TRIANGLE CONSTRAINTS FOR SIB-PAIR IDENTITY BY DESCENT PROBABILITIES UNDER A GENERAL MULTILOCUS MODEL FOR DISEASE SUSCEPTIBILITY

SANDRINE DUDOIT AND TERENCE P. SPEED

Department of Statistics
University of California, Berkeley
367 Evans Hall, #3860
Berkeley, CA 94720-3860

Technical Report No. 527
July 1998

Abstract. In this paper, we study sib-pair IBD probabilities under a general multilocus model for disease susceptibility which doesn't assume random mating, linkage equilibrium or Hardy-Weinberg equilibrium. We derive the triangle constraints satisfied by affected, discordant and unaffected sib-pair IBD probabilities, as well as constraints distinguishing between sharing of maternal and paternal DNA, under general monotonicity assumptions concerning the penetrance probabilities. The triangle constraints are valid for age and sex-dependent penetrances, and in the presence of parental imprinting. We study the parameterization of sib-pair IBD probabilities for common models, and present examples to demonstrate the impact of non-random mating and the necessity of our assumptions for the triangle constraints. We prove that the affected sib-pair possible triangle is covered by the IBD probabilities of two types of models, one with fixed mode of inheritance and general mating type frequencies, the other with varying mode of inheritance and random mating. Finally, we consider IBD probabilities at marker loci linked to disease susceptibility loci and derive the triangle constraints satisfied by these probabilities.

Key words. Identity by descent, linkage, disease genes, affected sib-pair method, triangle constraints, multilocus sib-pair IBD probabilities, Hardy-Weinberg equilibrium, random mating, linkage equilibrium, imprinting, age and sex-dependent penetrances.

1. Introduction. The affected sib-pair method is used routinely to test for linkage between a marker locus and a disease susceptibility (DS) locus. The method consists of sampling nuclear families with two affected children and establishing the number of chromosomes on which the two sibs share DNA identical by descent (IBD) at a marker. The observed distribution (N_0, N_1, N_2) for the number of affected sib-pairs sharing DNA IBD at the marker on 0, 1, and 2 chromosomes, respectively, is then compared to the expected proportions under random Mendelian segregation, $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Deviations from the $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ null distribution are taken as evidence of linkage between the marker and a DS locus. Several test statistics have been proposed to test for linkage, such as the "mean IBD" statistic,

$N_2 + \frac{1}{2}N_1$ (Blackwelder and Elston [2], Knapp *et al.* [16, 17, 18]). In order to increase the power to detect DS loci, other tests have been suggested which make use of the triangle constraints satisfied by the affected-sib pair IBD probabilities: $\pi_1 \leq \frac{1}{2}$ and $2\pi_0 \leq \pi_1$, where π_i is the probability that an affected sib-pair shares DNA IBD at a DS locus on i, $i = 0, 1, 2$, chromosomes (Holmans [13], Faraway [9], Holmans and Clayton [14], Cordell *et al.* [3], Knapp *et al.* [18]). For a single DS locus model, these constraints have only been proved under the stringent assumptions of random mating and Hardy-Weinberg equilibrium at the DS locus (Suarez *et al.* [35], Holmans [13]). For a model with two unlinked DS loci, the constraints were proved with the additional assumption of linkage equilibrium between the two loci (Cordell *et al.* [3]).

In this paper, we study sib-pair IBD probabilities under a general multilocus model for disease susceptibility which doesn't assume random mating, linkage equilibrium or Hardy-Weinberg equilibrium. We derive the triangle constraints satisfied by affected, discordant and unaffected sib-pair IBD probabilities, as well as constraints distinguishing between sharing of maternal and paternal DNA, under general monotonicity assumptions concerning the penetrance probabilities. The triangle constraints are valid for age and sex-dependent penetrances, and in the presence of parental imprinting. We study the parameterization of sib-pair IBD probabilities for common models, and present examples to demonstrate the impact of non-random mating and the necessity of our assumptions for the triangle constraints. We prove that the affected sib-pair possible triangle is covered by the IBD probabilities of two types of models, one with fixed mode of inheritance and general mating type frequencies, the other with varying mode of inheritance and random mating. In general, a triple (π_0, π_1, π_2) satisfying the triangle constraints corresponds to the IBD probabilities for many different modes of inheritance, thus it is inappropriate to estimate the IBD probabilities and solve for parameters such as penetrances and allele frequencies, unless one has knowledge of the mode of inheritance. Finally, we consider IBD probabilities at marker loci linked to disease susceptibility loci and derive the triangle constraints satisfied by these probabilities. Note that we do not address the question of linkage testing in this paper, but derive properties of the IBD probabilities which may be used subsequently in devising appropriate tests of linkage. Although this paper is concerned with sib-pair IBD probabilities, we present the basic definitions, models and derivations for sibships of arbitrary size and phenotype pattern in order to retain the generality of our approach. We used this approach to derive score tests of linkage for general sibships and optimal weights for combining such test statistics across the various types of sibships [7].

The remainder of this section presents basic definitions and an overview of the affected sib-pair method. We introduce the general multilocus model in Section 2 and derive the conditional distribution of inheritance vectors at DS loci given the phenotype vector of a sibship in Section 3. In Sections 4 and 5, we derive the triangle constraints for a single DS locus model and for a general multilocus model. In Section 6, we consider a single diallelic DS locus and study the parameterization of sib-pair IBD probabilities. Finally, in Section 7, we consider IBD probabilities at marker loci linked to disease susceptibility loci.

1.1. Identity by descent. DNA at the same locus on two homologous chromosomes is said to be *identical by descent (IBD)* if it originated from the same ancestral chromosome. If two homologous chromosomes from *different* people are IBD at some locus, the people are *related.* If two homologous chromosomes from the *same* person are IBD at some locus, this person is *inbred*, i.e. has related parents. Two people, neither of whom is inbred, can share DNA IBD at a particular locus on either 0, 1, or 2 chromosomes.

1.2. Inheritance vectors. Consider a sibship of $k \geq 2$ sibs, and suppose we wish to identify the parental origin of the DNA inherited by each sib at a particular autosomal locus, $\mathcal{L}$ say. Arbitrarily label the paternal chromosomes containing the locus of interest by (1,2), and similarly label the maternal chromosomes by (3,4). The labeling of parental chromosomes is done independently for unlinked loci. The *inheritance vector* (also called *gene-identity state* or *vector of segregation indicators*) of the sibship at the locus $\mathcal{L}$ is the $2k$-vector

$$X(\mathcal{L}) = x = (x_1, x_2, ..., x_{2k-1}, x_{2k}),$$

where for $1 \leq i \leq k$,

$$\begin{aligned} x_{2i-1} &= \text{label of paternal chromosome from which} \\ &\quad\ \text{sib } i \text{ inherited DNA at } \mathcal{L} \\ &= 1 \text{ or } 2; \\ x_{2i} &= \text{label of maternal chromosome from which} \\ &\quad\ \text{sib } i \text{ inherited DNA at } \mathcal{L} \\ &= 3 \text{ or } 4. \end{aligned}$$

According to the above definition there are 2^{2k} inheritance vectors for the sibship. However, inheritance vectors obtained by permuting the labels 1 and 2 and/or 3 and 4 represent the same IBD configuration in terms of sharing of paternal and maternal DNA, so there are only $2^{2(k-1)}$ distinct inheritance vectors. There may be further collapsing of the inheritance vectors into *IBD configurations* in the case of affected only sibships (Ethier and Hodge [8]). Note that the labels 1, 2, 3 and 4 for the parental chromosomes are only meaningful within a sibship and may correspond to different DNA sequences in different sibships. Hence, these "alleles" are neither transportable across families nor functional. If there is no inbreeding at $\mathcal{L}$, then 1, 2, 3 and 4 represent DNA sequences that are distinct by descent. Here we only consider IBD sharing resulting from the $2k$ segregations giving rise to the sibship, and allow inbreeding in the population. In practice, the inheritance vector of a sibship is determined by finding enough polymorphism in the parents to be able to identify the chromosomal fragments transmitted to individuals in the sibship. When IBD information is incomplete, partial information extracted from marker data may be summarized by the *inheritance distribution*, a conditional probability distribution over the possible inheritance vectors at the marker locus (Kruglyak and Lander [20], Kruglyak *et al.* [19]). Risch [33], Holmans [13] and Holmans and Clayton [14] also address the issue of incomplete marker polymorphism.

For sib-pairs, one usually considers three distinct IBD configurations at a particular locus, according to the number of chromosomes sharing DNA IBD at the locus. In some cases (e.g. parental imprinting), it may be appropriate to distinguish the parental origin of the DNA shared IBD by the sib-pair, and consider four IBD configurations as shown in Table 1.

TABLE 1
Sib-pair IBD configurations.

Number IBD	Representative inheritance vector
0	(1,3,2,4)
1 paternal	(1,3,1,4)
1 maternal	(1,3,2,3)
2	(1,3,1,3)

1.3. Phenotype vector. For a disease of interest, denote the *phenotype vector* of the k sibs by the vector ϕ

$$\phi = (\phi_1, ..., \phi_k),$$

where for $1 \leq i \leq k$, ϕ_i is the phenotype indicator of the ith sib

$$\phi_i = \begin{cases} 1, & \text{if the } i\text{th sib is affected,} \\ 0, & \text{if the } i\text{th sib is unaffected.} \end{cases}$$

For sib-pairs, there are three phenotype patterns, corresponding to the number of affected sibs.

TABLE 2
Sib-pair phenotype vectors.

Pattern (abbreviation)	Number of affected sibs	Phenotype vector
Affected sib-pair (ASP)	2	(1,1)
Discordant sib-pair (DSP)	1	(1,0) or (0,1)
Unaffected sib-pair (USP)	0	(0,0)

1.4. The affected sib-pair method. In general, there is an *association* between phenotype and IBD configuration of related individuals at loci linked to DS loci. This may be illustrated in a simple case, by considering IBD sharing and phenotype in sib-pairs for a fully recessive DS locus $\mathcal{D}$, with alleles D and d, where D is recessive with respect to d (i.e. only individuals with genotype DD are affected). To simplify computation, we further assume that both parents are heterozygous. Considering all 16 possible transmission patterns from these two parents to the two children, we build up tables of joint probabilities of phenotype and IBD configuration at the DS locus and at a locus unlinked to the DS locus. Table 3 clearly indicates an *association* between phenotype and IBD configuration *at* the DS locus, while Table 4 indicates *independence* of phenotype and IBD configuration at a locus *unlinked* to the DS locus.

TABLE 3

Joint probability of # affected sibs and # chromosomes sharing DNA IBD at *DS locus.*

		# affected sibs 0	1	2	
# chromosomes	0	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$
sharing DNA IBD	1	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
at DS locus	2	$\frac{3}{16}$	0	$\frac{1}{16}$	$\frac{1}{4}$
		$\frac{9}{16}$	$\frac{3}{8}$	$\frac{1}{16}$	

This association suggests the following strategy for mapping disease genes: take groups of related individuals with particular disease phenotypes and examine the frequencies with which specific IBD configurations arise at candidate DS loci. The most popular strategy is the *affected sib-pair method*, which studies IBD sharing between two sibs affected with the disease of interest. In 1975, Cudworth and Woodrow [5] considered the IBD distribution of the HLA haplotypes of 15 sib-pairs affected with juvenile-onset diabetes mellitus and compared this distribution to the proportions of $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ expected under random Mendelian segregation. They found a significant deviation from the $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ distribution and their study initiated a large body of research on the implication of HLA and other loci in insulin-dependent diabetes mellitus (IDDM) (Day and Simons [6], Thomson and Bodmer [36], Suarez *et al.* [35], Motro and Thomson [26], Louis

TABLE 4
Joint probability of # affected sibs and # chromosomes sharing DNA IBD at a locus unlinked *to the DS locus.*

		# affected sibs 0	1	2	
# chromosomes	0	$\frac{9}{64}$	$\frac{3}{32}$	$\frac{1}{64}$	$\frac{1}{4}$
sharing DNA IBD at a	1	$\frac{9}{32}$	$\frac{3}{16}$	$\frac{1}{32}$	$\frac{1}{2}$
locus *unlinked* to the DS locus	2	$\frac{9}{64}$	$\frac{3}{32}$	$\frac{1}{64}$	$\frac{1}{4}$
		$\frac{9}{16}$	$\frac{3}{8}$	$\frac{1}{16}$	

et al. [23, 22], Payami *et al.* [29, 28], Cox and Spielman [4], to name a few). Since then, the affected sib-pair method has been studied extensively, initially in the context of HLA-disease association and subsequently for various complex diseases (Alzheimer disease [30], schizophrenia [1], atopy [24]) and genome scans (Kruglyak and Lander [20], Feingold *et al.* [10], Feingold and Siegmund [11]). Day and Simons [6] derived the IBD distribution of affected sib-pairs and affected cousin-pairs at a single random mating diallelic DS locus for quasi-recessive and quasi-dominant modes of inheritance (see Section 6). Suarez *et al.* [35] derived the IBD distribution for the three types of sib-pairs (ASPs, DSPs and USPs) at a marker linked to a diallelic random mating DS locus. Motro and Thomson [26], Louis *et al.* [23, 22] and Payami *et al.* [29, 28] considered the problem of estimating the mode of inheritance and the allele frequency for a single diallelic DS locus with random mating, using IBD data from affected sib-pairs and affected sib-trios. Louis *et al.* [22] and Payami *et al.* [29] also studied the impact of selection, non-random mating, meiotic drive and recombination on the IBD probabilities. Risch [31, 32] considered multilocus models, with the usual assumptions of random mating and Hardy-Weinberg equilibrium at the DS loci and linkage equilibrium between the DS loci, and expressed the IBD probabilities of affected relative pairs in terms of λ_R, the risk ratio for relatives of type R compared with population prevalence.

Several test statistics have been suggested to test the null hypothesis of no linkage including the "mean IBD" statistic $N_2 + \frac{1}{2}N_1$ (Blackwelder and Elston [2], Knapp *et al.* [16, 17, 18]), $N_2 + \frac{1}{4}N_1$ (Feingold and Siegmund [11]), likelihood ratio statistics and χ^2 goodness-of-fit statistics, either unrestricted or restricted to the "possible triangle" (Risch [32, 33], Holmans [13], Faraway [9], Feingold *et al.* [10], Holmans and Clayton [14], Cordell *et al.* [3], Knapp *et al.* [18]).

Until now, the properties of the ASP method have been studied mainly under the population genetic assumptions of Hardy-Weinberg equilibrium, random mating, and linkage equilibrium for the DS loci. However, these population genetic assumptions are not only likely to be violated for most diseases of interest, but are hard to verify and their violation has a potentially large impact on the IBD probabilities (cf. Section 6.4). In what follows, we will be concerned with deriving the conditional probabilities of inheritance vectors given the phenotype vector of a sibship, as well as inequalities satisfied by these probabilities, under the general genetic model introduced in the next section.

