Ann. Hum. Genet. (1985), **49**, 303–314
Printed in Great Britain

The affected sib method. IV. Sib trios

BY H. PAYAMI, G. THOMSON, U. MOTRO*, E. J. LOUIS AND E. HUDES
Department of Genetics, University of California, Berkeley, California 94720, *U.S.A.*

SUMMARY

The classical sib pair method uses the expected and observed HLA (human leukocyte antigen) haplotype sharing distribution in sib pairs, who are affected with an HLA associated disease, to make inferences about the inheritance of the disease. In this paper we present the expected HLA haplotype sharing distributions in affected sib trios, and sib pairs, from families with three or more affected sibs. The underlying model for both distributions, as for the classical sib pair method, is that disease predisposition is determined by a single allele at an HLA-linked locus. The sib trio tests of hypotheses (additive and recessive), and disease parameter estimates (additive, recessive and intermediate), can be compared with those obtained from the classical sib pair analysis. In addition, the sib trio data allow parameter estimation for a general disease model to be made, if the data fall within the bounds of the expectation. This study forms the basis of later investigations which show that haplotype sharing of affected sib trios for two susceptibility alleles (negative complementation) model, which appears appropriate for insulin dependent diabetes mellitus (IDDM), moves outside the bound of the single susceptibility expectations outlined here, whereas haplotype sharing values for sib pairs are bound by the single susceptibility allele expectations. Available Caucasian IDDM data have been analysed. The results support genetic heterogeneity of IDDM.

INTRODUCTION

Over forty diseases have been shown to be associated with the human leukocyte antigen (HLA) system (Dausset & Svejgaard, 1977; Bodmer, 1978; Ryder *et al.* 1979; Ryder *et al.* 1981). The observed associations provide strong evidence for the existence of 'disease' genes within the HLA region. Close linkage of a 'disease' gene to the HLA region results in distortions of the HLA haplotype sharing frequency distributions in affected sibs. The distribution of the number of HLA haplotypes shared by sib pairs (affected sib pair method) (Day & Simons, 1976; Thomson & Bodmer, 1977*a*, *b*; Suarez, 1978; Suarez *et al.* 1978; Spielman *et al.* 1980; Thomson, 1981; Louis *et al.* 1983; Green *et al.* 1983; Payami *et al.* 1984; Motro & Thomson, 1985; Ewens & Clarke, 1984) can be used to make inferences both about the mode of inheritance of the disease and the population frequency of the HLA-linked disease predisposing allele.

Using sib pair data, estimates of the disease predisposing allele frequency can be obtained under the recessive and additive models, and the fit of the observed data to these hypotheses can be tested (Motro & Thomson, 1985). Assuming that the penetrance of non-disease allele genotypes is zero (intermediate model), the sib pair data may also allow estimation of the relative penetrance of heterozygous disease susceptible individuals to homozygotes, and the

* Permanent address: Department of Genetics, The Hebrew University, Jerusalem 91904, Israel.

disease allele frequency (Louis *et al.* 1983). A general disease model, where the penetrance of the genotype with no susceptibility allele is non-zero, has an infinite number of possible solutions (Suarez, 1978; Louis *et al.* 1983).

Sibships in which there are three or more affected do not have the same distribution of parental genotypes at the 'disease' locus as those with only two affected (Weitkamp, 1981; Rubinstein *et al.* 1981; Suarez *et al.* 1983; Motro & Thomson, 1985; Ewens & Clark, 1984). The analysis of only families with three or more affected sibs can therefore be used to give us additional information regarding the mode of inheritance of the HLA associated diseases.

Sib pair data from families with three or more affected sibs can be analysed, and the tests of hypothesis (additive and recessive), and parameter estimates (additive, recessive and intermediate) compared with those obtained from the classical sib pair analysis, which analyses sib pair haplotype sharing from families with two or more affected sibs. In addition to the additive, recessive and intermediate models, the analysis of sib trio data (the number of haplotypes shared by trios) from families with three or more affected, allows parameter estimation for a general disease model to be made, when the data fall within the bounds of the expectations. It is possible that the smaller sample sizes available with sib trio data compared with sib pair data may be compensated for if the trio data allow more discrimination in the estimation of the disease parameters. (One of the weaknesses of the sib pair method is that slight changes in the haplotype sharing values can lead to large changes in the parameter estimates (Spielman *et al.* 1980; Louis *et al.* 1983; Ewens & Clark, 1984).)

