coefficient of inbreeding *F*. The estimates of the mating types are then used for power calculations using the method provided by Knapp.[9] We then compare the sample sizes assumed using HWE to those calculated using our new method and show how these differences arise.

METHODS

Notation and terminology

Population stratification occurs when a population comprises two or more subpopulations, where there is random mating and HWE within subpopulations, but no mating between subpopulations. Assume that there are G separate subpopulations, where G, as well as the actual members of each subpopulation are unknown and let w_i be the proportion of the stratified population represented by subpopulation i. Consider a single biallelic marker or single nucleotide polymorphism (SNP) with two alleles, A and B, respectively. If p_i is the frequency of allele A in subpopulation i, then $\bar{p} = \sum_{i=1}^{G} w_i p_i$ is the allele A frequency (averaged over all subpopulations) in the overall population. Let $q_i = 1 - p_i$ be the frequency of the B allele in subpopulation i, so that the average allele frequency of allele B in the overall population is $1 - \bar{p} = \bar{q}$. The variance of the A allele frequency between subpopulations is defined as $\mathrm{Var}(p_i) = \sum_{i=1}^{G} w(p_i - \bar{p})^2$. As we assume random mating and HWE within subpopulations, the frequencies of the AA, AB and BB genotypes in subpopulation i are p_i^2, $2p_iq_i$ and q_i^2, respectively. Let the proportion of individuals in the overall stratified population with genotypes AA, AB and BB be μ_{AA}, μ_{AB} and μ_{BB}, respectively so that $\mu_{AA} = \sum_{i=1}^{G} w_i p_i^2$, $\mu_{AB} = \sum_{i=1}^{G} 2w_i p_i q_i$, $\mu_{BB} = \sum_{i=1}^{G} w_i q_i^2$.

The mating type is defined as the two-parent genotype combination. We assume symmetry between the mating types (ie AA×AB=AB×AA, etc.), so that instead of nine unique mating types, we have only six mating types. Let $P(\mathrm{AA}\times\mathrm{AA}) = m_1 = \sum_{i=1}^{G} w_i p_i^4$, $P(\mathrm{AA}\times\mathrm{AB}) = m_2 = \sum_{i=1}^{G} 4w_i p_i^3 q_i$, $P(\mathrm{AA}\times\mathrm{BB}) = m_3 = \sum_{i=1}^{G} 2w_i p_i^2 q_i^2$, $P(\mathrm{AB}\times\mathrm{AB}) = m_4 = \sum_{i=1}^{G} 4w_i p_i^2 q_i^2$, $P(\mathrm{AB}\times\mathrm{BB}) = m_5 = \sum_{i=1}^{G} 4w_i p_i q_i^3$ and $P(\mathrm{BB}\times\mathrm{BB}) = m_6 = \sum_{i=1}^{G} w_i q_i^4$.

We define ψ_2 and ψ_1 to be the respective genotypic relative risks (GRR) for the risk allele (allele A) homozygotes and heterozygotes compared to the BB

[1]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA; [2]Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, USA; [3]Research Division, Section on Genetics and Epidemiology, Joslin Diabetes Center, Boston, MA, USA
*Correspondence: Dr R Sebro, Institute for Human Genetics, University of California, San Francisco, 513 Parnassus Avenue, Suite S965, Box 0794, San Francisco, CA 94143-0794, USA. Tel: +1 415 476 1127; Fax: +1 415 476 2956; Email: rsebro@radmail.ucsf.edu
Received 22 December 2009; revised 23 March 2010; accepted 25 March 2010; published online 5 May 2010

homozygote. The GRR, ψ_i is defined as the ratio of the probability of disease in subjects with i risk alleles compared to subjects with 0 risk alleles. We assume that there is no segregation distortion and that the offspring genotype determines the offspring's disease risk. We also assume that the GRRs are the same in each subpopulation (absence of risk heterogeneity). Therefore, the difference in disease prevalence between subpopulations is only because of differences in the risk allele frequency between the subpopulations, that is the risk conferred by a given allele does not vary by subpopulation. For the sample size calculations done in this paper, the following modes of inheritance (MOI) were considered for comparison with Knapp: multiplicative model ($\psi_1=\gamma$ and $\psi_2=\gamma^2$), additive model ($\psi_1=\gamma$ and $\psi_2=2\gamma$), dominant model ($\psi_1=\psi_2=\gamma$) and the recessive model ($\psi_1=1$ and $\psi_2=\gamma$), where $\gamma \geq 1$ for all models.[9] We note that the multiplicative and additive MOI definitions were chosen to correspond to those defined previously by Camp.[10]

Population stratification results in a decrease in heterozygosity compared to that expected assuming HWE

If a population is divided into distinct subpopulations with random mating within each subpopulation, the decrease in the proportion of heterozygotes compared to that occurring under random mating of the whole population is proportional to the variance of the allele frequency between subpopulations. This phenomenon is known as the Wahlund Effect.[11] Wright's coefficient of inbreeding, F, is commonly used to describe population stratification, where $F = \frac{\mathrm{Var}(p_i)}{\bar{p}(1-\bar{p})}$, and the variance of the allele frequency between subpopulations, $\mathrm{Var}(p_i) = \sum_{i=1}^{G} w_i(p_i - \bar{p})^2$. Population stratification results in a loss of heterozygosity that is exactly 100 $(1-F)$% of that expected assuming HWE.