2. Genetic model. The genetic model consists of three main components: a model for disease susceptibility, connecting disease phenotypes to genotypes at DS loci in the groups of related individuals of interest (Section 2.1); a population genetic model, describing the population joint distribution of genotypes at the DS loci for the relevant founders (Section 2.2); and a segregation model, describing the segregation of alleles at the DS loci during meiosis (Section 2.3).

2.1. Model for disease susceptibility.

2.1.1. Basic model. In our general model, we consider L unlinked autosomal DS loci, $\mathcal{D}^1, \ldots, \mathcal{D}^L$, where $\mathcal{D}^l$ has m_l alleles, $D^l_1, \ldots, D^l_{m_l}$, $l = 1, \ldots, L$. We define the *multilocus penetrance* of a genotype at the L DS loci to be the conditional probability of affectedness given the multilocus genotype at the L DS loci, i.e.

$$f_{i_1 j_1, \ldots, i_L j_L} = pr(\text{Affected} \mid D^1_{i_1} D^1_{j_1}, \ldots, D^L_{i_L} D^L_{j_L}),$$

where $i_l,\ j_l = 1, \ldots, m_l,\ \ l = 1, \ldots, L$, and

$$\begin{aligned} D^l_{i_l} &= \text{paternally inherited allele at locus } \mathcal{D}^l, \\ D^l_{j_l} &= \text{maternally inherited allele at locus } \mathcal{D}^l. \end{aligned}$$

This definition allows the possibility of *parental imprinting*, i.e. different paternal and maternal contributions to disease susceptibility, as observed with Prader-Willi and Angelman syndromes (Lalande [21], Niikawa [27]). The involvement of imprinting was also suggested in the aetiology of atopy (Moffatt *et al.* [24]), IDDM and bipolar affective disorder (Lalande [21]). Special cases of these general penetrances include the multiplicative, additive and heterogeneity models of Risch [31].

In order to derive constraints satisfied by conditional IBD probabilities, we make the following assumption about the dependence structure of genotypes and phenotypes within a family:

▷ `Assumption G1.` Within a family, the phenotype of a particular sib is conditionally independent of the phenotypes and genotypes of his siblings, given his multilocus genotype at $\mathcal{D}^1, \ldots, \mathcal{D}^L$. That is, for a family of k sibs

$$pr(\phi_1, \ldots, \phi_k | sg_1, \ldots, sg_k) = \prod_{i=1}^{k} pr(\phi_i | sg_i),$$

where ϕ_i and sg_i denote the phenotype and multilocus genotype of the ith sib, respectively. This assumption rules out environmental covariance in the sib phenotypes.

Note that the *marginal* penetrances for genotypes at single DS loci depend on conditional genotype frequencies. For example, for $\mathcal{D}^1$

$$\begin{aligned} &pr(\text{Affected} \mid D^1_{i_1} D^1_{j_1}) \\ &\quad = \sum_{i_2 j_2, \ldots, i_L j_L} f_{i_1 j_1, \ldots, i_L j_L}\, pr(D^2_{i_2} D^2_{j_2}, \ldots, D^L_{i_L} D^L_{j_L} \mid D^1_{i_1} D^1_{j_1}). \end{aligned}$$

Also, we have assumed conditional independence of sib phenotypes given *multilocus* genotypes at *all* DS loci, and not given *marginal* genotypes at *individual* loci. Hence, even when computing conditional IBD probabilities at a single DS locus, we need to condition on the genotypes at all DS loci.

2.1.2. Age and sex-dependent penetrances. We may also consider a more general model that allows age and sex-dependent penetrances as follows:

$$f^a_{i_1 j_1, \ldots, i_L j_L} = pr(\text{Affected} \mid D^1_{i_1} D^1_{j_1}, \ldots, D^L_{i_L} D^L_{j_L}, a),$$

where a denotes the age and sex of a particular individual. Autosomal dominant inheritance with age-dependent penetrances has been used to explain the familial aggregation of Alzheimer disease (Pericak-Vance *et al.* [30]), and sex-dependent penetrances may be involved in IDDM (Morahan *et al.* [25]). `Assumption G1` then becomes

▷ `Assumption G1b.` Within a family, the phenotype of a particular sib is conditionally independent of any phenotype, genotype, age and sex data on his siblings, given his multilocus genotype at $\mathcal{D}^1, \ldots, \mathcal{D}^L$, and his age and sex. That is, for a family of k sibs

$$pr(\phi_1, \ldots, \phi_k | sg_1, \ldots, sg_k, a_1, \ldots, a_k) = \prod_{i=1}^{k} pr(\phi_i | sg_i, a_i),$$

where ϕ_i, sg_i and a_i denote the phenotype, multilocus genotype, age and sex of the ith sib, respectively. We can extend this model to accommodate other types of covariates.

2.2. Population genetic model. In order to derive the conditional distribution of inheritance vectors given the phenotype vector of a sibship, we will need to refer to the pairs of genotypes possessed by the parents at the DS loci. Let $pg^l = (pg_1^l, pg_2^l, pg_3^l, pg_4^l)$ denote the *ordered parental genotype* at the DS locus $\mathcal{D}^l$, $l = 1, \ldots, L$, where pg_i^l is the allele at $\mathcal{D}^l$ on the parental chromosome labeled i, $i = 1, 2, 3, 4$. For a DS locus with m alleles, there are $m^2 \times m^2$ ordered parental genotypes. These may be grouped into $(m(m+1)/2)^2$ *parental mating types*, by grouping genotypes which may be obtained from one another by permuting alleles 1 and 2 and/or 3 and 4. Let mt^l denote the parental mating type at the DS locus $\mathcal{D}^l$, and let $pg = (pg^1, \ldots, pg^L)$ and $mt = (mt^1, \ldots, mt^L)$ denote the multilocus ordered parental genotypes and mating types, respectively (see Table 5). For unlinked DS loci, because of the independent labeling of parental chromosomes, all ordered genotypes within a mating type have the same frequency. Hence

$$pr(pg^1, \ldots, pg^L) = \frac{pr(mt^1, \ldots, mt^L)}{\prod_{l=1}^{L} \#\{pg \in mt^l\}},$$

where $\#\{pg \in mt^l\}$ is the number of ordered parental genotypes which are part of the mating type mt^l.

TABLE 5
Representative parental mating type and ordered genotypes at $\mathcal{D}^l$.

Parental mating type mt^l	Parental genotypes pg^l
$[hijk]$	$D_h^l D_i^l \times D_j^l D_k^l$ $D_h^l D_i^l \times D_k^l D_j^l$ $D_i^l D_h^l \times D_j^l D_k^l$ $D_i^l D_h^l \times D_k^l D_j^l$

Most authors assume Hardy-Weinberg equilibrium and random mating at the DS loci, as well as linkage equilibrium between the loci. These assumptions would give expressions for the mating type frequencies in terms of a series $\{p_i^l\}$ of allele frequencies at each DS locus $\mathcal{D}^l$. We avoid these problematic assumptions as far as possible, and study the impact of non-random mating on the IBD probabilities in Section 6.4. Our general model also allows for the possibility of inbreeding at the DS loci and selective disadvantage on affected individuals (Louis *et al.* [22] and Payami *et al.* [29]). A lot may be gained in generality by dropping the usual population genetic assumptions, and with minimal cost in terms of added complexity.

2.3. Segregation model. We will make one last assumption in order to derive constraints satisfied by conditional probabilities of inheritance vectors. Let x^l denote the inheritance vector of a sibship of size k at the DS locus $\mathcal{D}^l$, $l = 1, \ldots, L$, and $x = (x^1, \ldots, x^L)$.

$\triangleright$ `Assumption G2`. There is no *segregation distortion*, i.e. for a sibship of size k with multilocus parental genotype pg and multilocus inheritance vector x at unlinked DS loci

$$pr(x|pg) = \left(\frac{1}{4^k}\right)^L.$$

3. Multilocus conditional distribution of inheritance vectors at DS loci given phenotype vector of sibship.

3.1. Basic model. We will compute the multilocus conditional distribution $pr(x|\phi)$ of the inheritance vectors at all L DS loci given the phenotype vector ϕ of a sibship of size k under our general model. By Bayes Theorem and under `Assumption G2`:

$$(1) \qquad pr(x|\phi) = \frac{\sum_{pg} pr(\phi|x,pg)\, pr(x|pg)\, pr(pg)}{\sum_x \sum_{pg} pr(\phi|x,pg)\, pr(x|pg)\, pr(pg)} = \frac{\sum_{pg} pr(\phi|x,pg)\, pr(pg)}{\sum_x \sum_{pg} pr(\phi|x,pg)\, pr(pg)}.$$

Now, x and pg together yield the multilocus ordered sib genotypes (distinguishing paternally and maternally inherited alleles), $sg_1, \ldots, sg_k$. Hence, under `Assumption G1`:

$$pr(x|\phi) = \frac{\sum_{pg^1} \cdots \sum_{pg^L} \{\prod_{i=1}^k pr(\phi_i|sg_i)\}\; pr(pg^1, \ldots, pg^L)}{\sum_{x^1} \cdots \sum_{x^L} \sum_{pg^1} \cdots \sum_{pg^L} \{\prod_{i=1}^k pr(\phi_i|sg_i)\}\; pr(pg^1, \ldots, pg^L)},$$

where $1 \le i \le k$ labels individual sibs and

$$pr(\phi_i|sg_i) = pr(\phi_i \mid pg^l_{x^l_{2i-1}} pg^l_{x^l_{2i}},\ l = 1, \ldots, L).$$

This multilocus conditional distribution is a function of the multilocus penetrances and the mating type frequencies which we denote by the global parameter ν.

The following proposition allows us to derive constraints satisfied by the IBD probabilities regardless of the mating type frequencies, by deriving a sufficient condition for the constraints which doesn't involve the population genetic model.

PROPOSITION 1. *Under* `Assumption G2`*, a sufficient condition for the following inequality*

$$(2) \qquad \forall\ x^2, \ldots, x^L \quad \sum_{x^1} c(x^1)\; pr(x^1, \ldots, x^L|\phi) \ge 0$$

is $\forall\ mt^1,\ \forall\ pg^2, \ldots, pg^L,\ \forall\ x^2, \ldots, x^L$

$$(3) \qquad \sum_{x^1} c(x^1) \left\{ \sum_{\{pg^1 \in mt^1\}} pr(\phi|x^1, \ldots, x^L, pg^1, \ldots, pg^L) \right\} \ge 0.$$

Proof. For all $x^2, \ldots, x^L$, equation (3) implies

$$\sum_{pg^2} \cdots \sum_{pg^L} \sum_{mt^1} \frac{pr(mt^1, \ldots, mt^L)}{\prod_{l=1}^{L} \#\{pg \in mt^l\}}$$

$$\times \sum_{x^1} c(x^1) \left\{ \sum_{\{pg^1 \in mt^1\}} pr(\phi | x^1, \ldots, x^L, pg^1, \ldots, pg^L) \right\} \geq 0$$

$$\Rightarrow \sum_{x^1} c(x^1) \sum_{pg} pr(\phi | x^1, \ldots, x^L, pg^1, \ldots, pg^L)\, pr(pg^1, \ldots, pg^L) \geq 0$$

$$\Rightarrow \sum_{x^1} c(x^1) \frac{\sum_{pg} pr(\phi | x^1, \ldots, x^L, pg)\, pr(pg)}{\sum_x \sum_{pg} pr(\phi | x^1, \ldots, x^L, pg)\, pr(pg)} \geq 0$$

$$\Rightarrow \sum_{x^1} c(x^1)\; pr(x^1, \ldots, x^L | \phi) \geq 0, \quad \text{which is (2).} \qquad \square$$

This generalizes in the obvious way for constraints at other loci.

3.2. Age and sex-dependent penetrances. In the presence of age and sex-dependent penetrances (or penetrances depending on other covariates), two approaches are possible. In the first, we stratify the IBD data according to phenotype, age and sex, and compute the conditional distribution of the inheritance vectors given the phenotype vector ϕ and the age and sex information a of the sibship. Then, by `Assumptions G1b, G2`:

$$\begin{aligned} pr(x|\phi, a) &= \frac{\sum_{pg} pr(\phi|x, pg, a)\, pr(x|pg, a)\, pr(pg|a)}{\sum_x \sum_{pg} pr(\phi|x, pg, a)\, pr(x|pg, a)\, pr(pg|a)} \\ &= \frac{\sum_{pg} \{\prod_{i=1}^{k} pr(\phi_i | sg_i, a_i)\}\, pr(pg)}{\sum_x \sum_{pg} \{\prod_{i=1}^{k} pr(\phi_i | sg_i, a_i)\}\, pr(pg)}. \end{aligned}$$

With the second approach, we compute the conditional distribution of the inheritance vectors given only ϕ as follows:

$$\begin{aligned} pr(x|\phi) &= \frac{\sum_{pg} \sum_a pr(\phi|x, pg, a)\, pr(x|pg, a)\, pr(pg|a)\, pr(a)}{\sum_x \sum_{pg} \sum_a pr(\phi|x, pg, a)\, pr(x|pg, a)\, pr(pg|a)\, pr(a)} \\ &= \frac{\sum_{pg} \sum_a \{\prod_{i=1}^{k} pr(\phi_i | sg_i, a_i)\}\, pr(pg)\, pr(a)}{\sum_x \sum_{pg} \sum_a \{\prod_{i=1}^{k} pr(\phi_i | sg_i, a_i)\}\, pr(pg)\, pr(a)}, \end{aligned}$$

where $pr(a)$ is the hypothetical probability of a sibship with age and sex a. $pr(x|\phi)$ may be expressed as a mixture of the IBD probabilities conditional on the age and sex information of the sibship,

$$pr(x|\phi) = \sum_a pr(a|\phi)\; pr(x|\phi, a).$$

Consequently, if there are large differences in penetrances between ages and/or sexes (Morahan *et al.* [25]), stratification as in the first approach may increase power to detect linkage in one of the age/sex groups. In either case, Proposition 1 may be modified to accommodate age and sex-dependent penetrances as follows:

PROPOSITION 2. *Under* `Assumption G2` *and for a model with age and sex-dependent penetrances, a sufficient condition for the following inequalities*

$$\forall\ x^2, \ldots, x^L \qquad \sum_{x^1} c(x^1)\, pr(x^1, \ldots, x^L|\phi) \geq 0$$

and

$$\forall\ a,\ \forall\ x^2, \ldots, x^L \qquad \sum_{x^1} c(x^1)\, pr(x^1, \ldots, x^L|\phi, a) \geq 0$$

is $\forall\ a,\ \forall\ mt^1,\ \forall\ pg^2, \ldots, pg^L,\ \forall\ x^2, \ldots, x^L$

$$\sum_{x^1} c(x^1) \left\{ \sum_{\{pg^1 \in mt^1\}} pr(\phi|x^1, \ldots, x^L, pg^1, \ldots, pg^L, a) \right\} \geq 0\ .$$

Hence, the age and sex data a on a sibship may be treated as the parental genotype at another unlinked DS locus, at least for the purpose of deriving constraints. In order to highlight the main ideas in our approach for deriving the triangle constraints, without the complexity of notation introduced by multilocus models with age and sex-dependent penetrances, we will defer the treatment of these models to Section 5. Unless specified otherwise, we will consider the basic model of Section 2.1.1.

