Am. J. Hum. Genet. 54:1104–1109, 1994

Computation of Identity-by-Descent Proportions Shared by Two Siblings

Sun-Wei Guo

Department of Biostatistics and Human Genome Center, University of Michigan, Ann Arbor

Summary

I provide a novel approach to computing the mean and variance of the proportion of genetic material shared identical by descent (IBD) by sibling pairs in a specified chromosomal region, conditional on observed marker data. I first show that each chromosome in an offspring can be represented by a two-state Markov chain, with the time parameter being the map distance along the chromosome. On this basis, I show that IBD proportion can be written as a stochastic integral and that the computation of its mean and variance can be reduced to evaluation of an integral of some elementary functions. In addition, I show how Goldgar's model can be extended to include dominance effects. Several examples are provided to illustrate the calculation.

Introduction

Mapping quantitative-trait loci (QTL) in humans is an important but challenging endeavor. Goldgar (1990) proposed an elegant statistical method—which is called the "multipoint identical-by-descent (IBD) method," or "MIM"—for detection of QTL by using multipoint marker data. MIM shows greater power than the Haseman-Elston sib-pair method (Haseman and Elston 1972) when multilocus marker data are available. In a subsequent paper, Goldgar and Oniki (1992) demonstrated by simulation that MIM is comparable in power to parametric multipoint linkage analysis but requires substantially less computation. Moreover, MIM performed better in the presence of polygenic or additional single-locus variation (Goldgar and Oniki 1992). Recently, they developed software that implements MIM and have distributed it to interested researchers (Goldgar 1993). More recently, Schork (1993) extended MIM by incorporating (possibly nonlinear) covariate effects, by analyzing multivariate traits, and by using a robust inference method. Goldgar et al. (1993) also have applied MIM to analyze discrete traits.

These methods and the corresponding software depend critically on correct estimation of R, the expected proportion of genetic material shared IBD by sibling pairs in a specified chromosomal region on the basis of marker information. In table 1 of his paper, Goldgar (1990) listed seven cases for different marker-data configurations and provided the mean and variance of IBD proportion shared by half-sib pairs (the mean and variance of IBD proportion shared by a full-sib pair are $\frac{1}{2}[E(R_m)+E(R_p)]$ and $\frac{1}{4}[V(R_m)+V(R_p)]$, respectively, where $E(R_m)$, $E(R_p)$, $V(R_m)$, and $V(R_p)$ are the means and variances of IBD proportions shared by two maternal and paternal chromosomes in the two sibs). Goldgar's seven cases completely cover the situations in which the marker data in parents are fully informative at either one or both marker loci and there are no missing marker data in the offspring. However, these seven cases *do not* cover situations in which (*a*) marker data in parents may be partially or completely uninformative and/or (*b*) there is a missing marker in the offspring, both of which situations frequently occur. For example, this occurs when a parent is homozygous at both marker loci. Although in this case the mean IBD proportion shared by two half-sibs is simply ½ regardless of the segment length, the variance is difficult to calculate (Thompson 1993). A less trivial example is A_1B_1/A_2B_2

Received December 16, 1993; accepted for publication February 11, 1994.

Address for correspondence and reprints: Dr. Sun-Wei Guo, Department of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029.

0002-9297/94/5406-0020$02.00

$\times$ A_3B_1/A_4B_2 mating. In this case, if one offspring's genotype is A_1B_1/A_4B_1 and the other is A_1A_3 at locus A and B_1B_2 at locus B, then none of the seven cases listed by Goldgar (1990) applies. This problem also occurs when the second offspring's genotype is missing at locus B. Note that while this can be handled easily by cases 8 and 11 in table 1, it should not be confused with cases of (I, U) and (N, U) in Goldgar's notation. Otherwise, the resultant relative difference is >40% for segment length $\lambda \leq 16$ cM.