In this paper we will present the expected HLA haplotype sharing distribution in pairs and in trios from families with three or more affected sibs. The expected HLA haplotype sharing distribution for sib quartets will also be presented. Data on insulin dependent diabetes mellitus (IDDM) will be analysed.

HLA HAPLOTYPE DISTRIBUTION IN SIB PAIRS FROM FAMILIES WITH THREE OR MORE AFFECTED SIBS

The underlying model which we consider throughout this work is that there is a gene with alleles D and d, which is involved in predisposing individuals to a particular disease, and that this gene is tightly linked to the HLA region such that recombination is negligible. The most general model of disease susceptibility allows all three genotypes DD, Dd and dd to be disease susceptible, with penetrance values for the three genotypes denoted f_2, f_1 and f_0 respectively. The distribution of the number of HLA haplotypes shared amongst affected sib pairs has been derived by Suarez (1978) for this general model (also see Louis *et al.* 1983).

If one considers only sib pairs obtained from families with at least k ($k = 3, 4, \ldots$) affected sibs, then the following distribution for the probability of sharing 2, 1 or 0 haplotypes denoted $X(k)$, $Y(k)$ and $Z(k)$ respectively, is obtained (equations 1).

$$X(k) = \Big[\tfrac{1}{4}p_D^4 + \frac{1}{2^{k-1}}p_D^3 p_d(1+\lambda_1)^{k-2}(1+\lambda_1^2) + \tfrac{1}{2}p_D^2 p_d^2 \lambda_1^k + \frac{1}{4^{k-1}}p_D^2 p_d^2(1+2\lambda_1+\lambda_0)^{k-2}(1+2\lambda_1^2+\lambda_0^2)$$
$$+ \frac{1}{2^{k-1}}p_D p_d^3(\lambda_1+\lambda_0)^{k-2}(\lambda_1^2+\lambda_0^2) + \tfrac{1}{4}p_d^4\lambda_0^k\Big] \Big/ \Sigma, \quad 1\,(\mathrm{i})$$

$$Y(k) = \frac{1}{2}\left[1 - \left(\frac{p_D^2 p_d^2}{4^{k-1}\Sigma}\right)(1+2\lambda_1+\lambda_0)^{k-2}(1-2\lambda_1+\lambda_0)^2\right], \quad 1\text{(ii)}$$

$$Z(k) = \left[\tfrac{1}{4}p_D^4 + \frac{1}{2^{k-2}}p_D^3 p_d(1+\lambda_1)^{k-2}\lambda_1 + \tfrac{1}{2}p_D^2 p_d^2 \lambda_1^k + \frac{2}{4^{k-1}}p_D^2 p_d^2(1+2\lambda_1+\lambda_0)^{k-2}(\lambda_1^2+\lambda_0) + \frac{1}{2^{k-2}}p_D p_d^3(\lambda_1+\lambda_0)^{k-2}\lambda_1\lambda_0 + \tfrac{1}{4}p_d^4\lambda_0^k\right]\Big/\Sigma, \quad 1\text{(iii)}$$

where

$$\Sigma = p_D^4 + \frac{1}{2^{k-2}}p_D^3 p_d(1+\lambda_1)^k + 2p_D^2 p_d^2\lambda^k + \frac{1}{4^{k-1}}p_D^2 p_d^2(1+2\lambda_1+\lambda_0)^k + \frac{1}{2^{k-2}}p_D p_d^3(\lambda_1+\lambda_0)^k + p_d^4\lambda_0^k,$$

p_D is the frequency of the disease predisposing allele D, with $p_d = 1-p_D$, $\lambda_1 = f_1/f_2$ and $\lambda_0 = f_0/f_2$.