4. Constraints for sib-pair conditional IBD probabilities under a general single autosomal disease locus model. In this section we consider a single autosomal DS locus $\mathcal{D}$, with m alleles, $D_1, \ldots, D_m$, and arbitrary mating type frequencies, and we are concerned with deriving constraints satisfied by sib-pair conditional IBD probabilities at this locus.

For affected sib-pairs, let

$$\begin{aligned}\pi_i^{ASP}(\nu) &= pr(\text{Sib-pair shares DNA IBD on } i \text{ chromosomes at } \mathcal{D}|ASP)\\ &= \sum_{\mathcal{C}_i} pr(x|ASP),\end{aligned}$$

where $\mathcal{C}_i = \{x:\ \delta(x_1, x_3) + \delta(x_2, x_4) = i\}$, $i = 0, 1, 2,\ \ x = (x_1, x_2, x_3, x_4)$ denotes the inheritance vector of the sib-pair at $\mathcal{D}$, and $\delta(k, l) = 1$ if $k = l$ and 0, otherwise. In some cases (e.g. parental imprinting), we may be interested in distinguishing between sharing of maternal and paternal DNA by the sib-pair, so let

$$\begin{aligned}\pi_{ij}^{ASP}(\nu) &= pr(\text{Sib-pair shares DNA IBD at } \mathcal{D} \text{ on } i \text{ paternal and}\\ &\qquad\qquad j \text{ maternal chromosomes}|ASP)\\ &= \sum_{\mathcal{C}_{ij}} pr(x|ASP),\end{aligned}$$

where $\mathcal{C}_{ij} = \{x : \ \delta(x_1, x_3) = i, \ \delta(x_2, x_4) = j\}, \ \ i, \ j = 0, 1$. Then

$$\pi_{10}^{ASP}(\nu) + \pi_{01}^{ASP}(\nu) = \pi_{1}^{ASP}(\nu).$$

The DSP and USP IBD probabilities are defined similarly, and we may drop the parameter ν and the sib-pair type to simplify notation when there is no ambiguity. Examples of models for which $\pi_{01} \neq \pi_{10}$ are given in Section 6. We will prove constraints satisfied by the sib-pair conditional IBD probabilities under our general single DS locus model and the following monotonicity assumption concerning the penetrances:

$\triangleright$ **`Assumption M1.`** $\forall \ i, \ j, \ k, \ l = 1, \ldots, m$

[M1a] $$(f_{ik} - f_{jk})(f_{il} - f_{jl}) \geq 0,$$

and

[M1b] $$(f_{ik} - f_{il})(f_{jk} - f_{jl}) \geq 0.$$

For symmetric penetrances (i.e. no parental imprinting), this is equivalent to the existence of an ordering of the alleles at the DS locus such that:

$$\forall \ i = 1, \ldots, m \qquad f_{i1} \leq f_{i2} \leq \ldots \leq f_{im}.$$

`Assumption M1` is satisfied by the usual diallelic recessive, dominant and additive modes of inheritance, but not by over-dominant modes of inheritance (e.g. $f_{11}, f_{22} < f_{12} = f_{21}$). It is also satisfied in the case of parental imprinting where paternally and maternally inherited alleles are ordered differently in terms of "severity" (e.g. $f_{21} \leq f_{11} \leq f_{12}$, $f_{21} \leq f_{22} \leq f_{12}$, where D_2 is protective if paternally inherited, but increases susceptibility if maternally inherited).

PROPOSITION 3. *Under a general single autosomal DS locus model with m alleles, arbitrary mating type frequencies, and* `Assumptions G1, G2, M1`*, the ASP conditional IBD probabilities at the DS locus satisfy the following constraints:*

(4) $$\pi_{10}^{ASP}(\nu) + \pi_{01}^{ASP}(\nu) \leq \pi_{00}^{ASP}(\nu) + \pi_{11}^{ASP}(\nu),$$

$$\pi_{00}^{ASP}(\nu) \leq \pi_{10}^{ASP}(\nu), \tag{5}$$

$$\pi_{00}^{ASP}(\nu) \leq \pi_{01}^{ASP}(\nu). \tag{6}$$

Consequently,

$$\pi_{1}^{ASP}(\nu) \leq \frac{1}{2}, \tag{7}$$

$$\pi_{1}^{ASP}(\nu) \geq 2\pi_{0}^{ASP}(\nu). \tag{8}$$

Hence, the ASP IBD probabilities at the DS locus fall in a triangle with vertices

$$\left(\frac{1}{4},\frac{1}{2},\frac{1}{4}\right), \quad \left(0,\frac{1}{2},\frac{1}{2}\right) \quad \text{and} \quad (0,0,1)$$

which we call the ASP possible triangle. The USP IBD probabilities satisfy the same constraints.

Note that the IBD probabilities satisfy the possible triangle constraints under the assumptions of random mating and Hardy-Weinberg equilibrium at the DS locus, without requiring `Assumption M1`. However, in real mapping situations, it is more likely to encounter non-random mating than modes of inheritance which violate `Assumption M1`. The ASP possible triangle is also referred to in the literature as Holmans' possible triangle.

Proof. The proof relies on Proposition 1 and the fact that if $\tilde{x}$ is obtained from x by permuting 1 and 2 and/or 3 and 4, then $\forall\ mt$

$$\sum_{\{pg \in mt\}} pr(\phi|\tilde{x}, pg) \ = \sum_{\{pg \in mt\}} pr(\phi|x, pg). \tag{9}$$

TABLE 6
Conditional probability of ASP given inheritance vector x and parental genotype pg for a representative mating type mt.

$pr(ASP|x, pg)$

Parental genotype	Inheritance vector x			
pg	(1,3,1,3)	(1,3,1,4)	(1,3,2,3)	(1,3,2,4)
$D_iD_j \times D_kD_l$	f_{ik}^2	$f_{ik}f_{il}$	$f_{ik}f_{jk}$	$f_{ik}f_{jl}$
$D_iD_j \times D_lD_k$	f_{il}^2	$f_{il}f_{ik}$	$f_{il}f_{jl}$	$f_{il}f_{jk}$
$D_jD_i \times D_kD_l$	f_{jk}^2	$f_{jk}f_{jl}$	$f_{jk}f_{ik}$	$f_{jk}f_{il}$
$D_jD_i \times D_lD_k$	f_{jl}^2	$f_{jl}f_{jk}$	$f_{jl}f_{il}$	$f_{jl}f_{ik}$

• To prove that $\pi_{10}^{ASP}(\nu)+\pi_{01}^{ASP}(\nu) \leq \pi_{00}^{ASP}(\nu)+\pi_{11}^{ASP}(\nu)$, it suffices to show that $\forall\ mt$

$$\sum_{\{pg \in mt\}} pr(ASP|(1,3,1,4), pg) \ + \sum_{\{pg \in mt\}} pr(ASP|(1,3,2,3), pg)$$

$$\leq \sum_{\{pg \in mt\}} pr(ASP|(1,3,2,4),pg) \ + \sum_{\{pg \in mt\}} pr(ASP|(1,3,1,3),pg).$$

This inequality is true since

$$\begin{aligned}
&\sum_{\{pg \in mt\}} pr(ASP|(1,3,2,4),pg) \ + \sum_{\{pg \in mt\}} pr(ASP|(1,3,1,3),pg) \\
&\quad - \sum_{\{pg \in mt\}} pr(ASP|(1,3,1,4),pg) \ - \sum_{\{pg \in mt\}} pr(ASP|(1,3,2,3),pg) \\
&= f_{ik}(f_{ik} + f_{jl} - f_{il} - f_{jk}) + f_{il}(f_{il} + f_{jk} - f_{ik} - f_{jl}) \\
&\quad + f_{jk}(f_{jk} + f_{il} - f_{jl} - f_{ik}) + f_{jl}(f_{jl} + f_{ik} - f_{jk} - f_{il}) \\
&= (f_{ik} + f_{jl} - f_{il} - f_{jk})^2 \geq 0.
\end{aligned}$$

- To prove that $\pi_{10}^{ASP}(\nu) \geq \pi_{00}^{ASP}(\nu)$, it suffices to show that $\forall\ mt$

$$\sum_{\{pg \in mt\}} pr(ASP|(1,3,1,4),pg) \ \geq \sum_{\{pg \in mt\}} pr(ASP|(1,3,2,4),pg).$$

This inequality is true since

$$\begin{aligned}
&\sum_{\{pg \in mt\}} pr(ASP|(1,3,1,4),pg) \ - \sum_{\{pg \in mt\}} pr(ASP|(1,3,2,4),pg) \\
&= f_{ik}(f_{il} - f_{jl}) + f_{il}(f_{ik} - f_{jk}) + f_{jk}(f_{jl} - f_{il}) + f_{jl}(f_{jk} - f_{ik}) \\
&= (f_{ik} - f_{jk})(f_{il} - f_{jl}) + (f_{il} - f_{jl})(f_{ik} - f_{jk}) \\
&= 2(f_{ik} - f_{jk})(f_{il} - f_{jl}) \geq 0 \text{ under } \texttt{Assumption M1a}.
\end{aligned}$$

- The proof of $\pi_{01}^{ASP}(\nu) \geq \pi_{00}^{ASP}(\nu)$ is similar and involves

$$\begin{aligned}
&\sum_{\{pg \in mt\}} pr(ASP|(1,3,2,3),pg) \ - \sum_{\{pg \in mt\}} pr(ASP|(1,3,2,4),pg) \\
&= 2(f_{ik} - f_{il})(f_{jk} - f_{jl}) \geq 0 \text{ under } \texttt{Assumption M1b}.
\end{aligned}$$

Equation (8) follows immediately from (5) and (6). Equation (4) implies $\pi_1 \leq \pi_0 + \pi_2 = 1 - \pi_1$, which is (7).

The proof for USPs is similar to that for ASPs, but with f_{ij} replaced by its complement $1 - f_{ij}$. □

PROPOSITION 4. *Under a general single autosomal DS locus model with m alleles, arbitrary mating type frequencies, and* `Assumptions G1, G2, M1`*, the DSP conditional IBD probabilities at the DS locus satisfy the* *following constraints:*

$$\pi_{00}^{DSP}(\nu) + \pi_{11}^{DSP}(\nu) \leq \pi_{10}^{DSP}(\nu) + \pi_{01}^{DSP}(\nu), \tag{10}$$

$$\pi_{10}^{DSP}(\nu) \leq \pi_{00}^{DSP}(\nu), \tag{11}$$

$$\pi_{01}^{DSP}(\nu) \leq \pi_{00}^{DSP}(\nu). \tag{12}$$

Consequently,

$$\pi_{1}^{DSP}(\nu) \geq \frac{1}{2}, \tag{13}$$

$$\pi_{1}^{DSP}(\nu) \leq 2\pi_{0}^{DSP}(\nu). \tag{14}$$

Hence, the DSP IBD probabilities at the DS locus fall in a triangle with vertices

$$\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right), \quad \left(\frac{1}{2}, \frac{1}{2}, 0\right) \quad \text{and} \quad \left(\frac{1}{3}, \frac{2}{3}, 0\right)$$

which we call the DSP possible triangle.

Proof. The proof relies on Proposition 1, equation (9) and the fact that $\forall\ x$ and $\forall\ mt$

$$\sum_{\{pg \in mt\}} pr((1,0)|x, pg) \ = \ \sum_{\{pg \in mt\}} pr((0,1)|x, pg).$$

TABLE 7

Conditional probability of DSP given inheritance vector x and parental genotype pg for a representative mating type mt.

$$pr((1,0)|x, pg)$$

Parental genotype	Inheritance vector x			
pg	(1,3,1,3)	(1,3,1,4)	(1,3,2,3)	(1,3,2,4)
$D_iD_j \times D_kD_l$	$f_{ik}(1-f_{ik})$	$f_{ik}(1-f_{il})$	$f_{ik}(1-f_{jk})$	$f_{ik}(1-f_{jl})$
$D_iD_j \times D_lD_k$	$f_{il}(1-f_{il})$	$f_{il}(1-f_{ik})$	$f_{il}(1-f_{jl})$	$f_{il}(1-f_{jk})$
$D_jD_i \times D_kD_l$	$f_{jk}(1-f_{jk})$	$f_{jk}(1-f_{jl})$	$f_{jk}(1-f_{ik})$	$f_{jk}(1-f_{il})$
$D_jD_i \times D_lD_k$	$f_{jl}(1-f_{jl})$	$f_{jl}(1-f_{jk})$	$f_{jl}(1-f_{il})$	$f_{jl}(1-f_{ik})$

The rest of the proof is similar to the proof for ASPs, and involves the following quantities:

- $$\sum_{\{pg \in mt\}} pr((1,0)|(1,3,1,4), pg) \ + \sum_{\{pg \in mt\}} pr((1,0)|(1,3,2,3), pg)$$
  $$- \sum_{\{pg \in mt\}} pr((1,0)|(1,3,2,4), pg) \ - \sum_{\{pg \in mt\}} pr((1,0)|(1,3,1,3), pg)$$
  $$= \ (f_{ik} + f_{jl} - f_{il} - f_{jk})^2 \geq 0.$$

- $\sum_{\{pg \in mt\}} pr((1,0)|(1,3,2,4),pg) \; - \sum_{\{pg \in mt\}} pr((1,0)|(1,3,1,4),pg)$
  $= \; 2(f_{ik} - f_{jk})(f_{il} - f_{jl}) \geq 0$ under `Assumption M1a`.

- $\sum_{\{pg \in mt\}} pr((1,0)|(1,3,2,4),pg) \; - \sum_{\{pg \in mt\}} pr((1,0)|(1,3,2,3),pg)$
  $= \; 2(f_{ik} - f_{il})(f_{jk} - f_{jl}) \geq 0$ under `Assumption M1b`. $\square$

Since the trinomial probabilities (π_0, π_1, π_2) must be nonnegative and add up to unity, the triple (π_0, π_1, π_2) corresponds to a point in the simplex

$$\mathcal{S} = \{(\pi_0, \pi_1, \pi_2) : \pi_i \geq 0, \; i = 0, 1, 2, \text{ and } \pi_0 + \pi_1 + \pi_2 = 1\}.$$

A convenient way of displaying the trinomial probabilities is using a *barycentric representation*. Barycentric coordinates in the plane represent the triple (π_0, π_1, π_2) by the vector $\pi_0 A_0 + \pi_1 A_1 + \pi_2 A_2$, where the A_i's are fixed vectors in the plane, such as the columns of the 2×3 matrix in the following equation:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & \sqrt{\frac{3}{2}} & 0 \end{bmatrix} \times \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \end{bmatrix}.$$

With this representation, $(0, 0, 1)$ is located at the origin and (π_0, π_1, π_2) are points in an equilateral triangle with sides of length $\sqrt{2}$. The vertices of the triangle correspond to one of the π's being unity, and along the sides of the triangle one of the π's is zero (see Figures 1, 2 p. 19). Holmans [13] uses a different representation for the trinomial probabilities which is two-dimensional and involves only (π_0, π_1). The boundaries of the space for (π_0, π_1) are $\pi_0 = 0$, $\pi_1 = 0$ and $\pi_1 + \pi_0 = 1$, and the boundaries of the ASP possible triangle are $\pi_0 = 0$, $\pi_1 = \frac{1}{2}$ and $\pi_1 = 2\pi_0$.