F values expected in practice

It is difficult to know *a priori* how much population stratification may exist in practice. Cavalli-Sforza[12] estimated that the worldwide F values averaged over all genes in the genome is approximately 0.139 ± 0.010. A study by Akey *et al*[13] analyzed 26 530 SNPs in three populations (African-American, East Asian and European-American). The F value for each locus was calculated and the distribution of these F values was noted. Of the 25 549 autosomal SNPs genotyped, 29.8% were common in all three populations, 26.8% were common in only two populations, 22.9% were unique to only one population and 5.1% were uncommon. Here, an SNP was defined as common if the minor allele frequency was greater than 20%. The average F value obtained in the coding, intronic and non-coding regions were 0.107, 0.118 and 0.123, respectively. The distribution of F values was skewed to the right, which suggested that F values in the range 0–0.15 are feasible values that should be considered in practice. We note that F values calculated by Akey *et al*[13] assume that equal proportions of each subpopulation are present in the overall stratified population, whereas we allow for differing proportions of each subpopulation in the overall stratified population. Akey's approach is best suited for calculating the genetic distance between subpopulations, whereas our approach is geared toward detecting the change in the genotype frequencies in the presence of stratification.

The effect of population stratification on mating type frequencies

Let Δ_i be the difference between the allele A frequency in subpopulation i and the allele A frequency in the entire population, where $\Delta_i = p_i - \bar{p}$, for $i=1, 2, \ldots G$. Furthermore, define the central moments of the allele frequency distribution, $\vartheta_j = \sum_{i=1}^{G} w_i(\Delta_i)^j$ for $j=1, 2, 3, 4$. Therefore,

$$\vartheta_1 = \sum_{i=1}^{G} w_i(\Delta_i) = 0 \tag{1}$$

$$\vartheta_2 = \sum_{i=1}^{G} w_i(\Delta_i)^2 = \mathrm{Var}(p_i) = F\bar{p}(1-\bar{p}) \tag{2}$$

$$\vartheta_3 = \sum_{i=1}^{G} w_i(\Delta_i)^3 \tag{3}$$

$$\vartheta_4 = \sum_{i=1}^{G} w_i(\Delta_i)^4 \tag{4}$$

Then,

$$\sum_{i=1}^{G} w_i p_i^3 = \sum_{i=1}^{G} w_i(\bar{p}+\Delta_i)^3 = \sum_{i=1}^{G} w_i \left[\sum_{j=0}^{3} \binom{3}{j} \bar{p}^{3-j}(\Delta_i)^j \right] \tag{5}$$

This yields

$$\sum_{i=1}^{G} w_i p_i^3 = \bar{p}^3 + 3\bar{p}^2\bar{q}F + \vartheta_3 \tag{6}$$

Similarly,

$$\sum_{i=1}^{G} w_i p_i^4 = \sum_{i=1}^{G} w_i(\bar{p}+\Delta_i)^4 = \sum_{i=1}^{G} w_i \left[\sum_{j=0}^{4} \binom{4}{j} \bar{p}^{4-j}(\Delta_i)^j \right] \tag{7}$$

This yields

$$\sum_{i=1}^{G} w_i p_i^4 = \bar{p}^4 + 6\bar{p}^3\bar{q}F + 4\bar{p}\vartheta_3 + \vartheta_4 \tag{8}$$

All of the six mating types can now be re-written in terms of the average risk allele frequency $\bar{p}$, Wright's coefficient of inbreeding F, ϑ_3 and ϑ_4 as shown in Table 1.

Change in the relative proportion of informative families because of population stratification

Three of the six mating types are informative for the TDT because they contain at least one heterozygous parent – AA×AB, AB×AB and AB×BB

Table 1 Mating type frequencies in the presence of population stratification parameterized in terms of the central moments of the allele frequency distribution

	AA	*AB*	*BB*
AA	$\bar{p}^4+6\bar{p}^3\bar{q}F+4\bar{p}\vartheta_3+\vartheta_4$	$2\bar{p}^3\bar{q}+6\bar{p}^2\bar{q}(1-2\bar{p})F+2(1-4\bar{p})\vartheta_3-2\vartheta_4$	$\bar{p}^2\bar{q}^2+F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2)+2(2\bar{p}-1)\vartheta_3+\vartheta_4$
AB	$2\bar{p}^3\bar{q}+6\bar{p}^2\bar{q}(1-2\bar{p})F+2(1-4\bar{p})\vartheta_3-2\vartheta_4$	$4\bar{p}^2\bar{q}^2+4F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2)+8(2\bar{p}-1)\vartheta_3+4\vartheta_4$	$2\bar{p}\bar{q}^3+6F\bar{p}\bar{q}(-1+3\bar{p}-2\bar{p}^2)+2(3-4\bar{p})\vartheta_3-2\vartheta_4$
BB	$\bar{p}^2\bar{q}^2+F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2)+2(2\bar{p}-1)\vartheta_3+\vartheta_4$	$2\bar{p}\bar{q}^3+6F\bar{p}\bar{q}(-1+3\bar{p}-2\bar{p}^2)+2(3-4\bar{p})\vartheta_3-2\vartheta_4$	$\bar{q}^4+6F\bar{p}\bar{q}(1-2\bar{p}+\bar{p}^2)+4(\bar{p}-1)\vartheta_3+\vartheta_4$
Total	$\bar{p}^2+F\bar{p}\bar{q}$	$2\bar{p}\bar{q}(1-F)$	$\bar{q}^2+F\bar{p}\bar{q}$