For a recessive model ($f_2 \neq 0, f_1 = f_0 = 0$, i.e. $\lambda_1 = \lambda_0 = 0$) the form of equations (1) is such that $4X(k)Z(k) = Y(k)^2$, $X(k) \geqslant \frac{1}{4}$. This implies (Motro & Thomson, 1985) that for the recessive case the haplotype sharing distribution in sib pairs from families with at least k affected sibs ($k \geqslant 3$), henceforth referred to as the k-sib pair distribution, falls on the 2-sib pair (pairs from sibships with two or more affected) recessive distribution. The point where it falls on the curve is such that $X(k) < X(k-1)$ if $0 < p_D < 1$ for all $k \geqslant 3$ and corresponds to a disease allele frequency p_D^* on the 2-sib pair distribution such that

$$p_D^* = \frac{p_D}{p_D + \frac{1}{2^{k-2}}p_d}, \quad (2)$$

where p_D is the true 'disease' allele frequency. It is easy to show that $p_D^* \geqslant p_D$ for all $k \geqslant 3$, with equality only if $p_D = 0$ or 1. For example, if $k = 3$ and $p_D = 0{\cdot}05$ then $p_D^* = 0{\cdot}0952$, if $p_D = 0{\cdot}3$ then $p_D^* = 0{\cdot}4615$. If $k = 4$ and $p_D = 0{\cdot}3$ then $p_D^* = 0{\cdot}6316$.

Similarly, for the additive case ($f_2 = 2f_1, f_0 = 0, \lambda_1 = \frac{1}{2}, \lambda_0 = 0$) the k-sib pair distribution also falls on the 2-sib pair additive distribution ($Y(k) = \frac{1}{2}$, $X(k) \geqslant \frac{1}{4}$). (Note that $Y(k) = \frac{1}{2}$ if and only if $1-2\lambda_1+\lambda_0 = 0$, that is, only for additive models.) Again $X(k) < X(k-1)$ if $0 < p_D < 1$ for all $k \geqslant 3$. The k-sib distribution corresponds to a 'disease' allele frequency p_D^* on the 2-sib distribution such that

$$p_D^* = p_D + p_D(1-p_D)\{[(\tfrac{1}{2}4^{k-1} - 2\times 3^{k-1} + 3\times 2^{k-1} - 2)p_D^2 + (4\times 3^{k-2} - 2^{k+1} + 4)p_D + 2^{k-1} - 2]/[(\tfrac{1}{2}4^{k-1} - 2\times 3^{k-1} + 3\times 2^{k-1} - 2)p_D^3 + (2\times 3^{k-1} - 3\times 2^k + 6)p_D^2 + (3\times 2^{k-1} - 6)p_D + 2]\}, \quad (3)$$

where p_D is the true disease allele frequency. The term in square brackets is always positive for $0 < p_D < 1$ and $k \geqslant 3$, hence $p_D^* \geqslant p_D$ with equality only if $p_D = 0$ or 1. If $k = 3$ and $p_D = 0{\cdot}005$ then $p_D^* = 0{\cdot}0099$, if $p_D = 0{\cdot}1$ then $p_D^* = 0{\cdot}1692$. *If* $k = 4$ and $p_D = 0{\cdot}1$ then $p_D^* = 0{\cdot}2561$.

Weitkamp (1981) made the prediction that there should be less HLA haplotype sharing among pairs from sibships with three or more affected than from sibships with two or more affected. He presented data on insulin dependent diabetes mellitus (IDDM) and multiple sclerosis that followed this pattern. Weitkamp considered data from families with unaffected parents. However, when considering theoretical predictions it is sometimes appropriate not to consider

the affectional status of the parent. This greatly simplifies the calculations in the present case and gives distributions which can be directly related to the classical sib pair distribution from families with two or more affected sibs. If the affectional status of the parent is not considered then Weitkamp's predictions are true, in terms of the number of sib pairs sharing two haplotypes (as well as the mean number of shared haplotypes, μ, see below), for recessive and additive models, but do not necessarily hold for all intermediate models.

To estimate the 'disease' allele frequency, p_D, for recessive and additive models we need to estimate μ, the mean number of shared haplotypes in the affected sib pairs in the population. For the k-sib distribution, we define

$$\mu(k) = 2X(k) + Y(k),$$

and the estimates $\hat{X}(k)$ and $\hat{Y}(k)$ from the data are used to estimate $\mu(k)$. The estimates $\hat{X}(k)$, $\hat{Y}(k)$ [and $\hat{Z}(k)$], hence $\hat{\mu}(k)$, must take account of the ascertainment procedure in the collection of the family data (Motro & Thomson, 1985).