5. Constraints for sib-pair conditional IBD probabilities under a general multilocus model. In this section we consider a general model with L unlinked autosomal DS loci and define sib-pair multilocus IBD probabilities as follows. For i_l, $j_l = 0, 1$, $\; l = 1, \ldots, L$, let

$$\begin{aligned} \pi^{ASP}_{i_1 j_1, \ldots, i_L j_L}(\nu) &= pr(\text{Sib-pair shares DNA IBD at } \mathcal{D}^l \text{ on } i_l \text{ paternal} \\ &\qquad \text{and } j_l \text{ maternal chromosomes}, l = 1, \ldots, L \mid ASP) \\ &= \sum_{\mathcal{C}_{i_1 j_1, \ldots, i_L j_L}} pr(x^1, \ldots, x^L | ASP), \end{aligned}$$

where

$$\mathcal{C}_{i_1 j_1, \ldots, i_L j_L} = \{(x^1, \ldots, x^L) : \; \delta(x_1^l, x_3^l) = i_l, \; \delta(x_2^l, x_4^l) = j_l, \; l = 1, \ldots, L\},$$

and x^l is the inheritance vector of the sib-pair at the DS locus $\mathcal{D}^l$. The marginal IBD probabilities at the DS locus $\mathcal{D}^l$, $l = 1, \ldots, L$, are defined by

$$\begin{aligned}\pi^{ASP}_{++,\ldots,i_l j_l,\ldots,++}(\nu) &= pr(\text{Sib-pair shares DNA IBD at } \mathcal{D}^l \text{ on } i_l \\ &\qquad \text{paternal and } j_l \text{ maternal chromosomes} \mid ASP) \\ &= \sum_{\mathcal{C}_{++,\ldots,i_l j_l,\ldots,++}} pr(x^1, \ldots, x^L | ASP),\end{aligned}$$

where $\mathcal{C}_{++,\ldots,i_l j_l,\ldots,++} = \{(x^1, \ldots, x^L) : \ \delta(x^l_1, x^l_3) = i_l, \ \delta(x^l_2, x^l_4) = j_l\}$. DSP and USP IBD probabilities are defined similarly.

We will derive constraints satisfied by the sib-pair multilocus IBD probabilities under our general multilocus model and the following monotonicity assumption which is a generalization of `Assumption M1` for multiple loci:

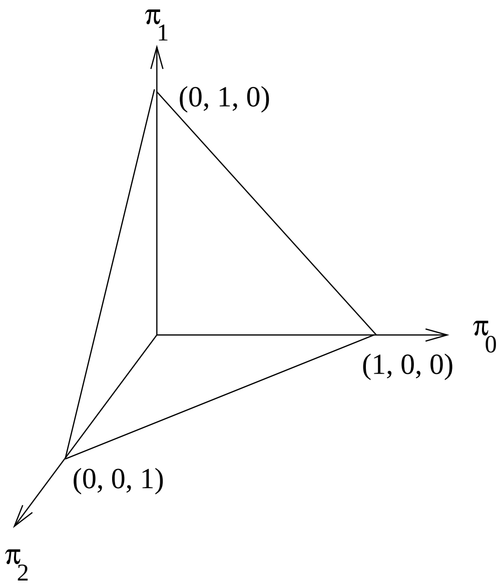

FIG. 1. *Simplex $\mathcal{S}$.*

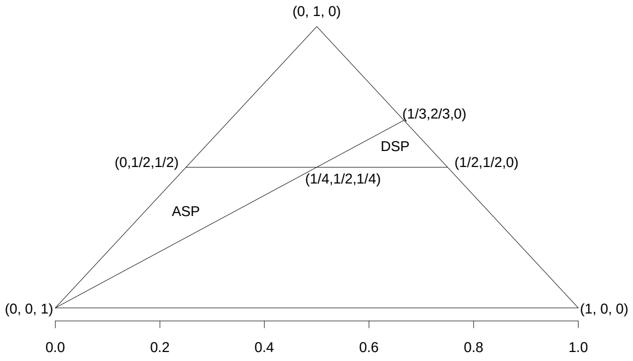

FIG. 2. *ASP and DSP possible triangles.*

$\triangleright$ `Assumption M2`. For each DS locus $\mathcal{D}^h$, $h = 1, \ldots, L$, let g, $\tilde{g}$ denote any two multilocus ordered genotypes at the remaining $L-1$ loci. Then $\forall\ h = 1, \ldots, L, \forall\ g,\ \tilde{g}$ and $\forall\ i,\ j,\ k,\ l = 1, \ldots, m_h$

$$\text{[M2a]} \qquad (f_{ik,g} - f_{jk,g})(f_{il,\tilde{g}} - f_{jl,\tilde{g}}) \geq 0,$$

$$\text{[M2b]} \qquad (f_{ik,g} - f_{il,g})(f_{jk,\tilde{g}} - f_{jl,\tilde{g}}) \geq 0,$$

$$\text{[M2c]} \qquad (f_{ik,g} + f_{jl,g} - f_{il,g} - f_{jk,g})(f_{ik,\tilde{g}} + f_{jl,\tilde{g}} - f_{il,\tilde{g}} - f_{jk,\tilde{g}}) \geq 0,$$

where $f_{ik,g}$ is the probability of affectedness given ordered genotype $D_i^h D_k^h$ at $\mathcal{D}^h$ and g at the remaining $L-1$ loci.

This assumption is satisfied by Risch's [31] multiplicative and additive penetrances, with the single locus constraints (`Assumption M1`) holding at each DS locus.

PROPOSITION 5. *Under a general multilocus model with arbitrary mating type frequencies and* `Assumptions G1,G2, M2`, *the ASP (and USP) multilocus IBD probabilities at the DS loci satisfy the following constraints. For each DS locus $\mathcal{D}^l$, $l=1, \ldots, L$, and $\forall\ i_h,\ j_h = 0, 1,\ h = 1, \ldots, L,\ h \neq l$*

$$(15) \qquad \pi^{ASP}_{i_1j_1,\ldots,00,\ldots,i_Lj_L}(\nu) \leq \pi^{ASP}_{i_1j_1,\ldots,10,\ldots,i_Lj_L}(\nu),$$

$$(16) \qquad \pi^{ASP}_{i_1j_1,\ldots,00,\ldots,i_Lj_L}(\nu) \leq \pi^{ASP}_{i_1j_1,\ldots,01,\ldots,i_Lj_L}(\nu),$$

$$(17) \qquad \pi^{ASP}_{i_1j_1,\ldots,01,\ldots,i_Lj_L}(\nu) + \pi^{ASP}_{i_1j_1,\ldots,10,\ldots,i_Lj_L}(\nu) \leq \pi^{ASP}_{i_1j_1,\ldots,00,\ldots,i_Lj_L}(\nu) + \pi^{ASP}_{i_1j_1,\ldots,11,\ldots,i_Lj_L}(\nu).$$

Consequently, for each DS locus, the marginal IBD probabilities satisfy

$$(18) \qquad \pi^{ASP}_{++,\ldots,00,\ldots,++}(\nu) \leq \pi^{ASP}_{++,\ldots,10,\ldots,++}(\nu),$$

$$(19) \qquad \pi^{ASP}_{++,\ldots,00,\ldots,++}(\nu) \leq \pi^{ASP}_{++,\ldots,01,\ldots,++}(\nu),$$

$$(20) \qquad \pi^{ASP}_{++,\ldots,01,\ldots,++}(\nu) + \pi^{ASP}_{++,\ldots,10,\ldots,++}(\nu) \leq \pi^{ASP}_{++,\ldots,00,\ldots,++}(\nu) + \pi^{ASP}_{++,\ldots,11,\ldots,++}(\nu),$$

and hence the single locus ASP possible triangle constraints are satisfied. The DSP multilocus IBD probabilities satisfy the reverse inequalities

$$(21) \qquad \pi^{DSP}_{i_1j_1,\ldots,00,\ldots,i_Lj_L}(\nu) \geq \pi^{DSP}_{i_1j_1,\ldots,10,\ldots,i_Lj_L}(\nu),$$

$$(22) \qquad \pi^{DSP}_{i_1j_1,\ldots,00,\ldots,i_Lj_L}(\nu) \geq \pi^{DSP}_{i_1j_1,\ldots,01,\ldots,i_Lj_L}(\nu),$$

$$(23) \qquad \pi^{DSP}_{i_1j_1,\ldots,01,\ldots,i_Lj_L}(\nu) + \pi^{DSP}_{i_1j_1,\ldots,10,\ldots,i_Lj_L}(\nu) \geq \pi^{DSP}_{i_1j_1,\ldots,00,\ldots,i_Lj_L}(\nu) + \pi^{DSP}_{i_1j_1,\ldots,11,\ldots,i_Lj_L}(\nu).$$

It follows that the marginal DSP IBD probabilities at each DS locus fall in the single locus DSP possible triangle.

Proof. Without loss of generality, we will prove the inequalities for $\mathcal{D}^1$ using Proposition 1. We will use the fact that if $\tilde{x}^1$ is obtained from x^1 by permuting the labels 1 and 2 and/or 3 and 4, then

$$\sum_{\{pg^1 \in mt^1\}} pr(\phi|\tilde{x}^1, \ldots, x^L, pg^1, \ldots, pg^L) = \sum_{\{pg^1 \in mt^1\}} pr(\phi|x^1, \ldots, x^L, pg^1, \ldots, pg^L).$$

TABLE 8

Conditional probability of ASP given inheritance vectors $x^1, \ldots, x^L$ and parental genotypes $pg^1, \ldots, pg^L$ for a representative mating type mt^1. sg_1 and sg_2 denote the multilocus genotypes of the two sibs at the last $L-1$ loci, as specified by $x^2, \ldots, x^L$ and $pg^2, \ldots, pg^L$.

$pr(ASP|x^1, \ldots, x^L, pg^1, \ldots, pg^L)$

	Inheritance vector at DS locus $\mathcal{D}^1$, x^1			
pg^1	(1,3,1,3)	(1,3,1,4)	(1,3,2,3)	(1,3,2,4)
$D_i^1 D_j^1 \times D_k^1 D_l^1$	$f_{ik,sg_1} f_{ik,sg_2}$	$f_{ik,sg_1} f_{il,sg_2}$	$f_{ik,sg_1} f_{jk,sg_2}$	$f_{ik,sg_1} f_{jl,sg_2}$
$D_i^1 D_j^1 \times D_l^1 D_k^1$	$f_{il,sg_1} f_{il,sg_2}$	$f_{il,sg_1} f_{ik,sg_2}$	$f_{il,sg_1} f_{jl,sg_2}$	$f_{il,sg_1} f_{jk,sg_2}$
$D_j^1 D_i^1 \times D_k^1 D_l^1$	$f_{jk,sg_1} f_{jk,sg_2}$	$f_{jk,sg_1} f_{jl,sg_2}$	$f_{jk,sg_1} f_{ik,sg_2}$	$f_{jk,sg_1} f_{il,sg_2}$
$D_j^1 D_i^1 \times D_l^1 D_k^1$	$f_{jl,sg_1} f_{jl,sg_2}$	$f_{jl,sg_1} f_{jk,sg_2}$	$f_{jl,sg_1} f_{il,sg_2}$	$f_{jl,sg_1} f_{ik,sg_2}$

• To prove that $\pi^{ASP}_{00,i_2j_2,\ldots,i_Lj_L}(\nu) \leq \pi^{ASP}_{10,i_2j_2,\ldots,i_Lj_L}(\nu)$, it suffices to show that $\forall\ mt^1,\ \forall\ pg^2, \ldots, pg^L,\ \forall\ x^2, \ldots, x^L$

$$\sum_{\{pg^1 \in mt^1\}} pr(ASP|(1,3,1,4), x^2, \ldots, x^L, pg^1, \ldots, pg^L) \geq \sum_{\{pg^1 \in mt^1\}} pr(ASP|(1,3,2,4), x^2, \ldots, x^L, pg^1, \ldots, pg^L).$$

This inequality is true since under `Assumption M2a`

$$\begin{aligned}
&\sum_{\{pg^1 \in mt^1\}} pr(ASP|(1,3,1,4), x^2, \ldots, x^L, pg^1, \ldots, pg^L) \\
&\quad - \sum_{\{pg^1 \in mt^1\}} pr(ASP|(1,3,2,4), x^2, \ldots, x^L, pg^1, \ldots, pg^L) \\
&= f_{ik,sg_1}(f_{il,sg_2} - f_{jl,sg_2}) + f_{il,sg_1}(f_{ik,sg_2} - f_{jk,sg_2}) \\
&\quad + f_{jk,sg_1}(f_{jl,sg_2} - f_{il,sg_2}) + f_{jl,sg_1}(f_{jk,sg_2} - f_{ik,sg_2}) \\
&= (f_{ik,sg_1} - f_{jk,sg_1})(f_{il,sg_2} - f_{jl,sg_2}) \\
&\quad + (f_{il,sg_1} - f_{jl,sg_1})(f_{ik,sg_2} - f_{jk,sg_2}) \geq 0.
\end{aligned}$$

• The proof of $\pi^{ASP}_{00,i_2j_2,\ldots,i_Lj_L}(\nu) \leq \pi^{ASP}_{01,i_2j_2,\ldots,i_Lj_L}(\nu)$ involves expressions similar to those above, and `Assumption M2b`.