In this paper, I provide a novel approach to computing the mean and variance of the proportion of genetic material shared IBD by sibling pairs in a specified chromosomal region, conditional on observed marker data. I first show that each chromosome in an offspring can be represented by a two-state Markov chain, with the time parameter being the map distance along the chromosome. On this basis, I will show that IBD proportion can be written as a stochastic integral and that the computation of its mean and variance can be reduced to evaluating an integral of some elementary functions. Moreover, I will show how Goldgar's model can be extended to include dominance effects. In addition, I will provide *nine* more cases that, together with Goldgar's seven cases, cover all possible marker-data configurations.

Markov Chain Representation of Offspring Chromosomes

To compute the mean and variance of IBD proportions, I make the following assumptions: (1) the crossover process along a chromosome is Poisson with intensity 1 morgan (Fisher 1965); (2) there is no sex difference in map length; (3) there is no mutation, translocation, conversion, deletion, or insertion; and (4) the nuclear family is outbred. Goldgar (1990) derived his results by assuming (1), but assumptions (2)–(4) are implicit in his derivation. Assumption (4) is particularly important if one wants to break down, into nuclear families, a pedigree with inbreeding loops, because failure to account for inbreeding will underestimate the true average IBD proportion.

For half-sibs with a common parent, their maternal (or paternal) chromosomes can be represented by a two-state Markov chain that takes values 0 (the parent's maternal chromosome) and 1 (the parent's paternal chromosome), with time parameter being the chromosome length. The transition matrix of the Markov chain is

$$[p_{ij}(t)] = \frac{1}{2}\begin{pmatrix} 1+e^{-2t} & 1-e^{-2t} \\ 1-e^{-2t} & 1+e^{-2t} \end{pmatrix} = \begin{pmatrix} 1-\theta & \theta \\ \theta & 1-\theta \end{pmatrix},$$

where $\theta = \frac{1}{2}(1-e^{-2t})$ is Haldane's (1919) mapping function. For half-sibs 1 and 2, the associated two-state Markov chains $h_1(t)$ and $h_2(t)$ are independent, since the gametogenesises that produced the two siblings are independent.

Computation of the Mean of IBD Proportions

Define the IBD proportion shared by two half-sibs on a chromosome segment of length λ as

$$R(\lambda) = \frac{1}{\lambda}\int_0^{\lambda} \delta[h_1(t), h_2(t)]dt\,, \qquad (1)$$

where $\delta(u, v) = 1$ if $u = v$, or $= 0$ otherwise. $R(\lambda)$ is a random variable; we are interested in calculating the mean and variance of $R(\lambda)$ with or without marker information. That is, we wish to compute

$$E[R(\lambda)\,|\,h_1(0)=i_1,\ h_2(0)=i_2,\ h_1(\lambda)=j_1,\ h_2(\lambda)=j_2]$$

and

$$E[R^2(\lambda)\,|\,h_1(0)=i_1,\ h_2(0)=i_2,\ h_1(\lambda)=j_1,\ h_2(\lambda)=j_2]\,.$$

Here, $i_1, i_2, j_1, j_2 = 0, 1$, which indicates the origin of the genetic material at point $t = 0$ and $t = \lambda$.

With this setup, it is easy to compute the mean and variance of the IBD proportions for half-sib pairs. For example, suppose that two half-sibs share an IBD maternal allele from their common parent at $t = 0$ and at $t = \lambda$ (this corresponds to the (I, I, NR, NR) case in Goldgar's table); then

$$\begin{aligned} &E[R(\lambda)\,|\,h_1(0)=0, h_2(0)=0, h_1(\lambda)=0, h_2(\lambda)=0] \\ &= \frac{1}{\lambda}\int_0^{\lambda} E\{\delta[h_1(t), h_2(t)]\,|\,h_1(0)=h_2(0)=h_1(\lambda) \\ &\qquad =h_2(\lambda)=0\}dt \\ &= \frac{1}{\lambda}\int_0^{\lambda} \sum_{k=0,1} \frac{[p_{0k}(t)p_{k0}(\lambda-t)]^2}{[p_{00}(\lambda)]^2}\,dt \\ &= \frac{1}{2\lambda(1+e^{-2\lambda})^2}\int_0^{\lambda} [(1+e^{-2\lambda})^2+(e^{-2t}+e^{-2(\lambda-t)})^2]dt \end{aligned}$$

$$= \frac{1}{2} + \frac{\theta}{4\lambda(1-\theta)} + \frac{1-2\theta}{4(1-\theta)^2},$$

which agrees with Goldgar (1990).

