

# True Pedigree Errors More Frequent Than Apparent Errors for Single Nucleotide Polymorphisms

Derek Gordon Simon C. Heath Jürg Ott

Laboratory of Statistical Genetics, Rockefeller University, New York, N.Y., USA

## Key Words

Single nucleotide polymorphisms · Pedigree errors · Probability theory · Haplotype relative risk sampling

## Abstract

Single nucleotide polymorphisms (SNPs) are currently being developed for use in disequilibrium analyses. These SNPs consist of two alleles with varying degrees of polymorphism. A natural design for use with SNPs is the 'haplotype relative risk' sampling design in which a father, mother, and child are typed at an SNP locus. Given such a trio of genotypes, we ask: what is the probability that a pedigree error (a change from one allele to the other) at an SNP locus will be detected using only Mendel's laws as a check? We calculate the probability of detecting such errors for a hypothetical SNP locus with varying degrees of polymorphism and for various true error rates. For the sets of allele frequencies considered, we find that the detection rates range between 25 and 30%, the detection rate being lowest when the two alleles have equal frequencies and the highest when one allele has a frequency of 10%. Based on this detection rate, we determine that the true error rate is roughly 3.3–4 times that of the apparent error rate at an SNP

locus. The greatest discrepancy between true and apparent error rates occurs when allele frequencies are equal.

## Introduction

For disease genes of individually small effects, instead of genetic linkage analysis, disequilibrium analysis may be more promising [1]. Sampling designs have been developed ('haplotypes relative risk' designs) in which the sampling unit is an affected offspring and his or her parents. Marker alleles transmitted to the offspring are contrasted to those not transmitted, where the latter form a contrived control sample [2–4]. Various efficient refinements of this approach have been developed [5, 6]. The data of interest for each of these tests are genotypes at a marker for a family trio, namely a father, mother, and child.

One of the aims in developing marker genes of a new type (single nucleotide polymorphisms; SNPs) is the detection of disease loci by disequilibrium analysis [7]. These SNPs have two alleles at each locus.

We consider the following problem: suppose that, for a random sample of families in which father, mother and

## KARGER

Fax +41 61 306 12 34  
E-Mail [karger@karger.ch](mailto:karger@karger.ch)  
[www.karger.com](http://www.karger.com)

© 1999 S. Karger AG, Basel  
0001-5652/99/0492-0065\$17.50/0

Accessible online at:  
<http://BioMedNet.com/karger>

Derek Gordon  
Laboratory of Statistical Genetics, Rockefeller University, Box 192  
1230 York Avenue, New York, NY 10021 (USA)  
Tel. +1 212 327 7987, Fax +1 212 327 7996, E-Mail [gordon@morgan.rockefeller.edu](mailto:gordon@morgan.rockefeller.edu)  
[heath@fisher.rockefeller.edu](mailto:heath@fisher.rockefeller.edu), [ott@complex.rockefeller.edu](mailto:ott@complex.rockefeller.edu)

child have been typed at an SNP locus, pedigree errors (any changes from one allele to another, e.g. due to sample swaps, genotyping error, etc.) occur randomly and independently with some fixed probability  $\alpha$ . We say that an error in a trio is *undetectable* if the genotypes of the trio still display Mendelian consistency. The question we address is: what is the probability that errors in our sample of family trios go undetected? Let us label this quantity by  $\beta$ . We investigate the question for a single SNP locus with varying degrees of polymorphism. Once we calculate  $\beta$ , we can also calculate the detection rate  $1 - \beta$ , and the apparent error rate,  $\alpha(1 - \beta)$ . We illustrate the meaning of the expression ‘apparent error rate’ with an example. Suppose we have 100 trios, so that we have 600 genotyped alleles. If the true error rate is  $\alpha = 0.05$ , or 5%, then 30 alleles are expected to have changed. If, say, the probability that an error goes undetected for  $\alpha = 0.05$  is  $\beta = 0.40$ , or 40%, the detection rate is  $1 - \beta = 1 - 0.4 = 0.6$ , or 60%. In this example, we will detect only 60% of the 30 errors, or 18 errors. Therefore, it will appear to us that the error rate in our set of genotypes is  $18/600 = 0.03$ , or 3%. This value of 3% is what we call the apparent error rate for this example.