• To prove that

$$\pi^{ASP}_{01,i_2j_2,\ldots,i_Lj_L}(\nu)+\pi^{ASP}_{10,i_2j_2,\ldots,i_Lj_L}(\nu)\leq\pi^{ASP}_{00,i_2j_2,\ldots,i_Lj_L}(\nu)+\pi^{ASP}_{11,i_2j_2,\ldots,i_Lj_L}(\nu)$$

it suffices to show that $\forall\ mt^1,\ \forall\ pg^2,\ldots,pg^L,\ \forall\ x^2,\ldots,x^L$

$$\begin{aligned}
&\sum_{\{pg^1\in mt^1\}} pr(ASP|(1,3,1,4),x^2,\ldots,x^L,pg^1,\ldots,pg^L)\\
&\quad+\sum_{\{pg^1\in mt^1\}} pr(ASP|(1,3,2,3),x^2,\ldots,x^L,pg^1,\ldots,pg^L)\\
\leq&\sum_{\{pg^1\in mt^1\}} pr(ASP|(1,3,1,3),x^2,\ldots,x^L,pg^1,\ldots,pg^L)\\
&\quad+\sum_{\{pg^1\in mt^1\}} pr(ASP|(1,3,2,4),x^2,\ldots,x^L,pg^1,\ldots,pg^L).
\end{aligned}$$

This inequality is true since under `Assumption M2c`

$$\begin{aligned}
&\sum_{\{pg^1\in mt^1\}} pr(ASP|(1,3,1,3),x^2,\ldots,x^L,pg^1,\ldots,pg^L)\\
&\quad+\sum_{\{pg^1\in mt^1\}} pr(ASP|(1,3,2,4),x^2,\ldots,x^L,pg^1,\ldots,pg^L)\\
&\quad-\sum_{\{pg^1\in mt^1\}} pr(ASP|(1,3,1,4),x^2,\ldots,x^L,pg^1,\ldots,pg^L)\\
&\quad-\sum_{\{pg^1\in mt^1\}} pr(ASP|(1,3,2,3),x^2,\ldots,x^L,pg^1,\ldots,pg^L)\\
=&\ f_{ik,sg_1}(f_{ik,sg_2}+f_{jl,sg_2}-f_{il,sg_2}-f_{jk,sg_2})\\
&\quad+f_{il,sg_1}(f_{il,sg_2}+f_{jk,sg_2}-f_{ik,sg_2}-f_{jl,sg_2})\\
&\quad+f_{jk,sg_1}(f_{jk,sg_2}+f_{il,sg_2}-f_{jl,sg_2}-f_{ik,sg_2})\\
&\quad+f_{jl,sg_1}(f_{jl,sg_2}+f_{ik,sg_2}-f_{jk,sg_2}-f_{il,sg_2})\\
=&\ (f_{ik,sg_1}+f_{jl,sg_1}-f_{il,sg_1}-f_{jk,sg_1})\\
&\ \cdot(f_{ik,sg_2}+f_{jl,sg_2}-f_{il,sg_2}-f_{jk,sg_2})\geq 0.
\end{aligned}$$

The proof of the USP inequalities follows immediately by replacing the penetrances by their complements. The inequalities for DSPs also follow immediately by replacing the penetrances involving sg_2 by their complements. □

The proof of corresponding constraints for age and sex-dependent penetrances, with or without conditioning on the age and sex information of the sibship, is similar and relies on Proposition 2. It involves penetrances of the form $f^a_{ik,g}$ and constraints similar to those in `Assumption M2`, namely: $\forall\ a,\tilde{a},\ \forall\ h=1,\ldots,L,\forall\ g,\ \tilde{g}$ and $\forall\ i,\ j,\ k,\ l=1,\ldots,m_h$

$$(f^{a}_{ik,g} - f^{a}_{jk,g})(f^{\tilde{a}}_{il,\tilde{g}} - f^{\tilde{a}}_{jl,\tilde{g}}) \geq 0,$$
$$(f^{a}_{ik,g} - f^{a}_{il,g})(f^{\tilde{a}}_{jk,\tilde{g}} - f^{\tilde{a}}_{jl,\tilde{g}}) \geq 0,$$
$$(f^{a}_{ik,g} + f^{a}_{jl,g} - f^{a}_{il,g} - f^{a}_{jk,g})(f^{\tilde{a}}_{ik,\tilde{g}} + f^{\tilde{a}}_{jl,\tilde{g}} - f^{\tilde{a}}_{il,\tilde{g}} - f^{\tilde{a}}_{jk,\tilde{g}}) \geq 0.$$

Under Risch's [31] multiplicative model with Hardy-Weinberg equilibrium, random mating and linkage equilibrium,

$$pr(x^1, \ldots, x^L|\phi) = \prod_{l=1}^{L} pr(x^l|\phi),$$

consequently,

$$\pi_{i_1 j_1, \ldots, i_L j_L} = \prod_{l=1}^{L} \pi_{++, \ldots, i_l j_l, \ldots, ++}.$$

6. Single diallelic DS locus models. In this section, we will study the parameterization of the ASP IBD probabilities under common models for disease susceptibility. We will give examples of genetic models for which the triangle constraints are violated and examine the impact of non-random mating on the IBD probabilities. Note that the models considered in this section may not always be realistic, but are nevertheless useful for our purpose.

Consider a single DS locus $\mathcal{D}$ with two alleles: a "disease" allele, D, and a "wild-type" allele, d, and define three penetrance probabilities as follows:

$$f_2 = pr(\text{Affected} \mid DD),$$
$$f_1 = pr(\text{Affected} \mid Dd) = pr(\text{Affected} \mid dD),$$
$$f_0 = pr(\text{Affected} \mid dd).$$

Common models for the penetrances are:

- *Strict-recessive:* $0 = f_0 = f_1 < f_2$ (Thomson and Bodmer [36]);
- *Quasi-recessive:* $0 < f_0 = f_1 \leq f_2 = rf_0$ (Day and Simons [6]);
- *Strict-dominant:* $0 = f_0 < f_1 = f_2$ (Thomson and Bodmer [36]);
- *Quasi-dominant:* $0 < f_0 \leq f_1 = f_2 = rf_0$ (Day and Simons [6]);
- *Additive:* $f_1 = \frac{f_0+f_2}{2}$ (Motro and Thomson [26]);
- *Intermediate:* $0 = f_0 \leq f_1 = sf_2 \leq f_2$ (Spielman *et al.* [34], Louis *et al.* [23]).

There are nine different mating types (Table 9), with frequencies denoted by P_{mt}, $mt = 1, \ldots, 9$.

6.1. Random mating and Hardy-Weinberg equilibrium models. Denote the frequencies of alleles D and d in the population of interest by p and $q = 1 - p$, respectively. Assume that mating is random at $\mathcal{D}$

TABLE 9

Parental mating types, ordered genotypes and their frequencies for a diallelic DS locus with random mating and Hardy-Weinberg equilibrium.

Parental mating type, mt	Parental genotypes, pg	Random mating and HW frequencies, P_{mt}
1	$DD \times DD$	p^4
2	$DD \times Dd$	p^3q
	$DD \times dD$	p^3q
3	$Dd \times DD$	p^3q
	$dD \times DD$	p^3q
4	$DD \times dd$	p^2q^2
5	$dd \times DD$	p^2q^2
6	$Dd \times Dd$	p^2q^2
	$Dd \times dD$	p^2q^2
	$dD \times Dd$	p^2q^2
	$dD \times dD$	p^2q^2
7	$Dd \times dd$	pq^3
	$dD \times dd$	pq^3
8	$dd \times Dd$	pq^3
	$dd \times dD$	pq^3
9	$dd \times dd$	q^4

and the three genotypes DD, Dd and dd have the Hardy-Weinberg (HW) frequencies p^2, $2p(1-p)$ and $(1-p)^2$, respectively. We will give a detailed treatment of ASP IBD probabilities only, since they are more frequently used and involve simpler expressions than DSP and USP IBD probabilities. The ASP IBD probabilities at the DS locus only depend on the allele frequency p and on the ratios of penetrances, f_0/f_2 and f_1/f_2. We will examine the ASP IBD probabilities under six common penetrance models, assuming random mating and Hardy-Weinberg equilibrium.

6.1.1. Strict-recessive model. (See Figure 3 p. 26.)

$$(\pi_0, \pi_1, \pi_2) = \left(\frac{p^2}{(1+p)^2}, \frac{2p}{(1+p)^2}, \frac{1}{(1+p)^2}\right), \ 0 < p \leq 1.$$

The curve traced by the recessive probabilities when p varies is the Hardy-Weinberg curve $\pi_1^2 = 4\pi_0\pi_2$ joining (0,0,1) to $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Note that these IBD probabilities are independent of f_2. More generally, it may be shown that the ASP IBD probabilities lie on the Hardy-Weinberg curve if $f_1^2 = f_0 f_2$ (Knapp *et al.* [16]).

6.1.2. Quasi-recessive model. (See Figures 3 and 5 pp. 26, 28.)

$$\pi_0 = \frac{p^4r^2 + (2p^2(1-p^2))r + (1-p^2)^2}{p^2(1+p)^2r^2 + 2p^2(1-p)(3+p)r + (1-p)(-p^3 - 3p^2 + 4p + 4)},$$

$$\pi_1 = \frac{2(1-2p^2+p^3+2p^2(1-p)r+p^3r^2)}{p^2(1+p)^2r^2+2p^2(1-p)(3+p)r+(1-p)(-p^3-3p^2+4p+4)},$$
$$\pi_2 = \frac{p^2r^2+1-p^2}{p^2(1+p)^2r^2+2p^2(1-p)(3+p)r+(1-p)(-p^3-3p^2+4p+4)}.$$

Note that f_0 cancels out, and as $r \to \infty$ we get the probabilities for the strictly recessive case. We proved that for fixed $p \in (0,1)$, the IBD probabilities lie on a line going from $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ to $(\frac{p^2}{(1+p)^2}, \frac{2p}{(1+p)^2}, \frac{1}{(1+p)^2})$, and given by

$$\pi_0 = \frac{1}{3+p}\left((1+p)-(1+3p)\pi_2\right),$$
$$\pi_1 = \frac{2}{3+p}\left(1+(p-1)\pi_2\right).$$

Hence, the trinomial probabilities may be re-parameterized as

$$t \longrightarrow t\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) + (1-t)\left(\frac{p^2}{(1+p)^2}, \frac{2p}{(1+p)^2}, \frac{1}{(1+p)^2}\right), \ 0 < t \leq 1,$$

where $t \to 0$ yields the strict-recessive case, and $t = 1$ corresponds to the case of no allele influencing DS at the candidate locus. The parameter t is used to obtain a simpler parameterization and has no direct genetic interpretation. t may be expressed as a function of (r, p)

$$t = \frac{4(1-2p^2+2p^2r)}{(4-7p^2+2p^3+p^4)+2p^2(3-2p-p^2)r+p^2(1+p)^2r^2}.$$

Each point strictly between the Hardy-Weinberg curve and the line $\pi_1 = 2\pi_0$ ($\pi_1^2 < 4\pi_0\pi_2$ and $\pi_1 > 2\pi_0$) corresponds to the ASP conditional IBD probabilities for a unique (up to f_0) quasi-recessive model. The parameters of this model are

$$p = \frac{1-3\pi_0-\pi_2}{3\pi_2+\pi_0-1} \in (0,1), \tag{24}$$

$$\begin{aligned} r = {} & \frac{4\pi_2(-2\pi_2^2+\pi_2(3-8\pi_0)+(-1+5\pi_0-6\pi_0^2))}{(-1+3\pi_0+\pi_2)(\pi_0^2+(-1+\pi_2)^2-2\pi_0(1+\pi_2))} \\ & + \frac{2(-1+\pi_0+3\pi_2)^2\sqrt{\pi_0}\sqrt{-1+2(\pi_0+\pi_2)}}{(-1+3\pi_0+\pi_2)(\pi_0^2+(-1+\pi_2)^2-2\pi_0(1+\pi_2))} \in (1,\infty). \end{aligned} \tag{25}$$

6.1.3. Additive model. (See Figure 3 p. 26.)

$$\pi_0 = \frac{(f_0-f_0p+f_2p)^2}{4f_0^2-7f_0^2p+6f_0f_2p+f_2^2p+3f_0^2p^2-6f_0f_2p^2+3f_2^2p^2},$$
$$\pi_1 = \frac{1}{2},$$
$$\pi_2 = \frac{2f_0^2-3f_0^2p+2f_0f_2p+f_2^2p+f_0^2p^2-2f_0f_2p^2+f_2^2p^2}{2(4f_0^2-7f_0^2p+6f_0f_2p+f_2^2p+3f_0^2p^2-6f_0f_2p^2+3f_2^2p^2)}.$$

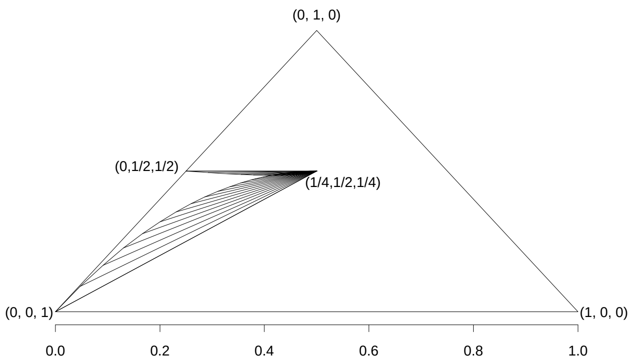

FIG. 3. *ASP quasi-recessive and quasi-dominant IBD probabilities. Strict-recessive model: the IBD probabilities lie on the Hardy-Weinberg curve joining (0,0,1) to* $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. *Quasi-recessive model: the lines under the Hardy-Weinberg curve are the IBD probabilities for fixed p and varying r. For fixed p, as r increases from 1 to* ∞, *the IBD probabilities move along a line from* $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ *to a point on the Hardy-Weinberg curve. Strict-dominant model: the IBD probabilities are on the curve joining* $(0, \frac{1}{2}, \frac{1}{2})$ *to* $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. *Quasi-dominant model: the lines above the strict-dominant curve are the IBD probabilities for fixed p and varying r.*

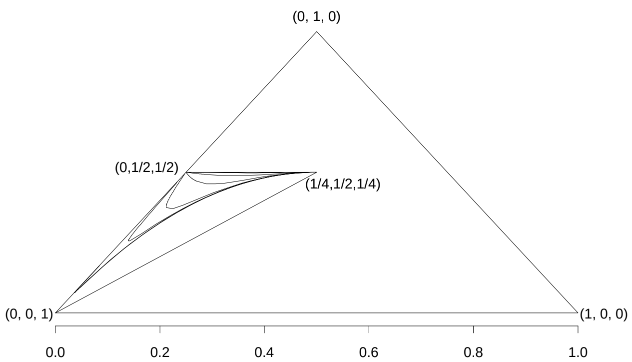

FIG. 4. *ASP intermediate IBD probabilities. From bottom to top, the curves are for fixed* $s = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4$ *and varying* $0 < p < 1$.

These probabilities lie on the line $\pi_1 = \frac{1}{2}$ joining the points $(0, \frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

6.1.4. Strict-dominant model. (See Figure 3 p. 26.)

For $0 < p \leq 1$,

$$(\pi_0, \pi_1, \pi_2) = \left(\frac{p(2-p)^2}{p^3 - 6p^2 + 5p + 4}, \frac{2(-p^2 + p + 1)}{p^3 - 6p^2 + 5p + 4}, \frac{2-p}{p^3 - 6p^2 + 5p + 4}\right).$$

The strict-dominant probabilities lie on a curve indexed by p and joining $(0, \frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Note that these IBD probabilities are independent of f_2 and are very close to the line $\pi_1 = \frac{1}{2}$. The strict-dominant and strict-recessive curves intersect at the point $(1, 2\sqrt{2}, 2)/(1 + \sqrt{2})^2$, which corresponds to the IBD probabilities for a strict-recessive model with $p = \sqrt{2}/2$ and a strict-dominant model with $p = 1 - \sqrt{2}/2$ (Louis *et al.* [23]).