$\bar{p}$ is average risk allele frequency in the stratified population, $\bar{q}=1-\bar{p}$.
p_i is risk allele frequency in subpopulation i, $q_i=1-p_i$.
μ_{AA} is expected proportion of AA homozygotes in stratified population.
μ_{AB} is expected proportion of AB heterozygotes in stratified population.
μ_{BB} is expected proportion of BB homozygotes in stratified population.
G is number of subpopulations present in the stratified population.
w_i is proportion of subjects sampled from subpopulation i.
F is Wright's coefficient of inbreeding.
$\vartheta_3 = \sum_{i=1}^{G} w_i(p_i - \bar{p})^3$, the third central moment of the risk allele frequency distribution.
$\vartheta_4 = \sum_{i=1}^{G} w_i(p_i - \bar{p})^4$, the fourth central moment of the risk allele frequency distribution.

(m_2, m_4 and m_5). However, as shown above, Wright's coefficient of inbreeding, F, is insufficient to appropriately calculate all mating type frequencies without ϑ_3 and ϑ_4. Under HWE the relative proportions of the informative mating types m_2:m_4:m_5 are clearly $\bar{p}^2$:$\bar{p}$ $(1-\bar{p})$:$(1-\bar{p})^2$; however, these proportions no longer hold in the presence of population stratification. To illustrate the changes in the relative proportions of the mating types, we consider a study population comprising two equal-sized subpopulations, where F=0 and 0.10, ϑ_3=0 and ϑ_4 is allowed to vary for a multiplicative MOI where γ=2 (only $\bar{p}$, F and ϑ_3 are required to completely describe a two-subpopulation model). The changes in the relative proportions of the informative mating types AA×AB, AB×AB and AB×BB are shown in Figure 1.

The power of the TDT and sample size calculations in the presence of population stratification

The TDT tests the null hypothesis of no association between a marker and disease in the presence of linkage. Knapp provided a method for reliably calculating the power of the TDT in a homogeneous population assuming HWE for affected child trios (ACTs).[9] We retain his symbols for ease of comparison with our method. At the heart of Knapp's method is the characterization of family types (genotypes of both parents as well as the affected offspring) for the TDT and the calculation of each family type probability (s_1, s_2,…, s_{10}), as shown in Table 2. Seven family types are informative for the TDT (s_1, s_2,…, s_7) and Knapp showed how their

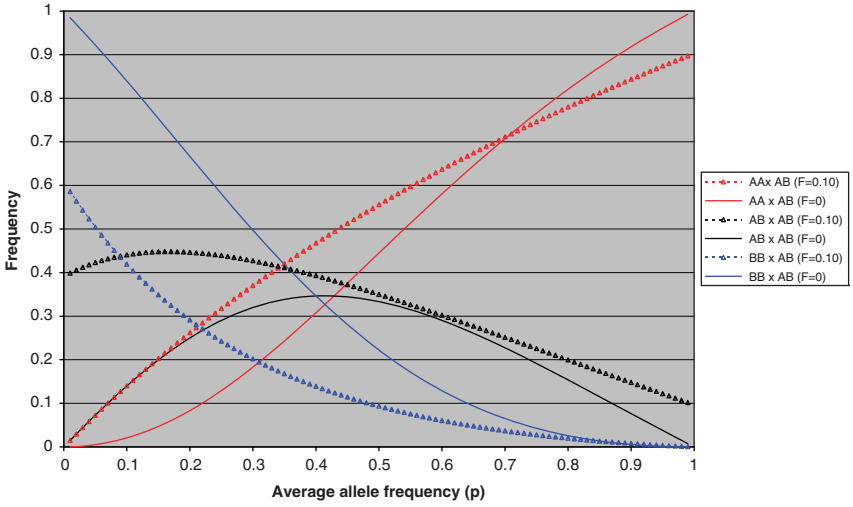

Figure 1 Relative proportion of the informative mating types for a multiplicative MOI where γ=2.