## Methods

For a diallelic marker with allele numbers 1 and 2, there are three possible genotypes: 1/1, 1/2, and 2/2. Assuming Mendel’s laws, there are exactly 15 ways in which parental genotypes at a diallelic marker may be transmitted to a child. We list these possibilities in the table 1. We define the ordered 3-tuple of genotypes (Paternal Allele 1/Paternal Allele 2, Maternal Allele 1/Maternal Allele 2, Child Allele 1/Child Allele 2) in which the set of alleles is consistent with Mendel’s laws, to be a *genotype trio*. In this definition we make no distinction between genotypes 2/1 and 1/2. Thus, we consider the trio (2/1, 2/1, 2/1) to be equal to (1/2, 1/2, 1/2), and so forth. For consistency, we shall always write the genotype 1/2. To distinguish between those 3-tuples that do and do not display Mendelian consistency, we shall use the term *general trio* to refer to *any* ordered 3-tuple of numbers in which each entry is either a 1- or a 2-allele, irrespective of whether the trio displays Mendelian consistency. From these definitions, we see that the set of genotype trios is a subset of the set of general trios. Also, we define the *conjugate* of a genotype trio  $M$  (denoted  $\bar{M}$ ) to be the genotype trio that results when we replace each value of 1 in the trio  $M$  by a 2, and we replace each value of 2 in  $M$  by a 1. For example, the conjugate of the trio (1/1, 1/1, 1/1) is (2/2, 2/2, 2/2), the conjugate of (1/2, 1/1, 1/1) is (2/1, 2/2, 2/2), which we write as (1/2, 2/2, 2/2), and so forth. The list of all conjugates of genotype trios may be found in table 1. We define an *error* in a genotype trio to be a change in one value of the trio. For example, if our original trio is (2/2, 2/2, 2/2) then we say we have introduced one error into that trio if we replace one and only one of the 2-alleles by a 1-allele. We extend our definition, in the obvious way, to speak of two or more errors in a genotype trio.

For each of the 15 trios, we can introduce anywhere from 0 to 6 errors. In the case of either 0 or 6 errors, the resulting general trio will always display Mendelian consistency. We are interested in calculating, for a collection of genotype trios in which at least one error has been introduced, the expected proportion of errors that go undetected. As above, we denote this quantity by  $\beta$ . Using basic probability theory [8], we have

$$\beta = \sum_{i=1}^6 \Pr(\text{undetected errors} \mid i \text{ errors in trio}) \Pr(i \text{ errors in trio}) \quad (1)$$

We do not include  $i = 0$  in our sum because we are only interested in the genotype trios in which errors have been introduced ( $\beta$  is a conditional probability). If we define  $B(\alpha, i)$  by

$$B(\alpha; i) = \binom{6}{i} \alpha^i (1 - \alpha)^{6-i},$$

then, because we assume that error introduction is random and independent for each allele in a genotype trio, and because we are only considering those genotype trios that contain at least one error, the quantity  $\Pr(i \text{ errors in trio})$  in formula (1) is given by

$$\Pr(i \text{ errors in trio}) = \frac{\binom{6}{i} \alpha^i (1 - \alpha)^{6-i}}{1 - (1 - \alpha)^6} = \frac{B(\alpha; i)}{\sum_{i=1}^6 B(\alpha; i)}.$$

Note that the expression  $B(\alpha, i)$  is the probability density function, evaluated at  $i$ ,  $1 \leq i \leq 6$ , for a binomial distribution with constant success rate  $\alpha$  in each of 6 independent experiments.