6.1.5. Quasi-dominant model. (See Figure 3 p. 26.)

$$\begin{aligned}
\pi_0 &= \frac{(p^2(2-p)^2)r^2 + (2p(1-p)^2(2-p))r + (1-p)^4}{Den}, \\
\pi_1 &= \frac{2p(1+p-p^2)r^2 + 4p(1-2p+p^2)r + 2(1-p)^3}{Den}, \\
\pi_2 &= \frac{p(2-p)r^2 + (1-p)^2}{Den},
\end{aligned}$$

where

$$Den = p(4+5p-6p^2+p^3)r^2 + 2p(4-9p+6p^2-p^3)r + 4 - 12p + 13p^2 - 6p^3 + p^4.$$

Again, f_0 cancels out, and as $r \to \infty$ we get the probabilities for the strictly dominant case. We proved that for fixed $p \in (0, 1)$, the IBD probabilities lie on a line going through $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, and given by

$$\begin{aligned}
\pi_0 &= \frac{1}{4-p}((2-p) + (3p-4)\pi_2), \\
\pi_1 &= \frac{2}{4-p}(1 - p\pi_2).
\end{aligned}$$

Hence, the trinomial probabilities may be re-parameterized as

$$\begin{aligned}
t \longrightarrow t & \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) \\
+ \ (1-t) & \left(\frac{p(2-p)^2}{p^3 - 6p^2 + 5p + 4}, \frac{2(-p^2 + p + 1)}{p^3 - 6p^2 + 5p + 4}, \frac{2-p}{p^3 - 6p^2 + 5p + 4}\right),
\end{aligned}$$

$0 < t \leq 1$, where $t \to 0$ yields the strict-dominant case, and $t = 1$ corresponds to the case of no allele influencing DS at the candidate locus.

Quasi-dominant probabilities are very close to the additive probabilities, i.e. to the line $\pi_1 = \frac{1}{2}$. Also, there is a small overlap between the IBD probabilities of quasi-recessive and quasi-dominant models, and a large region of the ASP triangle is not covered by either model.

6.1.6. Intermediate model. (See Figures 4 and 5 pp. 26, 28.)

$$\pi_0 = \frac{p\,(-2\,s + p\,(-1 + 2\,s))^2}{p + p^3\,(1 - 2\,s)^2 + 4\,p\,s + 4\,s^2 + p^2\,(2 - 8\,s^2)},$$
$$\pi_1 = \frac{2\,\left(p^2\,(1 - 2\,s) - p\,(-2 + s)\,s + s^2\right)}{p + p^3\,(1 - 2\,s)^2 + 4\,p\,s + 4\,s^2 + p^2\,(2 - 8\,s^2)},$$
$$\pi_2 = \frac{p + 2\,s^2 - 2\,p\,s^2}{p + p^3\,(1 - 2\,s)^2 + 4\,p\,s + 4\,s^2 + p^2\,(2 - 8\,s^2)}.$$

Special cases of the intermediate model include:

- $s = 0$: strict-recessive model;
- $s = \frac{1}{2}$: additive model with $f_0 = 0$
$$(\pi_0, \pi_1, \pi_2) = \left(\frac{p}{1+3p}, \frac{1}{2}, \frac{1+p}{2+6p}\right);$$
- $s = 1$: strict-dominant model.

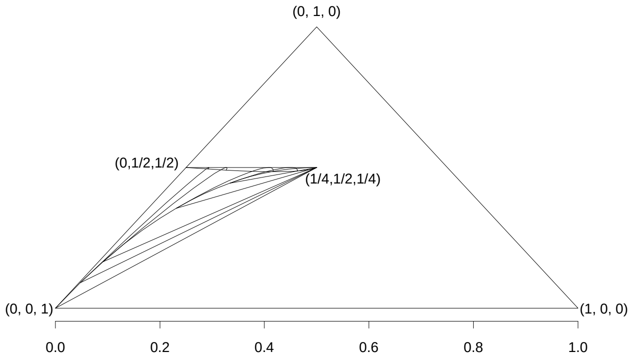

FIG. 5. *ASP quasi-recessive and intermediate IBD probabilities. Quasi-recessive model: as in Figure 3. Intermediate model: the curves joining the Hardy-Weinberg curve to the additive line are the IBD probabilities for intermediate models with fixed p and $0 \le s \le \frac{1}{2}$, while the curves joining the additive line to the strict-dominant curve are the IBD probabilities for intermediate models with fixed p and $\frac{1}{2} \le s \le 1$. From left to right $p = 0.05, 0.1, 0.3, 0.5$. These two models cover the interior of the ASP possible triangle.*

For fixed p, as s increases from 0 to $\frac{1}{2}$, the IBD probabilities trace a curve from the Hardy-Weinberg curve to the line $\pi_1 = \frac{1}{2}$, and as s increases from

$\frac{1}{2}$ to 1, the IBD probabilities trace a curve from the line $\pi_1 = \frac{1}{2}$ to the strict-dominant curve. For fixed $s \neq 0$, as p increases from 0 to 1, the IBD probabilities trace a curve from $(0, \frac{1}{2}, \frac{1}{2})$ to $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

The formulae for DSP and USP IBD probabilities are more complex since they involve one more parameter than the ASP probabilities. However, we proved that for a fixed disease allele frequency p, $0 < p < 1$, the quasi-recessive IBD probabilities for ASPs, DSPs and USPs fall on the *same line* going through $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Similarly for the quasi-dominant probabilities. For an additive model, $\pi_1 = \frac{1}{2}$, and this also holds with arbitrary mating type frequencies, but in the latter case we may have $\pi_{10} \neq \pi_{01}$.

6.2. Coverage of the ASP possible triangle. The affected sib-pair possible triangle is covered by the IBD probabilities of two types of models, one with fixed mode of inheritance and general mating type frequencies, the other with varying mode of inheritance and random mating. In general, a point in the possible triangle corresponds to the IBD probabilities for many different modes of inheritance, thus it is inappropriate to estimate the IBD probabilities and solve for parameters such as penetrances and allele frequencies, unless one has knowledge of the mode of inheritance.

6.2.1. Random mating and Hardy-Weinberg equilibrium models.

PROPOSITION 6. *The interior of the ASP possible triangle is covered by the ASP IBD probabilities under quasi-recessive and intermediate models with random mating and Hardy-Weinberg equilibrium.*

Proof. In Section 6.1.2 we proved that each point strictly between the Hardy-Weinberg curve and the line $\pi_1 = 2\pi_0$ ($\pi_1^2 < 4\pi_0\pi_2$, $\pi_1 > 2\pi_0$) corresponds to the ASP IBD probabilities for a unique (up to f_0) quasi-recessive model with parameters (r, p) given in equations (24) and (25). Each point strictly between the Hardy-Weinberg curve and the additive line ($\pi_1^2 > 4\pi_0\pi_2$, $\pi_1 < \frac{1}{2}$, $\pi_0 > 0$) corresponds to the ASP IBD probabilities for a unique (up to f_2) intermediate model with $0 < p < 1$ and $0 < s < \frac{1}{2}$. This follows from the Inverse Function Theorem applied to the function $(\pi_0(s,p), \pi_1(s,p))$ defined on the open set $(0, \frac{1}{2}) \times (0, 1)$. Here, $\pi_i(s,p)$ is the probability that an ASP shares DNA IBD at the DS locus on $i = 0, 1$ chromosomes under a random mating intermediate model with parameters (s, p) and fixed f_2. The Jacobian matrix for this function is

$$J(s,p) = \begin{bmatrix} \frac{\partial \pi_0}{\partial s} & \frac{\partial \pi_0}{\partial p} \\ \frac{\partial \pi_1}{\partial s} & \frac{\partial \pi_1}{\partial p} \end{bmatrix}$$

and the determinant of the Jacobian matrix is

$$\begin{aligned} |J(s,p)| &= \frac{8\,(-1+p)^2\,p\,s\,(-1+2\,s)\,(-p-2\,s+2\,p\,s)\,(-p-s+2\,p\,s)}{\left(p+p^3\,(1-2\,s)^2+4\,p\,s+4\,s^2+p^2\,(2-8\,s^2)\right)^3} \\ &\neq 0 \qquad \text{for } (s,p) \in (0, \tfrac{1}{2}) \times (0,1). \end{aligned}$$

Numerical solutions for (s, p) may be obtained by the Newton-Raphson method (Louis *et al.* [23]). Finally, the Hardy-Weinberg curve is covered by ASP strict-recessive IBD probabilities. □

6.2.2. Strict-recessive model.

PROPOSITION 7. *The ASP possible triangle is covered by the ASP IBD probabilities for a strict-recessive model with general mating type frequencies. That is, for each* (π_0, π_1, π_2) *with* $\pi_1 \leq \frac{1}{2}$, $\pi_0 \geq 0$ *and* $\pi_1 \geq 2\pi_0$, *we can find a family of mating type frequencies* $(P_1, \ldots, P_9)$ *such that* (π_0, π_1, π_2) *are the ASP IBD probabilities for the strict-recessive model with mating type frequencies* $(P_1, \ldots, P_9)$.

This illustrates the potentially large impact of non-random mating on the IBD probabilities.

Proof. The ASP strict-recessive IBD probabilities are independent of f_2, so, without loss of generality, we let $f_2 = 1$. These probabilities may be computed using Table 10 and are given by

$$\begin{aligned} \pi_0 &= \frac{P_1}{4P_1 + P_2 + P_3 + P_6/4}, \\ \pi_1 &= \frac{2P_1 + P_2/2 + P_3/2}{4P_1 + P_2 + P_3 + P_6/4}, \\ \pi_2 &= \frac{P_1 + P_2/2 + P_3/2 + P_6/4}{4P_1 + P_2 + P_3 + P_6/4}. \end{aligned} \tag{26}$$

Letting

$$\begin{aligned} P_1 &= c\pi_0, \\ P_2 = P_3 &= c(\pi_1 - 2\pi_0), \\ P_6 &= 4c(\pi_0 + \pi_2 - \pi_1), \end{aligned}$$

where c is a positive constant such that $P_1 + P_2 + P_3 + P_6 \leq 1$, yields (π_0, π_1, π_2). In particular, the boundaries of the ASP triangle are covered by the following models:

- $\pi_0 = 0$: Set $P_1 = 0$ and vary P_6 from 0 to 1;
- $\pi_1 = \frac{1}{2}$: Set $P_6 = 0$ and vary P_1 from 0 to 1;
- $\pi_1 = 2\pi_0$: Set $P_2 = P_3 = 0$ and vary P_1 from 0 to 1.

□

Note that for DSPs $\pi_2 = 0$ and hence the strict-recessive model doesn't cover the DSP possible triangle.

6.3. Some counterexamples. The following are two counterexamples demonstrating the necessity of the assumptions of our model for the triangle constraints, and examples of models for which $\pi_{01} \neq \pi_{10}$. Either environmental covariance in the sib phenotypes or over-dominance may lead to IBD probabilities that fall outside the ASP possible triangle.

TABLE 10

Conditional probability of ASP given inheritance vector x and parental genotype pg for a strict-recessive model, under `Assumption G1`. *For parental genotypes not listed in the table $pr(ASP|x,pg) = 0$.*

$$pr(ASP|x,pg)$$

Mating type mt	Parental genotype pg	Inheritance vector x			
		(1,3,1,3)	(1,3,1,4)	(1,3,2,3)	(1,3,2,4)
1	$DD \times DD$	1	1	1	1
2	$DD \times Dd$	1	0	1	0
	$DD \times dD$	0	0	0	0
3	$Dd \times DD$	1	1	0	0
	$dD \times DD$	0	0	0	0
	$Dd \times Dd$	1	0	0	0
6	$Dd \times dD$	0	0	0	0
	$dD \times Dd$	0	0	0	0
	$dD \times dD$	0	0	0	0

6.3.1. `Assumption M1` **- Monotonicity of penetrances.** Under random mating and Hardy-Weinberg equilibrium, the triangle constraints hold regardless of the penetrances. However, `Assumption M1` is necessary for the triangle constraints when we allow arbitrary mating type frequencies. An extreme example is that of a diallelic model with over-dominance, $f_0 = f_2 = 0, f_1 = 1$, and mating type frequency $P_6 = 1$. Then, for ASPs, $(\pi_0, \pi_1, \pi_2) = (\frac{1}{2}, 0, \frac{1}{2})$.

6.3.2. `Assumption G1` **- Conditional independence of phenotypes given genotypes.** Consider a genetic model with random mating and Hardy-Weinberg equilibrium at the DS locus, and the following extreme form of environmental covariance:

$$pr(\text{Both sibs are affected} \mid \text{At least one sib has genotype } DD) = 1,$$

$$pr(\text{Both sibs are affected} \mid \text{Neither of the sibs has genotype } DD) = 0.$$

For this model, the ASP IBD probabilities for $p \neq 0$ are given by

$$\pi_0 = \frac{p^2 + 4p(1-p) + 2(1-p)^2}{4p^2 + 12p(1-p) + 7(1-p)^2},$$
$$\pi_1 = \frac{2p^2 + 6p(1-p) + 4(1-p)^2}{4p^2 + 12p(1-p) + 7(1-p)^2},$$
$$\pi_2 = \frac{p^2 + 2p(1-p) + (1-p)^2}{4p^2 + 12p(1-p) + 7(1-p)^2}.$$

Hence, $2\pi_0 > \pi_1$ and $\pi_1 > \frac{1}{2}$ for $p \neq 1$, and `Assumption G1` is necessary for the triangle constraints.

Proof. The result follows from Tables 11 and 12, and equation (1).

TABLE 11

Conditional probability of ASP given inheritance vector x and parental genotype pg for a model which violates `Assumption G1`*. For parental genotypes not listed in the table $pr(ASP|x,pg) = 0$.*

$pr(ASP|x, pg)$

Parental genotype	Inheritance vector x			
pg	(1,3,1,3)	(1,3,1,4)	(1,3,2,3)	(1,3,2,4)
$DD \times DD$	1	1	1	1
$DD \times Dd$	1	1	1	1
$DD \times dD$	0	1	0	1
$Dd \times DD$	1	1	1	1
$dD \times DD$	0	0	1	1
$Dd \times Dd$	1	1	1	1
$Dd \times dD$	0	1	0	0
$dD \times Dd$	0	0	1	0
$dD \times dD$	0	0	0	1

TABLE 12

x	$\sum_{pg} pr(ASP\|x, pg)pr(pg)$
(1,3,1,3)	$p^4 + 2p^3q + p^2q^2$
(1,3,1,4)	$p^4 + 3p^3q + 2p^2q^2$
(1,3,2,3)	$p^4 + 3p^3q + 2p^2q^2$
(1,3,2,4)	$p^4 + 4p^3q + 2p^2q^2$

$\square$

6.3.3. Examples when $\pi_{01} \neq \pi_{10}$. Either parental imprinting or non-symmetry of parental mating type frequencies with respect to maternal and paternal genotypes may result in π_{01} being different from π_{10}.