Computation of the Variance of IBD Proportions

Variance computation can be carried out in a similar way. For example, if markers are completely uninformative at both loci, then

$$\begin{aligned} E[R^2(\lambda)] &= \frac{1}{\lambda^2}\int_0^\lambda \int_0^\lambda E\{\delta[h_1(t), h_2(t)]\delta(h_1(s), h_2(s)]\}dtds \\ &= \frac{2}{\lambda^2}\int_0^\lambda \int_0^s \sum_{k=0,1}\sum_{j=0,1} P[h_1(t)=k]P[h_2(t)=k] \\ &\quad \times p_{kj}^2(s-t)dtds \\ &= \frac{1}{\lambda^2}\int_0^\lambda \int_0^s [p_{00}^2(s-t)+p_{01}^2(s-t)]dtds \\ &= \frac{1}{\lambda^2}\int_0^\lambda \int_0^s [1-2p_{00}(s-t)p_{01}(s-t)]dtds \\ &= \frac{1}{4} + \frac{1}{8\lambda^2}[\lambda-\theta(1-\theta)]. \end{aligned}$$

This gives $V[R(\lambda)] = 1/8\lambda^2[\lambda-\theta(1-\theta)]$. I note that the approximation $V[R(\lambda)] \approx (1 - e^{-2\lambda})/8\lambda$, given by Thompson (1993), is quite good, especially when $\lambda \ll 1$ or $\lambda \gg 1$.

Table 1 gives a complete list of the means and variances of the IBD proportion shared by two half-sibs. It also corrects errors in the variance formulas for cases 6 and 7.

Inclusion of Dominance Effects in Goldgar's Model

Goldgar's model (1990) assumes additive effects both inside and outside a test chromosome region C. This may be too restrictive in some cases. I will show below that inclusion of dominance effects both inside and outside C is possible.

Denote $g_i(t)$ and $h_i(t)$ as the maternal and paternal chromosomes for sibs 1 and 2, respectively, where $0 \leq t \leq \lambda$. For convenience, I make no distinction between a chromosome and a chromosome segment, because computation for the chromosome with several segments can be broken down into computation for each segment (Goldgar 1990). If the gene affecting the trait is located at locus t, the genetic covariance between a pair of relatives can be derived (Malécot 1948; Kempthorne 1969). In particular, for sibs 1 and 2 with trait values X_1 and X_2, respectively, the genetic covariance between X_1 and X_2 is

$$\begin{aligned} \mathrm{cov}(X_1, X_2) &= \tfrac{1}{2}\{P[g_1(t)=g_2(t)] + P[h_1(t)=g_2(t)] \\ &\quad + P[g_1(t)=h_2(t)] + P[h_1(t)=h_2(t)]\}\sigma_a^2 \\ &\quad + \{P[g_1(t)=g_2(t), h_1(t)=h_2(t)] \\ &\quad + P[h_1(t)=g_2(t), g_1(t)=h_2(t)]\}\sigma_d^2, \end{aligned}$$

where σ_a^2 and σ_d^2 are the additive and dominance variances respectively, at locus t. Under the assumption of no epistasis, the contribution to genetic covariance because of genes unlinked to t also can be included easily. Under assumption (4)—of no inbreeding—$P[h_1(t)=g_2(t)] = P[g_1(t)=h_2(t)] = 0$. Thus,

$$\begin{aligned} \mathrm{cov}(X_1, X_2) &= \tfrac{1}{2}\{P[g_1(t)=g_2(t)] + P[h_1(t)=h_2(t)]\}\sigma_a^2 \\ &\quad + \{P[g_1(t)=g_2(t), h_1(t)=h_2(t)]\}\sigma_d^2. \end{aligned}$$

Note that

$$\begin{aligned} &\tfrac{1}{2}\{P[g_1(t)=g_2(t)]+P[h_1(t)=h_2(t)]\} \\ &\quad = P(\text{sibs 1 and 2 share 2 genes IBD at } t) \\ &\qquad + \tfrac{1}{2}P(\text{sibs 1 and 2 share 1 gene IBD at } t) = \pi(t), \end{aligned}$$

where $\pi(t)$ is the proportion of genes shared IBD by sibs 1 and 2 at locus t.