If ascertainment is by incomplete truncate selection (all families, independent of size and number of affected sibs, are equally likely to be included in the sample, Cavalli-Sforza & Bodmer, 1971, p. 853), then every sib pair from all families with 2 or more affected sibs contributes equally to the estimate of $\mu(2)$. For the more general case, in which families are selected with unequal probabilities, the contribution of sibs from families with different numbers of affected sibs must be appropriately weighted. For example, if families are ascertained by incomplete single selection, that is, ascertainment of a family is proportional to the number of affected sibs (Cavalli-Sforza & Bodmer, 1971, p. 853), then a family with say four affected sibs is 4/3 as likely to be ascertained as a family with three affected sibs and the data must be appropriately weighted (Motro & Thomson, 1985). It is important to note that the weighting procedure for the k-sib pair distribution is to weight families with more than k affected sibs to give them the appropriate k sib distribution, and then take sib pairs from this distribution. For example, in the case of incomplete truncate selection, sib quartets would contribute four sib trios to the sib trio distribution, and hence twelve sib pairs to the 3-sib pair distribution, as distinct from the six pairs they would contribute to the 2-sib pair distribution.

In order to take into account ascertainment procedures in the estimates of the disease parameters, we use cluster sampling techniques, the clusters being the different sibships. Because of this we use method of moment type estimators (MME) rather than maximum likelihood estimators (MLE) (see Motro & Thomson, 1985, for details).

In the recessive case the MME of p_D is

$$p_D = \begin{cases} \dfrac{2-\hat{\mu}}{2-\hat{\mu}+2^{k-1}(\hat{\mu}-1)} & \text{if} \quad \hat{\mu} \geqslant 1 \\ 1 & \text{if} \quad \hat{\mu} < 1 \end{cases} \tag{4}$$

which coincides with the MLE of p_D under the assumption of random sampling for this case.

For the additive case the MME of p_D is the solution of

$$\hat{\mu} = \left[\left(\tfrac{1}{2}p_D^3 + \frac{5\times 3^{k-2}}{2^{2k-2}}p_D^2 p_d + \frac{1}{2^{k-2}}p_D p_d^2 + \frac{1}{2^{2k-2}}p_d^3\right)\Big/ \left(p_D^3 + \frac{3^k}{2^{2k-2}}p_D^2 p_d + \frac{3}{2^{k-1}}p_D p_d^2 + \frac{1}{2^{2k-2}}p_d^3\right)\right] + \frac{1}{2}. \tag{5}$$

Table 1. *The expected HLA haplotype sharing frequencies in sib trios for a recessive model (p_D = susceptibility allele frequency)*

p_D	α	β	γ	δ
0·05	0·7562	0·2268	0·0057	0·0113
0·1	0·5917	0·3550	0·0178	0·0355
0·2	0·3906	0·4687	0·0469	0·0938
0·3	0·2770	0·4986	0·0748	0·1496
0·4	0·2066	0·4959	0·0992	0·1983
0·5	0·1600	0·4800	0·1200	0·2400
0·6	0·1275	0·4592	0·1378	0·2755
0·7	0·1040	0·4370	0·1530	0·3060
0·8	0·0865	0·4152	0·1661	0·3322
0·9	0·0730	0·3945	0·1775	0·3550
1·0	0·0625	0·3750	0·1875	0·3750

The MLE and MME of p_D are not identical for the additive model.

For the intermediate model ($\lambda_1 \neq 0$, $\lambda_0 = 0$) (Spielman *et al.* 1980) a Newton–Raphson procedure can be used to obtain estimates of λ_1 and p_D (Louis *et al.* 1983). For the general model ($\lambda_1, \lambda_0 \neq 0$) an observed set of k-sib pair haplotype sharing values is consistent with an infinite number of parameter set (p_D, λ_1, λ_0) solutions.

Sib Trios

There are four possible combinations of haplotype sharing among three affected sibs. If we denote the maternal haplotypes by a and b and the paternal by c and d, then these combinations (or their equivalents) are (ac, ac, ac), (ac, ac, ad), (ac, ac, bd) and (ac, ad, bd). These can be represented by the number of haplotypes shared by the sibs in the three possible pairwise comparisons, and we denote them by α (all 3 sibs share both haplotypes), β (one pair shares 2, and two pairs share one haplotype), γ (one pair shares 2, and two pairs share no haplotype) and δ (two pairs share 1, and one pair shares no haplotype) (Spielman *et al.* 1979; Suarez *et al.* 1982). For each mating type the probability that 3 children would be affected and have one of the above haplotype sharing configurations is given in Appendix I, and the procedure for deriving the α, β, γ and δ for the general model is given in Appendix I.