**Table 1.** List of all genotype trios and their conjugates

Genotype trio = $M$	Conjugate genotype trio $\bar{M}$
(1/1, 1/1, 1/1)	(2/2, 2/2, 2/2)
(1/1, 1/2, 1/1)	(2/2, 1/2, 2/2)
(1/1, 1/2, 1/2)	(2/2, 1/2, 1/2)
(1/2, 1/1, 1/1)	(1/2, 2/2, 2/2)
(1/2, 1/1, 1/2)	(1/2, 2/2, 1/2)
(1/1, 2/2, 1/2)	(2/2, 1/1, 1/2)
(1/2, 1/2, 1/1)	(1/2, 1/2, 2/2)
(1/2, 1/2, 1/2)	(1/2, 1/2, 1/2)
(1/2, 1/2, 2/2)	(1/2, 1/2, 1/1)
(2/2, 1/1, 1/2)	(1/1, 2/2, 1/2)
(1/2, 2/2, 1/2)	(1/2, 1/1, 1/2)
(1/2, 2/2, 2/2)	(1,2, 1/1, 1/1)
(2/2, 1/2, 1/2)	(1/1, 1/2, 1/2)
(2/2, 1/2, 2/2)	(1/1, 1/2, 1/1)
(2/2, 2/2, 2/2)	(1/1, 1/1, 1/1)

Next, we calculate the quantity  $\Pr(\text{undetected errors} \mid i \text{ errors in trio})$  in formula (1). Using basic probability theory, we have

$$\Pr(\text{undetected errors} \mid i \text{ errors in trio}) =$$

$$\sum_{M \in S} \Pr(N_0 \mid M, i) \Pr(M), \quad (2)$$

where  $S$  is the set of all 15 genotype trios (listed in the first column of table 1),  $N_0$  is the event that all errors in a trio go undetected,  $\Pr(N_0 \mid M, i)$  is the probability that  $i$  errors introduced into a genotype trio  $M$  go undetected, and  $\Pr(M)$  is the probability of occurrence of the genotype trio  $M$  in a population of genotype trios.  $\Pr(M)$  is calculated using the allele frequencies  $p$  for allele 1 and  $q (= 1 - p)$  for allele 2. For each genotype trio  $M$  in table 1, we calculate  $\Pr(M)$  in table 3. For each  $M$  and each  $i$ , the probability  $\Pr(N_0 \mid M, i)$  is equal to the proportion of resulting trios that, after  $i$  errors have been introduced, are consistent with Mendel's laws. For example, if  $M = (1/1, 1/1, 1/1)$  and  $i = 1$ ,  $\Pr(N_0 \mid M, i) = 4/6$ , or  $2/3$ ; of the 6 trios that result from a change in one of the alleles, a change in any of the 4 parental alleles will display Mendelian consistency and a change in either of the 2 child alleles will display Mendelian inconsistency.

We now state some lemmas whose application simplifies the calculation of  $\Pr(N_0 \mid M, i)$  for any of the 15 genotype trios  $M$  in table 1 and any  $i$ ,  $0 \leq i \leq 6$ . For the interested reader, we present proofs in the Appendix. In each of these lemmas,  $N_0$  is the event that no errors are detected in a genotype trio.

**Lemma 1.** For any genotype trio  $M$  and for any  $i$ ,  $0 \leq i \leq 6$ ,  $\Pr(N_0 \mid M, i) = \Pr(N_0 \mid \underline{M}, i)$ .

**Lemma 2.** For any genotype trio  $M$  and for any  $i$ ,  $0 \leq i \leq 6$ ,  $\Pr(N_0 \mid M, i) = \Pr(N_0 \mid M, 6 - i)$ .

**Lemma 3.** For genotype trios  $M = (a/b, c/d, e/f)$  and  $N = (c/d, a/b, e/f)$ ,  $\Pr(N_0 \mid M, i) = \Pr(N_0 \mid N, i)$  for any  $i$ ,  $0 \leq i \leq 6$ .

In table 2, we list the probability of detecting  $i$  errors,  $1 \leq i \leq 3$ , for each of 6 genotype trios. We determine these probabilities for each genotype trio  $M$  by listing all possible general trios that result when  $i$  errors have been introduced into  $M$ , and counting the subset of general trios that show Mendelian consistency. Probabilities for the remaining trios in table 1 can be calculated using lemmas 1 and 3. For  $4 \leq i \leq 6$ , we can determine  $\Pr(N_0 \mid M, i)$  using lemma 2.