Table 2 Family type probabilities given affected child

j	*Family type* *Parental mating*	*Affected child genotype*	*Probability assuming HWE (s_j)*	*Probability in presence of population stratification s^*_j*	*Estimation in presence of population stratification using only $\bar{p}$ and F $\hat{s}^*_j$*
1	AA×AB	AA	$2\bar{p}^3\bar{q}\psi_2/R$	$\frac{(2\bar{p}^3\bar{q}+6\bar{p}^2\bar{q}(1-2\bar{p})F+2(1-4\bar{p})\vartheta_3-2\vartheta_4)\psi_2}{\bar{R}}$	$\frac{(2\bar{p}^3\bar{q}+6\bar{p}^2\bar{q}(1-2\bar{p})F)\psi_2}{\bar{R}}$
2	AA×AB	AB	$2\bar{p}^3\bar{q}\psi_1/R$	$\frac{(2\bar{p}^3\bar{q}+6\bar{p}^2\bar{q}(1-2\bar{p})F+2(1-4\bar{p})\vartheta_3-2\vartheta_4)\psi_1}{\bar{R}}$	$\frac{(2\bar{p}^3\bar{q}+6\bar{p}^2\bar{q}(1-2\bar{p})F)\psi_1}{\bar{R}}$
3	AB×AB	AA	$\bar{p}^2\bar{q}^2\psi_2/R$	$\frac{(\bar{p}^2\bar{q}^2+F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2)+2(2\bar{p}-1)\vartheta_3+\vartheta_4)\psi_2}{\bar{R}}$	$\frac{(\bar{p}^2\bar{q}^2+F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2))\psi_2}{\bar{R}}$
4	AB×AB	AB	$2\bar{p}^2\bar{q}^2\psi_1/R$	$\frac{(2\bar{p}^2\bar{q}^2+2F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2)+4(2\bar{p}-1)\vartheta_3+\vartheta_4)\psi_1}{\bar{R}}$	$\frac{(2\bar{p}^2\bar{q}^2+2F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2))\psi_1}{\bar{R}}$
5	AB×AB	BB	$\bar{p}^2\bar{q}^2/R$	$\frac{(\bar{p}^2\bar{q}^2+F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2)+2(2\bar{p}-1)\vartheta_3+\vartheta_4)}{\bar{R}}$	$\frac{(\bar{p}^2\bar{q}^2+F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2))}{\bar{R}}$
6	AB×BB	AB	$2\bar{p}\bar{q}^3\psi_1/R$	$\frac{(2\bar{p}\bar{q}^3+6F\bar{p}\bar{q}(-1+3\bar{p}-2\bar{p}^2)+2(3-4\bar{p})\vartheta_3-2\vartheta_4)\psi_1}{\bar{R}}$	$\frac{(2\bar{p}\bar{q}^3+6F\bar{p}\bar{q}(-1+3\bar{p}-2\bar{p}^2))\psi_1}{\bar{R}}$
7	AB×BB	BB	$2\bar{p}\bar{q}^3/R$	$\frac{(2\bar{p}\bar{q}^3+6F\bar{p}\bar{q}(-1+3\bar{p}-2\bar{p}^2)+2(3-4\bar{p})\vartheta_3-2\vartheta_4)}{\bar{R}}$	$\frac{(2\bar{p}\bar{q}^3+6F\bar{p}\bar{q}(-1+3\bar{p}-2\bar{p}^2))}{\bar{R}}$
8	AA×AA	AA	$\bar{p}^4\psi_2/R$	$\frac{(\bar{p}^4+6\bar{p}^3\bar{q}F+4\bar{p}\vartheta_3+\vartheta_4)\psi_2}{\bar{R}}$	$\frac{(\bar{p}^4+6\bar{p}^3\bar{q}F)\psi_2}{\bar{R}}$
9	AA×BB	AB	$2\bar{p}^2\bar{q}^2\psi_1/R$	$\frac{(2\bar{p}^2\bar{q}^2+2F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2)+4(2\bar{p}-1)\vartheta_3+2\vartheta_4)\psi_1}{\bar{R}}$	$\frac{(2\bar{p}^2\bar{q}^2+2F\bar{p}\bar{q}(1-6\bar{p}+6\bar{p}^2))\psi_1}{\bar{R}}$
10	BB×BB	BB	$\bar{q}^4/R$	$\frac{(\bar{q}^4+6F\bar{p}\bar{q}(1-2\bar{p}+\bar{p}^2)+4(\bar{p}-1)\vartheta_3+\vartheta_4)}{\bar{R}}$	$\frac{(\bar{q}^4+6F\bar{p}\bar{q}(1-2\bar{p}+\bar{p}^2))}{\bar{R}}$

$\bar{p}$ is average risk allele frequency in stratified population, $\bar{q}=1-\bar{p}$.
G is number of subpopulations.
w_i is weight with which individuals are sampled from sub-population i.
$R=\psi_2\bar{p}^2+\psi_1 2\bar{p}\bar{q}+\bar{q}^2$.
$\bar{R}=\sum_{i=1}^{G} w_i R_i=\sum_{i=1}^{G} w_i(\Psi_2 p_i^2+2\Psi_1 p_i q_i+q_i^2)=\psi_2\bar{p}^2+2\bar{p}\bar{q}\psi_1+\bar{q}^2+(\psi_2-2\psi_1+1)F\bar{p}\bar{q}$.
$R_i=\psi_2 p_i^2+2\psi_1 p_i q_i+q_i^2$.
ψ_i is genotypic relative risk of i *vs* 0 risk alleles. *Derived from Knapp.[9]
F is Wright's coefficient of inbreeding.
$\vartheta_3=\sum_{i=1}^{G} w_i(p_i-\bar{p})^3$, the third central moment of the risk allele frequency distribution.
$\vartheta_4=\sum_{i=1}^{G} w_i(p_i-\bar{p})^4$, the fourth central moment of the risk allele frequency distribution.

multinomial frequencies could be calculated based only on ψ_1, ψ_2 and $\bar{p}$.[9] However, if population stratification exists, the average frequency of the risk allele A in the stratified population ($\bar{p}$) is not sufficient to calculate the mating type proportions. The true family type proportions (s^*_1, s^*_2,…, s^*_{10}) can be calculated from the correct mating type proportions based on the method described earlier (using $\bar{p}$, F, ϑ_3 and ϑ_4) and are shown in Table 2. However, ϑ_3 and ϑ_4 are almost always unknown, and though there are several published papers about F values expected in practice, there are no published reports of ϑ_3 and ϑ_4 values expected in practice. Yasuda[14] showed that the higher central moments (terms in ϑ_3 and ϑ_4) can be ignored if $\bar{p} > F$ and $1-\bar{p} > F$. Therefore, (s_1, s_2,…, s_7) can be estimated by ($\hat{s}^*_1$, $\hat{s}^*_2$,…, $\hat{s}^*_7$) using $\bar{p}$ and F as shown in Table 2 and these estimates can then be used in power calculations. GRRs are set assuming γ=2.0 and 1.5 for multiplicative, additive, dominant and recessive models. To evaluate our method for estimating sample sizes, we considered a stratified study population comprising two smaller subpopulations. In the first example, 70% of the trios are from subpopulation 1, and 30% of the trios are derived from subpopulation 2. In the second example, the study population comprises equal proportions of trios from both subpopulations. We then calculated the true sample size requirements to achieve 80% power, using a Type I error rate of $\alpha=1\times10^{-7}$ at various values of the average allele frequency, $\bar{p}$ (0.1, 0.3, 0.5 and 0.8), and at various levels of population stratification measured by F (0.01, 0.05 and 0.10). The true sample size estimates were compared to our estimates using only $\bar{p}$ and F, as well as compared to the sample size calculations assuming HWE. These sample size calculations are shown in Tables 3 and 4. A Type I error rate of $\alpha=1\times10^{-7}$ is used for comparison with Knapp,[9] and for comparision with sample size estimates from genomewide association studies.