The estimate of the true proportion of the chromosome segment C shared IBD, R^*, proposed by Goldgar (1990), can be rewritten as

$$\begin{aligned} R^* &= \tfrac{1}{2}[E(R_m)+E(R_p)] \\ &= \frac{1}{\lambda}\int_0^\lambda \tfrac{1}{2}\{P[g_1(t)=g_2(t)\,|\,M] \\ &\qquad + P[h_1(t)=h_2(t)\,|\,M]\}dt, \\ &= \frac{1}{\lambda}\int_0^\lambda \pi(t\,|\,M)dt, \end{aligned} \tag{2}$$

where M denotes the marker information. In other words, R^* is the averaged proportion of genes shared IBD by the two sibs in segment C, conditional on observed marker data. Analogously, when M is given, the

Table 1

E(R) and V(R) for Half-Sib Pair, Conditional on Marker Information

Case	i_1	i_2	j_1	j_2	Mean	Variance
1	i	i	i	i	$\frac{2(1-\theta)^2\lambda+\theta(1-\theta)+\lambda(1-2\theta)}{4(1-\theta)^2\lambda}$	$\frac{2\theta^3(1-\theta)\lambda+(1-2\theta)(1-2\theta+2\theta^2)\lambda^2-\theta^2(1-\theta)^2}{16(1-\theta)^4\lambda^2}$
2	i	$1-i$	i	$1-i$	$\frac{2(1-\theta)^2\lambda-\theta(1-\theta)-\lambda(1-2\theta)}{4(1-\theta)^2\lambda}$	$\frac{2\theta^3(1-\theta)\lambda+(1-2\theta)(1-2\theta+2\theta^2)\lambda^2-\theta^2(1-\theta)^2}{16(1-\theta)^4\lambda^2}$
3	i	i	$1-i$	$1-i$	$\frac{2\theta^2\lambda+\theta(1-\theta)-\lambda(1-2\theta)}{4\theta^2\lambda}$	$\frac{2\theta(1-\theta)^3\lambda-(1-2\theta)(1-2\theta+2\theta^2)\lambda^2-\theta^2(1-\theta)^2}{16\theta^4\lambda^2}$
4	i	$1-i$	$1-i$	i	$\frac{2\theta^2\lambda-\theta(1-\theta)+\lambda(1-2\theta)}{4\theta^2\lambda}$	$\frac{2\theta(1-\theta)^3\lambda-(1-2\theta)(1-2\theta+2\theta^2)\lambda^2-\theta^2(1-\theta)^2}{16\theta^4\lambda^2}$
5	$\left\{\begin{matrix}1-i\\ i\\ i\\ i\end{matrix}\right.$	$\begin{matrix}i\\ 1-i\\ i\\ i\end{matrix}$	$\begin{matrix}i\\ i\\ 1-i\\ i\end{matrix}$	$\left.\begin{matrix}i\\ i\\ i\\ 1-i\end{matrix}\right\}$	$\frac{1}{2}$	$\frac{\theta^2\lambda+(1-\theta)^2\lambda-\theta(1-\theta)}{16\theta(1-\theta)\lambda^2}$
6	$\left\{\begin{matrix}i\\ \dots\end{matrix}\right.$	$\begin{matrix}i\\ \dots\end{matrix}$	$\begin{matrix}\dots\\ i\end{matrix}$	$\left.\begin{matrix}\dots\\ i\end{matrix}\right\}$	$\frac{\lambda+\theta(1-\theta)}{2\lambda}$	$\frac{\lambda-2\theta^2(1-\theta)^2-\theta(1-\theta)}{8\lambda^2}$
7	$\left\{\begin{matrix}i\\ \dots\end{matrix}\right.$	$\begin{matrix}1-i\\ \dots\end{matrix}$	$\begin{matrix}\dots\\ i\end{matrix}$	$\left.\begin{matrix}\dots\\ 1-i\end{matrix}\right\}$	$\frac{\lambda-\theta(1-\theta)}{2\lambda}$	$\frac{\lambda-2\theta^2(1-\theta)^2-\theta(1-\theta)}{8\lambda^2}$
8	$\left\{\begin{matrix}\dots\\ i\\ i\\ i\end{matrix}\right.$	$\begin{matrix}i\\ \dots\\ i\\ i\end{matrix}$	$\begin{matrix}i\\ i\\ \dots\\ i\end{matrix}$	$\left.\begin{matrix}i\\ i\\ i\\ \dots\end{matrix}\right\}$	$\frac{(3-4\theta)\lambda+\theta(1-\theta)}{4\lambda(1-\theta)}$	$\frac{(1-\theta)(1-2\theta+4\theta^2)\lambda-\theta(1-\theta)^2(1+\theta)+(1-2\theta)\lambda^2}{16(1-\theta)^2\lambda^2}$