First, consider a single diallelic DS locus with random mating and Hardy-Weinberg equilibrium, and assume that $f_{00} = f_{01} = 0$ and $f_{10} = f_{11} = 1$. Then, for ASPs,

$$\sum_{pg} pr(ASP|(1,3,1,4), pg)\, pr(pg) = p^4 + 3p^3q + 3p^2q^2 + pq^3,$$
$$\sum_{pg} pr(ASP|(1,3,2,3), pg)\, pr(pg) = p^4 + 2p^3q + p^2q^2,$$

and hence, by equation (1), $\pi_{01} \neq \pi_{10}$.

Consider now a strict-recessive model with general mating type frequencies and $P_2 \neq P_3$. From the derivation of equation (26) for ASPs

$$\pi_{10} = \frac{P_1 + P_3/2}{4P_1 + P_2 + P_3 + P_6/4} \neq \frac{P_1 + P_2/2}{4P_1 + P_2 + P_3 + P_6/4} = \pi_{01}.$$

Hence, in some cases, distinguishing between maternal and paternal sharing may lead to more powerful tests of linkage.

6.4. Non-random mating and Hardy-Weinberg disequilibrium models. In order to investigate the impact of non-random mating and Hardy-Weinberg disequilibrium on the IBD probabilities, we consider the model used by Jin *et al.* [15] for the mating type frequencies at a single m-allele DS locus. Let $p_{ij,kl}$ denote the probability of a $D_iD_j \times D_kD_l$ mating, p_{ij} the probability of genotype D_iD_j, and p_i the allele frequency of D_i. The mating type frequencies are mixtures of the frequencies under random mating and complete dependence, with the same margins. They are given by

$$p_{ij,kl} = \begin{cases} (1-\delta_R)p_{ij}^2 + \delta_R p_{ij}, & \text{if } i = k \text{ and } j = l, \\ (1-\delta_R)p_{ij}p_{kl}, & \text{otherwise,} \end{cases}$$

where

$$p_{ij} = \begin{cases} p_i^2 + \delta_{HW} p_i(1-p_i), & \text{if } i = j, \\ 2p_ip_j(1-\delta_{HW}), & \text{if } i \neq j, \end{cases}$$

and $i \leq j$, $k \leq l$. Here, δ_{HW} is a parameter representing deviation from Hardy-Weinberg equilibrium, while δ_R represents deviation from random mating. Positive δ_{HW} correspond to a deficiency of heterozygotes compared to Hardy-Weinberg frequencies, and similarly, positive δ_R correspond to a deficiency of different genotype matings compared to random mating frequencies. Louis *et al.* [22] used this model with $\delta_{HW} = 0$ to study the impact of positive assortative mating on ASP IBD probabilities and Weir [37] used this model with $\delta_R = 0$ as a one-parameter class of alternatives to Hardy-Weinberg equilibrium. For the usual two-allele model, with alleles D and d and allele frequencies p and $q = 1 - p$, respectively, the mating type frequencies are given in Table 13. The parameters (δ_R, δ_{HW}) are constrained to yield non-negative genotype and mating type frequencies, in particular, $\delta_R \leq 1$ and $\delta_{HW} \leq 1$.

PROPOSITION 8. Impact of non-random mating on ASP strict-recessive curve.
For a strict-recessive model ($f_0 = f_1 = 0, f_2 = 1$), the ASP IBD probabilities satisfy the following:

- *If $\delta_R = 1$, then $\pi_1 = 2\pi_0$.*

TABLE 13
Mating type frequencies for non-random mating model.

Mating type mt	Mating type frequency P_{mt}
1	$(1-\delta_R)(p^2+\delta_{HW}pq)^2+\delta_R(p^2+\delta_{HW}pq)$
2	$(1-\delta_R)(p^2+\delta_{HW}pq)(2pq(1-\delta_{HW}))$
3	$(1-\delta_R)(p^2+\delta_{HW}pq)(2pq(1-\delta_{HW}))$
4	$(1-\delta_R)(p^2+\delta_{HW}pq)(q^2+\delta_{HW}pq)$
5	$(1-\delta_R)(p^2+\delta_{HW}pq)(q^2+\delta_{HW}pq)$
6	$(1-\delta_R)(2pq(1-\delta_{HW}))^2+\delta_R(2pq(1-\delta_{HW}))$
7	$(1-\delta_R)(2pq(1-\delta_{HW}))(q^2+\delta_{HW}pq)$
8	$(1-\delta_R)(2pq(1-\delta_{HW}))(q^2+\delta_{HW}pq)$
9	$(1-\delta_R)(q^2+\delta_{HW}pq)^2+\delta_R(q^2+\delta_{HW}pq)$

- *When $\delta_R = 0$, i.e. mating is random, then regardless of δ_{HW}, the ASP strict-recessive IBD probabilities lie on the Hardy-Weinberg curve $\pi_1^2 = 4\pi_0\pi_2$.*
- *When $0 < \delta_R \leq 1$, the ASP strict-recessive IBD probabilities are between the Hardy-Weinberg curve and the line $\pi_1 = 2\pi_0$, i.e. $\pi_1^2 \leq 4\pi_0\pi_2$ and $\pi_1 \geq 2\pi_0$. Also, if $0 \leq \delta_{HW} \leq 1$*

$$\lim_{p\to 1}(\pi_0,\pi_1,\pi_2) = \left(\frac{1}{4},\frac{1}{2},\frac{1}{4}\right),$$
$$\lim_{p\to 0, p\neq 0}(\pi_0,\pi_1,\pi_2) = \frac{1}{\frac{7}{2}\delta_{HW}+\frac{1}{2}}\left(\delta_{HW}, 2\delta_{HW}, \frac{1}{2}(\delta_{HW}+1)\right),$$

  hence the limit of the IBD probabilities as $p \to 0$ is on the line $\pi_1 = 2\pi_0$ and is independent of δ_R.
- *When $\delta_R < 0$, then regardless of δ_{HW}, the ASP strict-recessive IBD probabilities are between the Hardy-Weinberg curve and the line $\pi_1 = \frac{1}{2}$, i.e. $\pi_1^2 \geq 4\pi_0\pi_2$ and $\pi_1 \leq \frac{1}{2}$.*

These observations are illustrated in Figures 6 and 7 p. 36.

Proof. For the strict-recessive model, the ASP probabilities are given by equation (26). Since $P_3 = P_2$, then

$$\begin{aligned} 4\pi_0\pi_2 - \pi_1^2 &= \frac{P_1P_6 - P_2^2}{(4P_1+2P_2+P_6/4)^2} \\ &= \frac{\delta_R p_{DD} p_{Dd}(1-p_{dd}+p_{dd}\delta_R)}{(4P_1+2P_2+P_6/4)^2}, \end{aligned}$$

and we have the following cases:
- $\delta_R = 0$: $\pi_1^2 = 4\pi_0\pi_2$.
- $0 < \delta_R \leq 1$: $\pi_1^2 \leq 4\pi_0\pi_2$.

- $\delta_R < 0$: Since $P_1 \geq 0$, then $\delta_R \geq \frac{-p_{DD}}{1-p_{DD}}$ and hence $1 - p_{dd} + p_{dd}\delta_R \geq \frac{1-p_{DD}-p_{dd}}{1-p_{DD}} \geq 0$. It follows that $\pi_1^2 \geq 4\pi_0\pi_2$.
- $\delta_R = 1$: $P_2 = P_3 = 0$ and it follows that $\pi_1 = 2\pi_0$.

$\square$

7. Multilocus conditional distribution of inheritance vectors at marker loci linked to the DS loci given phenotype vector of sibship. Consider L unlinked markers, $\mathcal{M}^1, \ldots, \mathcal{M}^L$, where $\mathcal{M}^l$ is linked to $\mathcal{D}^l$ at recombination fraction θ_l, $l = 1, \ldots, L$. Let y^l denote the inheritance vector of the sibship at $\mathcal{M}^l$, $l = 1, \ldots, L$, and let $y = (y^1, \ldots, y^L)$. We wish to compute $pr(y|\phi)$, the conditional distribution of the inheritance vectors at the marker loci given the phenotype vector of the sibship. This distribution is obtained by conditioning on all possible recombination patterns in the sibship between the markers and the DS loci. Since the inheritance vectors at the marker loci are conditionally independent of the phenotype vector given the inheritance vectors at the DS loci, then

$$pr(y|\phi) = \sum_x pr(y, x|\phi) = \sum_x pr(y|x, \phi)\ pr(x|\phi) = \sum_x pr(y|x)\ pr(x|\phi).$$

Now, the number of coordinates at which x^l and y^l differ, $\Delta(x^l, y^l) = \sum_{i=1}^{2k}(1 - \delta(x_i^l, y_i^l))$, is the total number of recombinants between $\mathcal{D}^l$ and $\mathcal{M}^l$. The chance that a coordinate differs between x^l and y^l is the chance of a recombination between $\mathcal{D}^l$ and $\mathcal{M}^l$, i.e. the recombination fraction θ_l. Since recombination events are independent for unlinked loci and across meioses, then

$$\begin{aligned} pr(y|\phi) &= \sum_x \left\{ \prod_{l=1}^{L} pr(y^l|x^l) \right\} pr(x|\phi) \\ &= \sum_x \left\{ \prod_{l=1}^{L} \theta_l^{\Delta(x^l,y^l)} (1-\theta_l)^{2k-\Delta(x^l,y^l)} \right\} pr(x|\phi). \end{aligned}$$

Hence, the conditional distribution $pr(y|\phi)$ of inheritance vectors at markers linked to DS loci in the manner described above may be obtained from the conditional distribution $pr(x|\phi)$ of the inheritance vectors at the DS loci by means of the transition matrix

$$T(\theta_1, \ldots, \theta_L) = T(\theta_1) \otimes \ldots \otimes T(\theta_L).$$

$T(\theta_l)$ is the Kronecker power of the 2×2 transition matrices corresponding to transitions in each of the $2k$ coordinates between $\mathcal{D}^l$ and $\mathcal{M}^l$

$$T(\theta_l) = \begin{bmatrix} 1-\theta_l & \theta_l \\ \theta_l & 1-\theta_l \end{bmatrix}^{\otimes 2k}.$$

This matrix representation separates the contributions of the genetic model for disease susceptibility ($pr(x|\phi)$) and of linkage (θ's). Note that

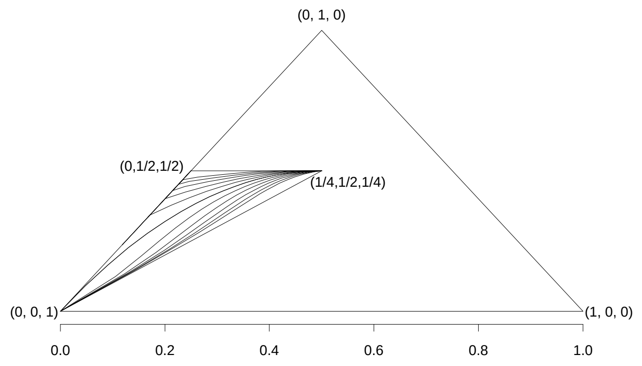

FIG. 6. *ASP strict-recessive curves for non-random mating model with* $\delta_{HW} = 0$ *and* $\delta_R = -0.5, -0.4, \ldots, 0.5$ *(from top to bottom). For fixed* δ_R*, as* $p \to 1$*, the ASP IBD probabilities move along a curve toward* $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

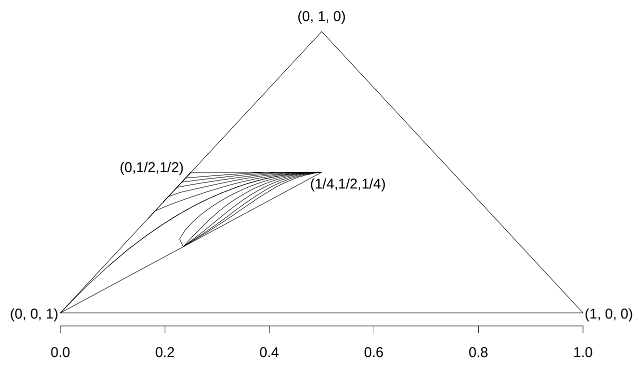

FIG. 7. *ASP strict-recessive curves for non-random mating model with* $\delta_{HW} = 0.1$ *and* $\delta_R = -0.5, -0.4, \ldots, 0.5$ *(from top to bottom). For fixed* δ_R*, as* $p \to 1$*, the ASP IBD probabilities move along a curve toward* $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

$T(0)$ is the $2^{2k} \times 2^{2k}$ identity matrix and the entries of $T(\frac{1}{2})$ are all equal to $\frac{1}{2^{2k}}$.

Let us now consider the simple case of sib-pairs and a single marker $\mathcal{M}$ linked to a DS locus $\mathcal{D}$ at recombination fraction θ ($\mathcal{D}$ could be one of several unlinked DS loci). For ASPs and $i = 0, 1, 2$, let

$$\pi_i^{ASP}(\nu) = pr(\text{Sib-pair shares DNA IBD on } i \text{ chromosomes at } \mathcal{D}|ASP)$$

and let

$$\rho_i^{ASP}(\theta,\nu) = pr(\text{Sib-pair shares DNA IBD on } i \text{ chromosomes at } \mathcal{M}|ASP).$$

The IBD probabilities π_{ij} and ρ_{ij}, i, $j = 0, 1$, distinguishing between sharing of maternal and paternal DNA, are defined at $\mathcal{D}$ and $\mathcal{M}$ as in Section 4. The same notation is used for DSPs and USPs. By Proposition 5, the π's satisfy the triangle constraints.

It may be shown that for each type of sib-pair

$$\begin{bmatrix} \rho_{00}(\theta,\nu) \\ \rho_{01}(\theta,\nu) \\ \rho_{10}(\theta,\nu) \\ \rho_{11}(\theta,\nu) \end{bmatrix} = \begin{bmatrix} \psi^2 & \psi\bar{\psi} & \psi\bar{\psi} & \bar{\psi}^2 \\ \psi\bar{\psi} & \psi^2 & \bar{\psi}^2 & \psi\bar{\psi} \\ \psi\bar{\psi} & \bar{\psi}^2 & \psi^2 & \psi\bar{\psi} \\ \bar{\psi}^2 & \psi\bar{\psi} & \psi\bar{\psi} & \psi^2 \end{bmatrix} \times \begin{bmatrix} \pi_{00}(\nu) \\ \pi_{01}(\nu) \\ \pi_{10}(\nu) \\ \pi_{11}(\nu) \end{bmatrix}, \tag{27}$$

where $\psi = \theta^2 + (1-\theta)^2$ and $\bar{\psi} = 1 - \psi = 2\theta\bar{\theta}$. When we do not distinguish between maternal and paternal sharing, the transition matrix $T(\theta)$ collapses into a 3×3 matrix

$$\begin{bmatrix} \rho_0(\theta,\nu) \\ \rho_1(\theta,\nu) \\ \rho_2(\theta,\nu) \end{bmatrix} = \begin{bmatrix} \psi^2 & \psi\bar{\psi} & \bar{\psi}^2 \\ 2\psi\bar{\psi} & \psi^2+\bar{\psi}^2 & 2\psi\bar{\psi} \\ \bar{\psi}^2 & \psi\bar{\psi} & \psi^2 \end{bmatrix} \times \begin{bmatrix} \pi_0(\nu) \\ \pi_1(\nu) \\ \pi_2(\nu) \end{bmatrix}. \tag{28}$$

This 3×3 transition matrix is given in Haseman and Elston [12] and Suarez *et al.* [35].