RESULTS

Population stratification alters the frequency of the mating types when compared to those calculated assuming HWE. The change in the distribution of the mating type frequencies directly changes the family type frequencies (mother–father–offspring genotype combinations). When HWE is assumed for TDT power calculations, the expected value and variance of the test statistic under the alternative hypothesis as well as the variance of the test statistic under the null hypothesis differ from that calculated when accounting for population stratification. The variance of the TDT in the presence of stratification under the null hypothesis could be larger or smaller than that calculated assuming HWE depending on the mating type parameters ($\bar{p}$, F, ϑ_3, ϑ_4). The difference between the variance of the TDT under the alternative hypothesis accounting for population stratification and the variance of the TDT assuming HWE depends on the model (ψ_2, ψ_1) and mating type parameters ($\bar{p}$, F, ϑ_3, ϑ_4). The current methods existing for calculating TDT sample sizes ignore the fact that parental genotypes used in the TDT statistic are no longer in HWE if there is population stratification. Furthermore, these methods ignore the fact that the parents cannot be considered independent as population stratification induces non-random mating, when considered on a whole-population basis.

Knapp showed that the power of the TDT is a function of the genetic model, genotypic risk parameters and the risk allele frequency.[9] Our model for TDT sample size calculations in the presence of population stratification based on estimations of the mating type frequencies using the average allele frequency, $\bar{p}$ and Wright's coefficient of inbreeding F allows sampling from any number of subpopulations with any ascertainment scheme, and is generalizeable to any stratified study population. In addition, we show that the change in the power of the TDT in a stratified population is due to three main factors: (1) the loss in heterozygosity, which is 100 $(1-F)$% of that predicted assuming HWE; (2) the change in the distribution of mating type frequencies; and (3) the change in the relative proportion

Table 3 Sample size necessary to gain 80% power in TDT with singletons ($\alpha=10^{-7}$, γ=2), comparing Knapp's first approximation assuming HWE to the sample size estimates assuming population stratification

				Multiplicative MOI			*Additive MOI*			*Recessive MOI*			*Dominant MOI*		
$\bar{p}$	F	ϑ_3	ϑ_4	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
0.1	0.01	2.3×10^{-5}	1.4×10^{-6}	689	692 (100.4%)	692	689	692 (100.4%)	692	45071	39168 (86.9%)	39354	949	978 (103.1%)	977
		0	8.1×10^{-7}			692			692			39168			978
0.1	0.05	2.6×10^{-4}	3.5×10^{-5}	689	702 (101.9%)	704	689	702 (101.9%)	702	45071	24102 (53.5%)	25141	949	1110 (117.0%)	1101
		0	2.0×10^{-5}			702			732			24102			1110
0.1	0.10	7.5×10^{-4}	1.4×10^{-4}	689	714 (103.6%)	720	689	714 (103.6%)	720	45071	14797 (32.8%)	16273	949	1324 (139.5%)	1292
		0	8.1×10^{-5}			714			714			14798			1324
0.3	0.01	8.4×10^{-5}	7.8×10^{-6}	349	352 (100.9%)	352	349	352 (100.9%)	352	2546	2513 (98.7%)	2519	907	927 (102.2%)	925
		0	4.4×10^{-6}			352			352			2513			927
0.3	0.05	9.4×10^{-4}	1.9×10^{-4}	349	364 (104.3%)	365	349	364 (104.3%)	365	2546	2383 (93.6%)	2449	907	1015 (111.9%)	999
		0	1.1×10^{-4}			364			364			2383			1015
0.3	0.10	2.7×10^{-3}	7.8×10^{-4}	349	380 (108.9%)	384	349	380 (108.9%)	384	2546	2227 (87.5%)	2406	907	1150 (126.8%)	1097
		0	4.4×10^{-4}			380			380			2227			1149
0.5	0.01	1.1×10^{-4}	1.1×10^{-5}	338	342 (101.2)	342	338	342 (101.2)	342	957	969 (101.3%)	971	1839	1855 (100.9%)	1852
		0	6.3×10^{-6}			342			342			969			1855
0.5	0.05	1.2×10^{-3}	2.8×10^{-4}	338	358 (105.9%)	359	338	358 (105.9%)	359	957	1017 (106.3%)	1037	1839	1923 (104.6%)	1881
		0	1.6×10^{-4}			358			358			1017			1923
0.5	0.10	3.5×10^{-3}	1.1×10^{-3}	338	380 (112.4%)	384	338	380 (112.4%)	384	957	1084 (113.3%)	1147	1839	2017 (109.7%)	1889
		0	6.3×10^{-4}			380			380			1084			2017
0.8	0.01	5.6×10^{-5}	4.5×10^{-6}	634	643 (101.4%)	643	634	643 101.4%)	643	851	872 (102.5%)	872	21998	20879 (94.9%)	20879
		0	2.6×10^{-6}			643			643			872			20879
0.8	0.05	6.2×10^{-4}	1.1×10^{-4}	634	682 (107.6%)	683	634	682 (107.6%)	683	851	965 (113.4%)	975	21998	17047 (77.5%)	16435
		0	6.4×10^{-5}			682			682			965			17047
0.8	0.10	1.8×10^{-3}	4.5×10^{-4}	634	736 (116.1%)	741	634	736 (116.1%)	741	851	1111 (130.6%)	1145	21998	13379 (60.8%)	12178
		0	2.6×10^{-4}			736			736			1111			13378