9	$\left\{\begin{matrix}\dots\\ i\\ i\\ 1-i\end{matrix}\right.$	$\begin{matrix}i\\ \dots\\ 1-i\\ i\end{matrix}$	$\begin{matrix}1-i\\ i\\ \dots\\ 1-i\end{matrix}$	$\left.\begin{matrix}i\\ 1-i\\ 1-i\\ \dots\end{matrix}\right\}$	$\frac{\lambda-\theta(1-\theta)}{4\lambda(1-\theta)}$	$\frac{(1-\theta)(1-2\theta+4\theta^2)\lambda-\theta(1-\theta)^2(1+\theta)+(1-2\theta)\lambda^2}{16(1-\theta)^2\lambda^2}$
10	$\left\{\begin{matrix}\dots\\ i\\ i\\ i\end{matrix}\right.$	$\begin{matrix}i\\ \dots\\ i\\ i\end{matrix}$	$\begin{matrix}1-i\\ 1-i\\ 1-i\\ \dots\end{matrix}$	$\left.\begin{matrix}1-i\\ 1-i\\ \dots\\ 1-i\end{matrix}\right\}$	$\frac{(4\theta-1)\lambda+\theta(1-\theta)}{4\theta\lambda}$	$\frac{\theta(3-6\theta+4\theta^2)\lambda-\theta^2(1-\theta)(2-\theta)-(1-2\theta)\lambda^2}{16\theta^2\lambda^2}$
11	$\left\{\begin{matrix}\dots\\ i\\ i\\ 1-i\end{matrix}\right.$	$\begin{matrix}i\\ \dots\\ 1-i\\ i\end{matrix}$	$\begin{matrix}i\\ 1-i\\ 1-i\\ \dots\end{matrix}$	$\left.\begin{matrix}1-i\\ i\\ \dots\\ 1-i\end{matrix}\right\}$	$\frac{\lambda-\theta(1-\theta)}{4\theta\lambda}$	$\frac{\theta(3-6\theta+4\theta^2)\lambda-\theta^2(1-\theta)(2-\theta)-(1-2\theta)\lambda^2}{16\theta^2\lambda^2}$
12	$\left\{\begin{matrix}i\\ \dots\end{matrix}\right.$	$\begin{matrix}\dots\\ i\end{matrix}$	$\begin{matrix}\dots\\ i\end{matrix}$	$\left.\begin{matrix}i\\ \dots\end{matrix}\right\}$	$1-\theta$	$\frac{\lambda-\theta(1-\theta)-2(1-2\theta)^2\lambda^2}{8\lambda^2}$
13	$\left\{\begin{matrix}i\\ \dots\end{matrix}\right.$	$\begin{matrix}\dots\\ i\end{matrix}$	$\begin{matrix}\dots\\ 1-i\end{matrix}$	$\left.\begin{matrix}1-i\\ \dots\end{matrix}\right\}$	θ	$\frac{\lambda-\theta(1-\theta)-2(1-2\theta)^2\lambda^2}{8\lambda^2}$
14	$\left\{\begin{matrix}i\\ \dots\end{matrix}\right.$	$\begin{matrix}\dots\\ i\end{matrix}$	$\begin{matrix}i\\ \dots\end{matrix}$	$\left.\begin{matrix}\dots\\ i\end{matrix}\right\}$	$\frac{1}{2}$	$\frac{\lambda-\theta(1-\theta)-2(1-2\theta)\lambda^2}{16(1-\theta)\lambda^2}$
15	$\left\{\begin{matrix}i\\ \dots\end{matrix}\right.$	$\begin{matrix}\dots\\ i\end{matrix}$	$\begin{matrix}1-i\\ \dots\end{matrix}$	$\left.\begin{matrix}\dots\\ 1-i\end{matrix}\right\}$	$\frac{1}{2}$	$\frac{\lambda-\theta(1-\theta)-2(1-2\theta)\lambda^2}{16\theta\lambda^2}$
16	$\left\{\begin{matrix}i\\ \dots\\ \dots\\ \dots\\ \dots\end{matrix}\right.$	$\begin{matrix}\dots\\ i\\ \dots\\ \dots\\ \dots\end{matrix}$	$\begin{matrix}\dots\\ \dots\\ i\\ \dots\\ \dots\end{matrix}$	$\left.\begin{matrix}\dots\\ \dots\\ \dots\\ i\\ \dots\end{matrix}\right\}$	$\frac{1}{2}$	$\frac{\lambda-\theta(1-\theta)}{8\lambda^2}$