PROPOSITION 9. *Under our general multilocus model with arbitrary mating type frequencies and* `Assumptions G1,G2, M2`

- $(\rho_0^{ASP}(\theta,\nu), \rho_1^{ASP}(\theta,\nu), \rho_2^{ASP}(\theta,\nu))$, *the ASP IBD probabilities at a marker linked to the DS locus $\mathcal{D}$ at a recombination fraction θ, fall in a triangle with vertices*

$$\left(\frac{1}{4},\frac{1}{2},\frac{1}{4}\right), \quad (\bar{\psi}^2, 2\psi\bar{\psi}, \psi^2) \quad \text{and} \quad \frac{1}{2}\left(\bar{\psi}, 1, \psi\right)$$

  where $\psi = \theta^2 + (1-\theta)^2$;
- *this triangle is contained in the ASP possible triangle ($\theta = 0$);*
- *as $\theta \longrightarrow \frac{1}{2}$ the triangles shrink toward $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ along the line $\rho_1 = \frac{1}{2}$.*

Similarly,

- $(\rho_0^{DSP}(\theta,\nu), \rho_1^{DSP}(\theta,\nu), \rho_2^{DSP}(\theta,\nu))$, *the DSP IBD probabilities at a marker linked to the DS locus $\mathcal{D}$ at a recombination fraction θ, fall in a triangle with vertices*

$$\left(\frac{1}{4},\frac{1}{2},\frac{1}{4}\right),\quad \frac{1}{2}(\psi,1,\bar{\psi}) \ \text{ and } \ \frac{1}{3}(1-\bar{\psi}^2, 2(\psi\bar{\psi}+\psi^2+\bar{\psi}^2), 1-\psi^2);$$

- *this triangle is contained in the DSP possible triangle ($\theta = 0$);*
- *as $\theta \rightarrow \frac{1}{2}$ the triangles shrink toward $(\frac{1}{4},\frac{1}{2},\frac{1}{4})$ along the line $\rho_1 = \frac{1}{2}$.*

Again, USP IBD probabilities satisfy the same constraints as ASP IBD probabilities.

Proof. For ASPs (USPs), the proof relies on $\frac{1}{2} \leq \psi \leq 1$, on equations (7) and (8), and on the relationship between IBD probabilities at the marker and at the DS locus

$$\begin{bmatrix} \rho_0 \\ \rho_1 \\ \rho_2 \end{bmatrix} = \pi_0 \begin{bmatrix} \psi^2 \\ 2\psi\bar{\psi} \\ \bar{\psi}^2 \end{bmatrix} + \pi_1 \begin{bmatrix} \psi\bar{\psi} \\ \psi^2+\bar{\psi}^2 \\ \psi\bar{\psi} \end{bmatrix} + \pi_2 \begin{bmatrix} \bar{\psi}^2 \\ 2\psi\bar{\psi} \\ \psi^2 \end{bmatrix}, \tag{29}$$

whereby (ρ_0, ρ_1, ρ_2) is a convex combination of 3 trinomial probability vectors contained in the triangle with vertices (1,0,0), (0,1,0), (0,0,1). The proof is similar for DSPs. ◻

The possible triangles are shown in Figure 8 for various values of the recombination fraction θ. Figure 9 shows the impact of recombination on the IBD probabilities for four models. The following can easily be shown with representation (29):

- If $\pi_1^2 = 4\pi_0\pi_2$, then $\rho_1^2 = 4\rho_0\rho_2$. Hence, strict-recessive random mating ASP IBD probabilities at the marker $\mathcal{M}$ remain on the Hardy-Weinberg curve.
- If $\pi_1 = \frac{1}{2}$, then $\rho_1 = \frac{1}{2}$. Hence additive probabilities at the marker $\mathcal{M}$ remain on the additive line.

8. Open questions. In this paper, we studied sib-pair IBD probabilities under general multilocus models for disease susceptibility. We proved the possible triangle constraints under general monotonicity assumptions concerning the penetrances, and without the problematic assumptions of random mating, Hardy-Weinberg equilibrium and linkage equilibrium. We also studied the parameterization of sib-pair IBD probabilities for common genetic models and showed that the ASP possible triangle was covered by the IBD probabilities of at least two types of models. Finally, we illustrated the potentially large impact of non-random mating on IBD probabilities.

Several open questions regarding IBD probabilities come to mind. Firstly, the problem of *linked* DS loci remains to be addressed. An obvious complication in the derivation of constraints for this type of models arises since $pr(x|pg)$ now involves recombination fractions between the linked DS loci and hence does not cancel out of equation (1).

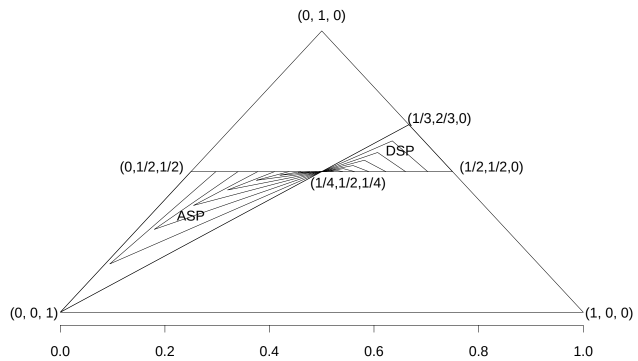

FIG. 8. *ASP and DSP possible triangles for IBD probabilities at a marker θ away from a DS locus, $\theta = 0, 0.05, \ldots, 0.5$.*

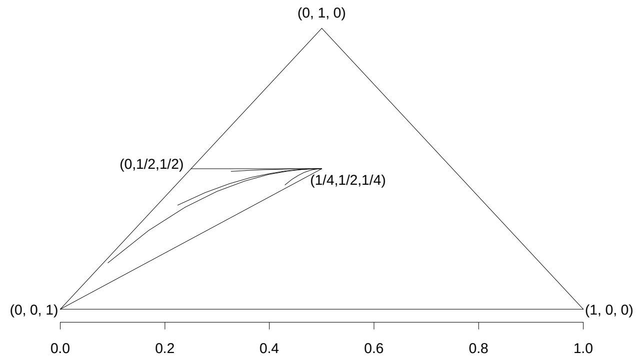

FIG. 9. *Curves traced by ASP IBD probabilities at a marker, as recombination fraction θ between the marker and a DS locus varies between 0 and $\frac{1}{2}$. The 4 models considered involve a single diallelic random mating DS locus. Starting from the top curve, the models are: strict-dominant with $p = 0.1$, intermediate with $s = 0.1$, $p = 0.1$, strict-recessive with $p = 0.1$, and quasi-recessive with $r = 10$, $p = 0.1$.*

This invalidates Proposition 1 for linked loci. A second open problem is the coverage of the space for *multilocus* IBD probabilities. For two loci, for example, what region of (3×3)-1$-$dimensional space is covered by π_{i_1,i_2}, the probability that an ASP shares DNA IBD on i_1 and i_2 chromosomes at locus 1 and 2, respectively? A third question is the generalization of the constraints to larger sibships with both affected and unaffected individuals. We have already derived constraints for affected sib-trio IBD probabilities, however generalizing the constraints to arbitrary sibships is challenging.

We used the matrix representation of the IBD probabilities at markers linked to DS loci (transition matrix $T(\theta)$) to derive a score test of the null hypothesis of no linkage between a marker and a DS locus for sibships of various sizes and phenotype patterns [7]. This test is locally most powerful in the recombination fraction θ between the marker and a DS locus, and provides optimal weights for combining the test statistics across the different types of sibships. For affected only sibships, the score statistic is the usual statistic S_{pairs}, obtained by forming all possible pairs of affected sibs and averaging the proportions of chromosomes on which they share DNA IBD. The matrix representation and score tests may be extended to other types of pedigrees.

Acknowledgments. We would like to thank David Nelson for kindly providing the program for barycentric plots as well as the simplex figure. We are also grateful to participants at the IMA meeting at which a version of these results were presented. Their lively discussion spurred us to further develop some of our ideas on age and sex-dependent penetrances. Finally, we wish to thank the referee for helpful comments. This work was supported in part by a PMMB Burroughs-Wellcome Fellowship, and by the NIH through grant RO1 HG01093.

REFERENCES

[1] SCHIZOPHRENIA COLLABORATIVE LINKAGE GROUP (CHROMOSOME 22), A combined analysis of D22S278 marker alleles in affected sib-pairs: support for a susceptibility locus for schizophrenia at chromosome 22q12. *Am. J. Med. Genet.*, 67:40–45, 1996.

[2] W.C. BLACKWELDER AND R.C. ELSTON, A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet. Epidemiol.*, 2:85–97, 1985.

[3] H.J. CORDELL, J.A. TODD, S.T. BENNETT, Y. KAWAGUCHI, AND M. FARRALL, Two-locus maximum lod score analysis of a multifactorial trait: Joint consideration of IDDM2 and IDDM4 with IDDM1 in type I diabetes. *Am. J. Hum. Genet.*, 57:920–934, 1995.

[4] N.J. COX AND R.S. SPIELMAN, The insulin gene and susceptibility to IDDM. *Genet. Epidemiol.*, 6:65–69, 1989.

[5] A.G. CUDWORTH AND J.C. WOODROW, Evidence for HLA-linked genes in "juvenile" diabetes mellitus. *Br. Med. J.*, 3:133–135, 1975.

[6] N.E. DAY AND M.J. SIMONS, Disease-susceptibility genes — their identification by multiple case family studies. *Tissue Antigens*, 8:109–119, 1976.

[7] S. DUDOIT AND T.P. SPEED, A score test for linkage using identity by descent data from sibships. In preparation.

[8] S.N. Ethier and S.E. Hodge, Identity-by-descent analysis of sibship configurations. *Am. J. Med. Genet.*, 22:263–272, 1985.

[9] J.J. Faraway, Improved sib-pair linkage test for disease susceptibility loci. *Genet. Epidemiol.*, 10:225–233, 1993.

[10] E. Feingold, P.O. Brown, and D. Siegmund, Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Am. J. Hum. Genet.*, 53:234–251, 1993.

[11] E. Feingold and D. Siegmund, Strategies for mapping heterogeneous recessive traits by allele sharing methods. *Am. J. Hum. Genet.*, 60:965–978, 1997.

[12] J.K. Haseman and R.C. Elston, The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*, 2:3–19, 1972.

[13] P. Holmans, Asymptotic properties of affected-sib-pair linkage analysis. *Am. J. Hum. Genet.*, 52:362–374, 1993.

[14] P. Holmans and D. Clayton, Efficiency of typing unaffected relatives in an affected-sib-pair linkage study with single-locus and multiple tightly linked markers. *Am. J. Hum. Genet.*, 57:1221–1232, 1995.

[15] K. Jin, T.P. Speed, and G. Thomson, Tests of random mating for a highly polymorphic locus: Application to HLA data. *Biometrics*, 51:1064–1076, 1995.

[16] M. Knapp, S.A. Seuchter, and M.P. Baur, Linkage analysis in nuclear families 1: Optimality criteria for affected sib-pair tests. *Hum. Hered.*, 44:37–43, 1994.

[17] M. Knapp, S.A. Seuchter, and M.P. Baur, Linkage analysis in nuclear families 2: Relationship between affected sib-pair tests and lod score analysis. *Hum. Hered.*, 44:44–51, 1994.

[18] M. Knapp, S.A. Seuchter, and M.P. Baur, Two-locus disease models with two marker loci: The power of affected sib-pair tests. *Am. J. Hum. Genet.*, 55:1030–1041, 1994.

[19] L. Kruglyak, M.J. Daly, M.P. Reeve-Daly, and E.S. Lander, Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, 58:1347–1363, 1996.

[20] L. Kruglyak and E.S. Lander, Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.*, 57:439–454, 1995.

[21] M. Lalande, Parental imprinting and human disease. *Annu. Rev. Genet.*, 30:173–195, 1997.

[22] E.J. Louis, H. Payami, and G. Thomson, The affected sib method. V. Testing the assumptions. *Ann. Hum. Genet.*, 51:75–92, 1987.

[23] E.J. Louis, G. Thomson, and H. Payami, The affected sib method. II. The intermediate model. *Ann. Hum. Genet.*, 47:225–243, 1983.

[24] M.F. Moffat, P.A. Sharp, J.A. Faux, R.P. Young, W.O.C.M. Cookson, and J.M. Hopkin, Factors confounding genetic linkage between atopy and chromosome 11q. *Clin. Exp. Allergy*, 22:1046–1051, 1992.

[25] G. Morahan, D. Huang, B.D. Tait, P.G. Colman, and L.C. Harrison, Markers on distal chromosome 2q linked to insulin-dependent diabetes mellitus. *Science*, 272:1811–1813, 1996.

[26] U. Motro and G. Thomson, The affected sib method. I. Statistical features of the affected sib-pair method. *Genetics*, 110:525–538, 1985.

[27] N. Niikawa, Genomic imprinting and its relevance to genetics diseases. *Jpn. J. Human Genet.*, 41:351–361, 1996.

[28] H. Payami, E.J. Louis G. Thomson, U. Motro, and E. Hudes, The affected sib method. IV. Sib trios. *Ann. Hum. Genet.*, 49:303–314, 1985.

[29] H. Payami, G. Thomson, and E.J. Louis, The affected sib method. III. Selection and recombination. *Am. J. Hum. Genet.*, 36:352–362, 1984.

[30] M.A. Pericak-Vance, J.L. Bebout, P.C. Gaskell, Jr., L.H. Yamaoka, W.Y. Hung, M.J. Alberts, A.P. Walker, R.J. Bartlett, C.A. Maynes, K.A. Welsh, N.L. Earl, A. Heyman, C.M. Clark, and A.D. Roses, Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am. J. Hum. Genet.*, 48:1034–1050, 1991.

[31] N. Risch, Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.*, 46:222–228, 1990.

[32] N. Risch, Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. Hum. Genet.*, 46:229–241, 1990.

[33] N. Risch, Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am. J. Hum. Genet.*, 46:242–253, 1990.

[34] R.S. Spielman, L. Baker, and C.M. Zmijewski, Gene dosage and susceptibility to insulin-dependent diabetes. *Ann. Hum. Genet., Lond.*, 44:135–150, 1980.

[35] B.K. Suarez, J. Rice, and T. Reich, The generalized sib-pair IBD distribution: its use in the detection of linkage. *Ann. Hum. Genet., Lond.*, 42:87–94, 1978.

[36] G. Thomson and W.F. Bodmer, The genetics of HLA and disease associations. In J. Dausset and A. Svejgaard, editors, *HLA and Disease*, pages 545–563. Munksgaard, Copenhagen, 1977.

[37] B.S. Weir, *Genetic Data Analysis II.* Sinauer Associates, 1996.