A is Knapp's first approximation assuming HWE.
B is estimate in presence of population stratification, based on $\bar{p}$ and F only.
C is true sample size requirement.

Table 4 Sample size necessary to gain 80% power in TDT with singletons ($\alpha=10^{-7}$, $\gamma=1.5$), comparing Knapp's first approximation assuming HWE to the sample size estimates assuming population stratification

				Multiplicative MOI			*Additive MOI*			*Recessive MOI*			*Dominant MOI*		
$\bar{p}$	F	ϑ_3	ϑ_4	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
0.1	0.01	2.3×10^{-5}	1.4×10^{-6}	2210	2224 (100.6%)	2225	1755	1738 (99.0%)	1739	174 695	151 448 (86.7%)	152 181	2897	2983 (103.0%)	2981
		0	8.1×10^{-7}			2224			1738			151 448			2983
0.1	0.05	2.6×10^{-4}	3.5×10^{-5}	2210	2283 (103.3%)	2287	1755	1671 (95.2%)	1683	174 695	92 276 (52.8%)	96 328	2897	3375 (116.5%)	3349
		0	2.0×10^{-5}			2283			1671			92 277			3375
0.1	0.10	7.5×10^{-4}	1.4×10^{-4}	2210	2361 (106.8%)	2371	1755	1591 (90.7%)	1623	174 695	55 922 (32.0%)	61 637	2897	4011 (138.5%)	3916
		0	8.1×10^{-5}			2361			1591			55 924			4011
0.3	0.01	8.4×10^{-5}	7.8×10^{-6}	1037	1046 (100.9%)	1047	608	611 (100.5%)	611	9097	8963 (98.5%)	8986	2440	2493 (102.2%)	2490
		0	4.4×10^{-6}			1046			611			8963			2493
0.3	0.05	9.4×10^{-4}	1.9×10^{-4}	1037	1084 (104.5%)	1086	608	621 (102.1%)	626	9097	8443 (92.8%)	8686	2440	2730 (111.9%)	2690
		0	1.1×10^{-4}			1084			621			8443			2730
0.3	0.10	2.7×10^{-3}	7.8×10^{-4}	1037	1135 (109.5%)	1142	608	633 (104.1%)	649	9097	7821 (86.0%)	8473	2440	3092 (126.7%)	2956
		0	4.4×10^{-4}			1135			633			7822			3092
0.5	0.01	1.1×10^{-4}	1.1×10^{-5}	947	957 (101.1%)	957	464	469 (101.1%)	470	3099	3133 (101.1%)	3138	4568	4610 (100.9%)	4602
		0	6.3×10^{-6}			957			469			3133			4610
0.5	0.05	1.2×10^{-3}	2.8×10^{-4}	947	999 (105.5%)	1001	464	492 (106.0%)	495	3099	3279 (105.8%)	3343	4568	4789 (104.8%)	4687
		0	1.6×10^{-4}			999			492			3279			4788
0.5	0.10	3.5×10^{-3}	1.1×10^{-3}	947	1057 (111.6%)	1063	464	523 (112.7%)	535	3099	3479 (112.3%)	3690	4568	5034 (110.2%)	4724
		0	6.3×10^{-4}			1057			523			3479			5033
0.8	0.01	5.6×10^{-5}	4.5×10^{-6}	1658	1678 (101.2%)	1679	698	710 (101.7%)	710	2356	2415 (102.5%)	2417	50 826	48 284 (95.0%)	48 116
		0	2.6×10^{-6}			1678			710			2415			48 284
0.8	0.05	6.2×10^{-4}	1.1×10^{-4}	1658	1767 (106.6%)	1769	698	763 (109.3%)	766	2356	2680 (113.8%)	2707	50 826	39 566 (77.8%)	38 166
		0	6.4×10^{-5}			1767			763			2680			39 566
0.8	0.10	1.8×10^{-3}	4.5×10^{-4}	1658	1891 (114.1%)	1899	698	840 (120.3%)	852	2356	3095 (131.4%)	3194	50 826	31 208 (61.4%)	28 448
		0	2.6×10^{-4}			1891			840			3095			31 206

A is Knapp's first approximation assuming HWE.
B is estimate in presence of population stratification, based on $\bar{p}$ and F only.
C is true sample size requirement.

of the informative mating types to each other. These factors should not be ignored if there is the possibility of even mild population stratification.