NOTE.—Cases 1–7 were derived by Goldgar (1990). Haldane's (1919) map function is used: $\theta = \frac{1}{2}(1-e^{-2\lambda})$. $i = 0, 1$, depending whether the maternal or the paternal chromosome is inherited.

averaged proportion of segment C that the two sibs share *two* genes IBD can be calculated as

$$S^* = \frac{1}{\lambda}\int_0^\lambda P[g_1(t)=g_2(t), h_1(t)=h_2(t)\,|\,M]dt$$
$$= E\left\{\frac{1}{\lambda}\int_0^\lambda \delta[g_1(t), g_2(t)]\delta[h_1(t), h_2(t)]dt\,|\,M\right\}$$
$$= \frac{1}{\lambda}\int_0^\lambda E\{\delta[g_1(t), g_2(t)]\,|\,M\}E\{\delta[h_1(t), h_2(t)]\,|\,M\}dt, \quad (3)$$

which can be calculated easily. For example, consider the case where the segment C is flanked by two markers λ Morgans apart. For a $A_1B_1/A_2B_1 \times A_2B_2/A_3B_2$ mating, if their two offsprings' genotypes are A_1A_3 and A_2A_3 (the genotypes at locus B are omitted because of their noninformativeness), respectively, then

$$S^* = \frac{1}{\lambda}\int_0^\lambda E\{\delta[g_1(t), g_2(t)]\,|\,g_1(0)=g_2(0)=0\}$$
$$\times E\{\delta[h_1(t), h_2(t)]\,|\,h_1(0)=1, h_2(0)=0\}$$
$$= \frac{1}{\lambda}\int_0^\lambda \sum_{j=0,1} p_{0j}^2(t) \sum_{i=0,1} p_{0i}(t)p_{1i}(t)dt$$
$$= \frac{1}{4\lambda}\int_0^\lambda (1-e^{8t})dt = \frac{1}{4\lambda}\left(\lambda - \frac{1-e^{8\lambda}}{8}\right)$$
$$= \frac{1}{4} - \frac{\theta(1-\theta)(2\theta^2-2\theta+1)}{\lambda}.$$