Given the probabilities in table 2 and lemmas 1–3, we are now able to calculate the value  $\Pr(\text{undetected error} \mid i \text{ errors})$  in formula (2). For each value of  $i$ , this probability will be the sum of the entries

**Table 2.** Conditional probability that a genotype trio  $M$  displays Mendelian consistency if  $i$  errors are introduced,  $1 \leq i \leq 3$

Trio = $M$	$\Pr(N_0 \mid M, 1)$	$\Pr(N_0 \mid M, 2)$	$\Pr(N_0 \mid M, 3)$
(1/1, 1/1, 1/1)	2/3	4/5	3/5
(1/2, 1/1, 1/1)	5/6	3/5	4/5
(1/2, 1/1, 1/2)	2/3	4/5	3/5
(1/1, 2/2, 1/2)	2/3	3/5	9/10
(1/2, 1/2, 1/1)	2/3	4/5	3/5
(1/2, 1/2, 1/2)	1	3/5	7/10

**Table 3.** Expected proportion of undetected errors, given that  $i$  errors are introduced into genotype trio,  $1 \leq i \leq 3$

Trio = $M$	$\Pr(M)$	$\Pr(N_0 \mid M, i)$			$\Pr(N_0 \mid M, i) \Pr(M)$		
		$i = 1$	$i = 2$	$i = 3$	$i = 1$	$i = 2$	$i = 3$
(1/1, 1/1, 1/1)	$p^4$	2/3	4/5	3/5	(2/3) $p^4$	(4/5) $p^4$	(3/5) $p^4$
(1/1, 1/2, 1/1)	$p^3q$	5/6	3/5	4/5	(5/6) $p^3q$	(3/5) $p^3q$	(4/5) $p^3q$
(1/1, 1/2, 1/2)	$p^3q$	2/3	4/5	3/5	(2/3) $p^3q$	(4/5) $p^3q$	(3/5) $p^3q$
(1/2, 1/1, 1/1)	$p^3q$	5/6	3/5	4/5	(5/6) $p^3q$	(3/5) $p^3q$	(4/5) $p^3q$
(1/2, 1/1, 1/2)	$p^3q$	2/3	4/5	3/5	(2/3) $p^3q$	(4/5) $p^3q$	(3/5) $p^3q$
(1/1, 2/2, 1/2)	$p^2q^2$	2/3	3/5	9/10	(2/3) $p^2q^2$	(3/5) $p^2q^2$	(9/10) $p^2q^2$
(1/2, 1/2, 1/1)	$p^2q^2$	2/3	4/5	3/5	(2/3) $p^2q^2$	(4/5) $p^2q^2$	(3/5) $p^2q^2$
(1/2, 1/2, 1/2)	$2 p^2q^2$	1	3/5	7/10	$2 p^2q^2$	(6/5) $p^2q^2$	(7/5) $p^2q^2$
(1/2, 1/2, 2/2)	$p^2q^2$	2/3	4/5	3/5	(2/3) $p^2q^2$	(4/5) $p^2q^2$	(3/5) $p^2q^2$
(2/2, 1/1, 1/2)	$p^2q^2$	2/3	3/5	9/10	(2/3) $p^2q^2$	(3/5) $p^2q^2$	(9/10) $p^2q^2$
(1/2, 2/2, 1/2)	$pq^3$	2/3	4/5	3/5	(2/3) $pq^3$	(4/5) $pq^3$	(3/5) $pq^3$
(1/2, 2/2, 2/2)	$pq^3$	5/6	3/5	4/5	(5/6) $pq^3$	(3/5) $pq^3$	(4/5) $pq^3$
(2/2, 1/2, 1/2)	$pq^3$	2/3	4/5	3/5	(2/3) $pq^3$	(4/5) $pq^3$	(3/5) $pq^3$
(2/2, 1/2, 2/2)	$pq^3$	5/6	3/5	4/5	(5/6) $pq^3$	(3/5) $pq^3$	(4/5) $pq^3$
(2/2, 2/2, 2/2)	$q^4$	2/3	4/5	3/5	(2/3) $q^4$	(4/5) $q^4$	(3/5) $q^4$
Total	1	Not Applicable			See Formulas (2.2a) and (2.2b)		