Interestingly, and somewhat contradictory to initial expectation, population stratification does not always result in decreased power of the TDT because of the loss in heterozygosity. Tables 3 and 4 show that the power of the TDT can be increased in the presence of population stratification because of the change in the relative proportion of the informative mating types and the disease model. If the disease model is a multiplicative MOI, then sample sizes calculated assuming HWE are generally smaller than those actually required. For a dominant, additive or recessive model, the relative frequencies of the AA×AB, AB×BB and AB×AB mating types dictate the change in sample size requirements as the relative information content for these three mating types may not be in the ratio 1:1:2. For example, if the disease is inherited in a dominant MOI, the mating type AA×AB has no information and the mating type AB×BB has the most information.

In a stratified population, the risk allele frequency and mating type frequencies in a single subpopulation may cause subjects from that subpopulation to be more informative than subjects from other subpopulations. This results in one subpopulation that heavily influences the overall TDT sample size calculations. The subjects from other subpopulations may not be as informative and add very little to the overall TDT statistic. This suggests that it may be useful to partition a study population into its component subpopulations prior to analysis.

Larger values of Wright's F were correlated larger discrepancies in the sample size requirements compared to those calculated assuming HWE. For example, in Table 3 where $F=0.01$ and $\gamma=2$, estimates of the sample size required varied from 3% larger to approximately 15% smaller than the sample size calculated using Knapp's first approximation. However, when $F=0.1$ and $\gamma=2$, estimates of the sample size required varied from 39% larger to about 68% smaller than the sample size calculated using Knapp's first approximation.

Wright's F cannot capture all the information about population stratification pertaining to the distortion in mating type frequencies; however, reasonable estimates of the mating type frequencies can be made using F and average allele frequency $\bar{p}$ when $F \leq \min(\bar{p}, 1-\bar{p})$.

To estimate sample sizes for the TDT we ignore the third and higher order risk allele frequency moments (ϑ_3 and ϑ_4). However, small changes in ϑ_3 and ϑ_4 are important and result in sample size discrepancies when our estimates are compared to the true sample size estimates for a fixed F, as seen in Tables 3 and 4.

One potential limitation of our method is that it does not take into account population admixture, as we assume both parents are sampled from the same subpopulation. Population admixture occurs when the study population comprises multiple subpopulations, but there is mating within and between subpopulations. Population admixture is complex, and the rate of admixture depends on socioeconomic, racial, ethnic, linguistic, migratory factors and several other factors. Additional work is needed to assess how the power of the TDT changes in recently admixed populations.

Knapp uses the disease prevalence as a normalizing factor so that the sum of the probabilities of all family types with an affected offspring sums to 1.[9] We note that the disease prevalence, calculated assuming HWE denoted by R, is different from the disease prevalence $\bar{R}$ calculated assuming population stratification. The disease prevalence in the stratified population, $\bar{R} = \sum_{i=1}^{G} w_i R_i = (\psi_2 - 2\psi_1 + 1)F\bar{p}\bar{q} + \psi_2\bar{p}^2 + 2\psi_1\bar{p}\bar{q} + \bar{q}^2$, where $R_i = \psi_2 p_i^2 + 2\psi_1 p_i q_i + q_i$ is the disease prevalence in subpopulation i. We also

note that the normalization factor, $\bar{R}$ in the stratified population is greater in magnitude than the normalization factor in the population in HWE, R, if $\psi_2+1>2\psi_1$, however, if $\psi_2+1<2\psi_1$ the normalization factor in the stratified population, $\bar{R}$ is smaller in magnitude than that of the population in HWE. The difference in the normalization factors between the stratified population and the population in HWE, $F\bar{p}\bar{q}$ $(\psi_2-2\psi_1+1)$, depends on the genetic model chosen and the GRRs. The difference is due to the altered proportion of risk genotypes because of population stratification.

Distortion in the mating type frequencies can be seen when there is underlying population stratification. In a randomly mating population in HWE, $F=\vartheta_3=\vartheta_4=0$. Distortion in the mating type frequencies can also be observed in the parents of an affected child. For example, the parents of a child with a rare recessive disease will most likely be both heterozygous at the disease locus (AB×AB). Similarly, for a rare dominant disorder, the parental genotypes of affected children are most likely a heterozygote and a wild type (AB×BB). This distortion in mating type frequencies becomes most apparent when the risk allele is rare (less than 1% frequency in the population) but confers a substantial risk of disease and the allele explains the majority of the variability seen in the disease and the disease can be considered a simple Mendelian disease from a genetic perspective. The method proposed accounts for the genetic MOI, and takes into account the distortions in the mating type frequencies seen because the parents are the parents of an affected child. Our method is suitable for complex diseases (where the GRR conferred by any single polymorphism is generally less than 3), and the risk allele is relatively common (ie the risk allele frequency in the population is greater than 1%).

DISCUSSION

The TDT was proposed as the solution to the challenge of finding a disease susceptibility gene in a stratified study population. This method was proposed primarily because of its robustness to population stratification and admixture, which made it superior to the traditional case–control tests, which are susceptible to false-positive results. Typically, in the design of family-based studies, HWE is assumed when calculating adequate sample sizes required for a prespecified power level. However, when there is population stratification, there is an increase in homozygosity beyond that expected by HWE. The homozygous parents in the study population do not contribute to the TDT statistic.[15] For example, in a sample of 500 ACTs, in which the estimated disease allele frequency in the parents is 0.1, 180 heterozygous parents would be expected assuming HWE. However, if there is a large degree of population substructure where $F=0.25$, only 135 heterozygous parents are expected under the null hypothesis (the locus of interest is not associated with the disease of interest).