Thus, Goldgar's model can be extended to include dominance effects inside and/or outside C. Specifically, for sibship data, the model can be formulated as

$$X = G_c + G_d + G_A + G_D + E,$$

with $E(G_c) = E(G_d) = E(G_A) = E(G_D) = E(E) = 0$, and

$$\mathrm{cov}(X_i, X_j) = \begin{cases} R_{ij}V_c + S_{ij}V_d + {}^1/_2V_A + {}^1/_4V_D & \text{if } i \neq j \\ V_c + V_d + V_A + V_D + V_E & \text{if } i = j, \end{cases}$$

where V_c and V_d (V_A and V_D) are the additive and dominance variances, respectively, due to genes inside (outside) the region C, and R_{ij} and S_{ij} can be replaced by their estimations, R_{ij}^* and S_{ij}^*, respectively. Incidentally, if marker data are uninformative, or if the maternal and paternal marker configurations are the same, then the integrands in equations (2) and (3) are the arithmetic and geometric means of $\delta[g_1(t), g_2(t)]$ and $\delta[h_1(t), h_2(t)]$, respectively, since $\sqrt{\delta[g_1(t), g_2(t)]\delta[h_1(t), h_2(t)]} = \delta[g_1(t), g_2(t)]\delta[h_1(t), h_2(t)]$.

Discussion

I have provided a novel approach to computing the mean and variance of the proportion of genetic material shared IBD by sibling pairs in a specified chromosomal region, conditional on observed marker data. I first showed that each chromosome in an offspring can be represented by a two-state Markov chain, with the time parameter being the map distance along the chromosome. On this basis, I showed that IBD proportion can be written as a stochastic integral and that the computation of its mean and variance can be reduced to evaluating an integral of some elementary functions. In addition, I showed how dominance effects also can be included in Goldgar's model.

Throughout this paper, I have assumed that there are no missing marker data in parents and that the phase information is known. This may not be true in reality. In cases when this is not so, one can compute the IBD proportion for any given genotype/phase combination and add up all computed IBD proportions, weighted by genotype frequencies and/or phase probabilities. Note that marker data from other family members would help to determine the genotype/phase distributions.

The proposed method easily can be extended to more than two siblings. This can be accomplished by replacing $\delta(u_1, u_2)$ in equation (1) with $\delta(u_1, u_2, \ldots, u_k)$, where $\delta(u_1, u_2, \ldots, u_k) = 1$ if $u_1 = u_2 = \cdots = u_k$, or = 0 otherwise. Note that these computations serve as an infrastructure upon which many other statistical methods for gene mapping can be developed. For example, with some modifications, the method presented in this paper for computation of IBD proportions shared by two sibs can be extended to compute the expected IBD proportions shared by a group of relatives, on the basis of their marker data (Guo, submitted-*a*). This provides the foundation for a novel gene-mapping approach to mapping complex genetic traits (Guo, submitted-*b*).

Acknowledgments

This research was supported in part by National Institutes of Health grants HG00209 and HG00376 and by the University of Michigan Diabetes Research and Training Center Pilot/Feasibility grant. I would like to thank Dr. Michael

Boehnke for his helpful comments. I also would like to thank two anonymous reviewers for their helpful comments.

References

Fisher RA (1965) The theory of inbreeding, 2d ed. Academic Press, New York.

Goldgar DE (1990) Multipoint analysis of human quantitative genetic variation. Am J Hum Genet 47:957–967

——— (1993) Documentation of MIM software

Goldgar DE, Lewis CM, Gholami K (1993) Analysis of discrete phenotypes using a multipoint identity by descent method: application to the Alzheimer's data set. Proceedings of Genetic Analysis Workshop 8. Genet Epidemiol 10:383–388

Goldgar DE, Oniki RS (1992) Comparison of a multipoint identity-by-descent method with parametric multipoint linkage analysis for mapping quantitative traits. Am J Hum Genet 50:598–606

Guo SW Identity-by-descent proportions: concept, computation, and applications (submitted-*a*)

——— Genomic similarity mapping: a novel approach to mapping complex diseases (submitted-*b*)

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and marker locus. Behav Genet 2:3–19

Kempthorne O (1969) An introduction to genetic statistics. The Iowa State University Press, Ames

Malécot G (1948) Les mathématiques de l'hérédité. Masson, Paris

Schork NJ (1993) Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. Am J Hum Genet 53:1306–1319

Thompson EA (1993) Genetic importance and genomic descent. In Ballou JD, Gilpin M, Foose TJ (eds) Population management for survival and recovery. University of Chicago Press, Chicago