**Table 4.** Detection rate  $1 - \beta$  and apparent error rate  $\alpha(1 - \beta)$  for various values of true error rate  $\alpha$  and various allele frequencies  $p$  at diallelic locus

True error rate, $\alpha$	Frequency of 1-allele = 0.1		Frequency of 1-allele = 0.5	
	detection rate, $1 - \beta$	apparent error rate, $\alpha(1 - \beta)$	detection rate, $1 - \beta$	apparent error rate, $\alpha(1 - \beta)$
0.0010	0.3032	0.0003	0.2501	0.0003
0.0050	0.3025	0.0015	0.2506	0.0013
0.0100	0.3017	0.0030	0.2512	0.0025
0.0200	0.3001	0.0060	0.2525	0.0050
0.0500	0.2960	0.0148	0.2562	0.0128
0.1000	0.2909	0.0291	0.2622	0.0262
0.2000	0.2862	0.0572	0.2732	0.0546
0.3000	0.2865	0.0860	0.2820	0.0846

under the appropriate heading in the column  $\Pr(N_0 | M, i) \Pr(M)$  in table 3. We see that:

$$\begin{aligned} \sum \Pr(N_0 | M, 1) \Pr(M) &= 2/3 p^4 + 3 p^3 q + 14/3 p^2 q^2 + 3 p q^3 + 2/3 q^4, \\ \sum \Pr(N_0 | M, 2) \Pr(M) &= 4/5 p^4 + 14/5 p^3 q + 4 p^2 q^2 + 14/5 p q^3 + 4/5 q^4, \end{aligned} \quad (2.2a)$$

$$\sum \Pr(N_0 | M, 3) \Pr(M) = 3/5 p^4 + 14/5 p^3 q + 22/5 p^2 q^2 + 14/5 p q^3 + 3/5 q^4.$$

Applying lemma 2, we observe the remarkable fact that

$$\begin{aligned} \sum \Pr(N_0 | M, 4) \Pr(M) &= \sum \Pr(N_0 | M, 2) \Pr(M), \\ \sum \Pr(N_0 | M, 5) \Pr(M) &= \sum \Pr(N_0 | M, 1) \Pr(M). \end{aligned} \quad (2.2b)$$

We substitute the questions (2.2a) and 2.2b) into formula (2), which we then substitute into formula (1) to determine our desired quantity  $\beta$ , the probability that errors in a randomly sampled genotype trio go undetected. In table 4, we calculate this quantity for various values of  $\alpha$  and  $p$ .

## Results and Discussion

From table 4, we observe that, if we use only Mendel's laws to check whether or not a pedigree error has occurred, then the true error rate is roughly 3.3–4 times the magnitude of the apparent error rate. While this magnitude might not appear to be much for very small values of  $\alpha$ , the effect becomes quite significant when one looks at apparent error rates of 0.05 or more. From results in table 4, an apparent error rate of approximately, 0.06 corresponds to a true error rate of approximately 0.20, regardless of the allele frequencies used. Note that for each value of the true error rate  $\alpha$ , the apparent error rates for both sets of allele frequencies are equal to two decimal places. Also note that, when one allele has frequency 0.1, the detection rate *decreases* from 0.3032 when  $\alpha$  is 0.001

to 0.2865 when  $\alpha$  is 0.300, a difference in detection rates of 0.0167. By contrast, when both allele frequencies are 0.5, the detection rate *increases* from 0.2501 when  $\alpha$  is 0.001 to 0.2820 when  $\alpha$  is 0.300, a difference in detection rates of 0.0319, almost twice that of the former difference.

As a comparison to the SNPs scenario, we performed simulations in which we generated 100,000 genotype trios for a marker with 10 equally frequent alleles. We generated the simulated data using the SIMULATE program [9]. We created errors in the data by increasing a 1-allele to 2, decreasing a 10-allele to 9, and for every other allele, either increasing the allele by one or decreasing the allele by one, each with 50% probability. As in the SNPs case, we assumed that the errors occurred independently and randomly at some rate  $\alpha$ . For each setting of  $\alpha$ , we generated a new set of 100,000 genotype trios. The results of these simulations are found in table 5. We also calculate an empirical  $\beta$  by considering the quotient (total number of undetected errors)/(total number of errors introduced), and we report the empirical detection rate,  $1 - \beta$ .