Although there are several methods for calculating the power of the TDT,[9,16,17] none of these methods takes into consideration the power of the TDT in the presence of population stratification. Despite the fact that the TDT maintains the correct Type I error rate, the power of the TDT is significantly affected in the presence of population stratification. This is extremely important, because the TDT is often used when population stratification is possible. Studies with sample sizes calculated ignoring population stratification might be underpowered, and fail to detect putative disease genes. Our method is the first method that proposes incorporating Wright's coefficient of inbreeding, F as a measure of population stratification to approximate the mating type frequencies in the presence of population stratification for TDT sample size calculation.

One of our major points is that in the presence of population stratification, there is considerable variation in the power of the TDT to detect an association between a locus and a disease. The power of the TDT depends on the underlying genetic structure of the study population. We suggest that in future, researchers present estimates of the parameters required to describe the population structure ($\bar{p}$, F, ϑ_3 and ϑ_4) as well as the standard errors of these parameter estimates with all TDT findings. Studies using the TDT to replicate previously published findings may be more challenging than previously thought, as two studies with the same number of subjects could have very different powers to detect the same genetic association depending on the underlying population substructure. This finding may shed light on the cornucopia of studies that have failed to replicate previously published positive findings of association at certain loci. However, one must note that the initial finding may have been spurious.

In addition to possibly losing power because of population stratification (as a result of altered proportions and relative ratios of informative family types), the power of the TDT can be reduced if there is genetic risk heterogeneity between the subpopulations. Genetic risk heterogeneity occurs if the relative risk (RR) of disease conferred by the putative deleterious allele varies from one subpopulation to another. Quantitative risk heterogeneity occurs when the effect of the deleterious allele is not homogenous across all subpopulations, but is more profound in some subpopulations. A classical example for quantitative risk heterogeneity exists between ApoE and Alzheimer's disease, where the association exists pan-ethnically but is strongest in Caucasians and Asians, and weaker in Hispanics and African-Americans.[18] The results of the TDT remain valid when there is quantitative risk heterogeneity, but the power of the TDT to detect this association with a given sample size will vary depending on the ethnic composition of the study sample.

Qualitative risk heterogeneity occurs when one allele is deleterious in one subpopulation ($RR>1$), but is protective in another ($RR<1$). An important point to be emphasized is that the power of the TDT diminishes greatly if there is qualitative risk heterogeneity. In fact, the case–control genomic control method proposed by Devlin and Roeder[3] and the population stratification model (STRAT) proposed by Pritchard *et al*[19] are both more powerful than the TDT per genotyped individual in the presence of qualitative risk heterogeneity as illustrated in Table 2 of Pritchard and Donnelly.[6]

In summary, statistical geneticists and genetic epidemiologists should carefully identify their study population and based on a conservative level of population stratification, follow the guidelines proposed when calculating sample sizes in anticipation of genetic analysis using the TDT and other family-based tests.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Dr Nan Laird, Dr Christoph Lange, Dr Neil Risch and the reviewers for their valuable comments in preparing this manuscript. Ronnie Sebro was supported by a Howard Hughes Medical Institute Pre-Doctoral Fellowship.

1 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1518.
2 Devlin B, Roeder K, Bacanu SA: Unbiased methods for population-based association studies. *Genet Epidemiol* 2001; **21**: 273–284.
3 Devlin B, Roeder K: Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.
4 Knowler WC, Williams RC, Pettitt DJ, Steinberg AG: Gm3-5,13,14 and type-2 diabetes mellitus – an association in American-Indians with genetic admixture. *Am J Hum Genet* 1988; **43**: 520–526.

5 Lander ES, Schork N: Genetic dissection of complex traits. *Science* 1994; **265**: 2037–2048.
6 Pritchard JK, Donnelly P: Case-control studies of association in structured or admixed populations. *Theor Popul Biol* 2001; **60**: 227–237.
7 Pritchard JK, Rosenberg NA: Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999; **65**: 220–228.
8 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.
9 Knapp M: A note on power approximations for the transmission/disequilibrium test. *Am J Hum Genet* 1999; **64**: 1177–1185.
10 Camp NJ: Genomewide transmission/disequilibrium testing – consideration of the genotypic relative risks at disease loci. *Am J Hum Genet* 1997; **61**: 1424–1430.
11 Elandt-Johnson RC: *Probability Models and Statistical Methods in Genetics*. New York: Wiley, 1971.
12 Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press, 1994.
13 Akey JM, Zhang G, Zhang K, Jin L, Shriver MD: Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 2002; **12**: 1805–1814.
14 Yasuda N: An extension of Wahlund's principle to evaluate mating type frequency. *Am J Hum Genet* 1968; **20**: 1–23.
15 Risch N, Teng J: The relative power of family-based and case-control designs for association studies of complex human diseases. I. DNA pooling. *Genome Res* 1998; **8**: 1273–1288.
16 Iles MM: On calculating the power of a TDT study – comparison of methods. *Ann Hum Genet* 2002; **66**: 323–328.
17 Chen WM, Deng HW: A general and accurate approach for computing the statistical power of the transmission disequilibrium test for complex disease genes. *Genet Epidemiol* 2001; **21**: 53–67.
18 Risch N: Searching for genetic determinants in the new millennium. *Nature* 2000; **405**: 847–856.
19 Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: Association mapping in structured populations. *Am J Hum Genet* 2000; **67**: 170–181.