Note that in the case of 10 equally frequent alleles, the lowest detection rate is 0.5850, when the true error rate is 0.0010. By contrast, for the same true error rate, in table 4, using two equally frequent alleles, the detection rate is 0.2500. Comparing tables 4 and 5, we see that in the SNP case, the true error rate is approximately 3.5–4 times that of the apparent error rate when equal allele frequencies are used. In the case of 10 equally frequent alleles, however, the true error rate is approximately 1.3–1.7 times the apparent error rate.

What are possible solutions to the problem of low detection rates for SNPs? One possible solution is to check for pedigree errors by performing multipoint analysis, and looking for multiple recombinations in an interval

**Table 5.** Detection rate  $1 - \beta$  and apparent error rate  $\alpha(1 - \beta)$  for various values of true error rate  $\alpha$  at locus with 10 equally frequent alleles – Simulation results

True error rate, $\alpha$	Empirical detection rate, $1 - \beta$	Apparent error rate, $\alpha(1 - \beta)$
0.0010	0.5850	0.0006
0.0050	0.5900	0.0030
0.0100	0.6010	0.0060
0.0200	0.6050	0.0121
0.0500	0.6290	0.0310
0.1000	0.6630	0.0660
0.2000	0.7160	0.1430
0.3000	0.7510	0.2250

of less than 30 cM [10]. However, with only two alleles at each of the SNP markers, the likelihood of observing true recombinations over a short interval is most probably reduced. The extent of this reduction needs to be researched.

Will genotyping additional siblings help? Possibly, but note that if genotyping is performed at only one locus, additional sibs can aid in finding errors in the parents, not in the first sibling. Also, if one of the alleles has a very high frequency, then the additional sib will provide even less information in determining errors. For example, consider the case where the 1 allele has a population frequency of 0.9. The majority of genotype trios will be (1/1, 1/1, 1/1). Let us assume that only one error occurs in this trio, as is the case when the true error rate is low. If it occurs in one of the parents, then the error goes undetected regardless of which sibling is typed. In the case where multiple loci are genotyped, additional sibs may determine phase in the parents and hence double recombinants in the first child. However, the extent to which genotyping additional sibs helps in finding SNP errors also needs to be researched.

Given these results, we suggest that researchers who look for pedigree errors in SNPs use additional methods beyond checking for Mendelian inconsistency. Suggestions include genotyping trios twice and genotyping multiple loci and looking for double recombinants.

## Appendix

In this section, we present proofs of the three lemmas we stated in the body of the manuscript. We comment that in some of the proofs, we present equations that are equalities of *sets*. Briefly, two sets  $A$  and  $B$  are equal if they contain precisely the same elements. We shall use

the notation  $A \equiv B$  to indicate equality of sets  $A$  and  $B$ . Also, we shall use the notation  $|A|$  to denote the cardinality, or size of the set  $A$ . For example, if the set  $A$  has three elements, then  $|A| = 3$ . Finally, if  $C$  represents a set of general trios, then we denote by  $\underline{C}$  the set of all conjugates of elements of  $C$ . For further information on set theory notation and definitions, please see Hungerford [11].

*Lemma 1.* For any genotype trio  $M$  and for any  $i$ ,  $0 \leq i \leq 6$ ,  $\Pr(N_0 | M, i) = \Pr(N_0 | \underline{M}, i)$ .

*Proof.* Let  $S(M, i)$  be the set of all general trios that result by introducing  $i$  errors into a genotype trio  $M$ . Also, let  $U(M, i)$  be the subset of  $S(M, i)$  of all general trios that are genotype trios, i.e., the trios that display Mendelian consistency. Using the identity  $S(\underline{M}, i) \equiv S(\underline{M}, i)$ , which is straightforward to check, and the fact that a general trio displays Mendelian consistency if and only if its conjugate displays Mendelian consistency, we have the identity  $|U(M, i)| = |U(\underline{M}, i)|$ . Using this identity and the fact that  $|S(M, i)| = |S(\underline{M}, i)|$  we can write:

$$\Pr(N_0 | M, i) = \frac{|U(M, i)|}{|S(M, i)|} = \frac{|U(\underline{M}, i)|}{|S(\underline{M}, i)|} = \Pr(N_0 | \underline{M}, i),$$

which proves the lemma.

*Lemma 2.* For any genotype trio  $M$  and for any  $i$ ,  $0 \leq i \leq 6$ ,  $\Pr(N_0 | M, i) = \Pr(N_0 | M, 6 - i)$ .

*Proof.* We prove this lemma by first proving the identity that, for any genotype trio  $M$  and any  $i$ ,  $0 \leq i \leq 6$ ,

$$\Pr(N_0 | M, i) = \Pr(N_0 | \underline{M}, 6 - i). \quad (3)$$

To prove this identity, note first that we have the equality  $S(M, i) \equiv S(\underline{M}, 6 - i)$ . From this equality it follows immediately that  $U(M, i) \equiv U(\underline{M}, 6 - i)$ . If two sets are equal, then clearly they have the same cardinality. Thus,  $|S(M, i)| = |S(\underline{M}, 6 - i)|$ ,  $|U(M, i)| = |U(\underline{M}, 6 - i)|$ , and

$$\Pr(N_0 | M, i) = \frac{|U(M, i)|}{|S(M, i)|} = \frac{|U(\underline{M}, 6 - i)|}{|S(\underline{M}, 6 - i)|} = \Pr(N_0 | \underline{M}, 6 - i),$$

so the identity is proved. Using identity (3), lemma 1, and the fact that the conjugate of  $\underline{M}$  is  $M$ , we have

$$\Pr(N_0 | M, i) = \Pr(N_0 | \underline{M}, 6 - i) = \Pr(N_0 | M, 6 - i),$$

which proves lemma 2.

*Lemma 3.* For genotype trios  $M = (a/b, c/d, e/f)$  and  $N = (c/d, a/b, e/f)$ ,  $\Pr(N_0 | M, i) = \Pr(N_0 | N, i)$  for any  $i$ ,  $0 \leq i \leq 6$ .

*Proof.* We have  $|S(M, i)| = |S(N, i)|$ . From the fact that a general trio  $M$  displays Mendelian consistency if and only if the general trio  $N$ , determined by reassigning the labels of father and mother in  $M$ , displays Mendelian consistency, it follows that  $|U(M, i)| = |U(N, i)|$ . From these two equalities and the fact that  $\Pr(N_0/A, i) = |U(A, i)|/|S(A, i)|$  for any genotype trio  $A$  and any  $i$ ,  $0 \leq i \leq 6$ , the lemma follows.

## Acknowledgments

This material is based upon work supported by the US Army Medical Research Acquisition Activity under award No. DAMD17-94-J-4406. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the US Army Medical Research Acquisition Activity.

## References

- 1 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996;273:516–1517.
- 2 Falk CT, Rubinstein P: Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* 1987;51:227–233.
- 3 Ott J: Statistical properties of the haplotype relative risk. *Genet Epidemiol* 1989;6:127–130.
- 4 Thomson G: Mapping disease genes: Family-based association studies. *Am J Hum Genet* 1995;57:487–498.
- 5 Terwilliger JD, Ott J: A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum Hered* 1992;42:337–346.
- 6 Spielman RS, Ewens WJ: The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 1996;59:983–989.
- 7 Collins FS, Guyer MS, Charkravarti A: Variations on a theme: Cataloging human DNA sequence variation. *Science* 1997;278:1580–1581.
- 8 Hogg RV, Craig AT: *Introduction to Mathematical Statistics*. New York, Macmillan, 1978.
- 9 Terwilliger JD, Ott J: *Handbook of Human Genetic Linkage*. Baltimore, John Hopkins University Press, 1994.
- 10 Ott J: *Analysis of Human Genetic Linkage*. Baltimore, John Hopkins University Press, 1991.
- 11 Hungerford TW: *Algebra*. New York, Springer, 1984.