For the recessive model ($\lambda_1 = \lambda_0 = 0$) the expected distribution of haplotype sharing in sib trios is given by

$$\alpha = \frac{1}{(1+3p_D)^2}, \quad \beta = \frac{6p_D}{(1+3p_D)^2}, \quad \gamma = \frac{3p_D^2}{(1+3p_D)^2}, \quad \delta = \frac{6p_D^2}{(1+3p_D)^2}. \tag{6}$$

Some representative values of this distribution are given in Table 1. As p_D approaches 1·0 the (α, β, γ, δ) distribution tends to (1/16, 6/16, 3/16, 6/16), the values expected under random segregation of the parental haplotypes and the disease.

If we denote the observed values of the classes α, β, γ, and δ by $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$ respectively, then the MME of p_D, in the recessive case, is

$$p_D^* = \begin{cases} \dfrac{2-\hat{\mu}}{3\hat{\mu}-2} & \text{if} \quad 3\hat{\alpha}+\hat{\beta}-\hat{\gamma}-\hat{\delta} \geqslant 0 \\ 1 & \text{if} \quad 3\hat{\alpha}+\hat{\beta}-\hat{\gamma}-\hat{\delta} \leqslant 0, \end{cases} \tag{7}$$

Table 2. *The expected HLA haplotype sharing frequencies in sib trios for an additive model (p_D = susceptibility allele frequency)*

p_D	α	β	γ	δ
0·05	0·1962	0·6425	0·0538	0·1075
0·1	0·1658	0·5816	0·0842	0·1684
0·2	0·1316	0·5132	0·1184	0·2368
0·3	0·1120	0·4741	0·1380	0·2759
0·4	0·0989	0·4478	0·1511	0·3022
0·5	0·0893	0·4286	0·1607	0·3214
0·6	0·0818	0·4136	0·1682	0·3364
0·7	0·0757	0·4014	0·1743	0·3486
0·8	0·0706	0·3912	0·1794	0·3588
0·9	0·0663	0·3825	0·1837	0·3675
1·0	0·0625	0·3750	0·1875	0·3750

Table 3. *HLA haplotype sharing in 538 insulin dependent diabetic siblings*

Number affected sibs	Families	Haplotype sharing
2	481	$(a/c, a/c) = 265$; $(a/c, a/d) = 183$; $(a/c, b/d) = 33$
3	46	$(a/c, a/c, a/c) = 14$; $(a/c, a/c, a/d) = 24$; $(a/c, a/c, b/d) = 4$; $(a/c, a/d, b/d) = 4$
4	7	$(a/c, a/c, a/c, a/c) = 1$; $(a/c, a/c, a/c, a/d) = 2$; $(a/c, a/c, a/d, a/d) = 3$; $(a/c, a/c, a/d, b/c) = 1$
5	2	$(a/c, a/c, a/c, a/d, a/d) = 1$; $(a/c, a/c, b/d, a/d, b/c) = 1$
6	2	$(a/c, a/c, a/c, a/c, a/c, a/d) = 1$; $(a/c, a/c, a/c, b/d, b/d, a/d) = 1$

Sources: Anderson, 1983; Barbosa *et al.* 1977, 1980; Bengsch *et al.* 1978; Berger *et al.* 1980; Bertrams *et al.* 1984; Clerget-Darpoux *et al.* 1981; Contu *et al.* 1977, 1983; Cudworth & Woodrow, 1975; Hsu *et al.* 1977; Jawarski *et al.* 1980; Krawisz *et al.* 1978; Ludwig *et al.* 1977; Platz *et al.* 1981; Rubinstein *et al.* 1976, 1977; Ryder *et al.* 1979; Savi *et al.* 1977; Serjeantson *et al.* 1980; Spielman *et al.* 1980; Spielman, 1982; Sucio-Foca *et al.* 1979; Thomsen *et al.* 1975; Torfs, 1984; Walker & Cudworth, 1980; Wolf *et al.* 1983; Cudworth, pers. comm.; 8th International Histocompatibility Workshop kindly provided by Dr M. Baur (only data from families with more than two affected sibs included).

To avoid overlaps only the original data from each report were included.

where $\hat{\mu} = (6\hat{\alpha}+4\hat{\beta}+2\hat{\gamma}+2\hat{\delta})/[3(\hat{\alpha}+\hat{\beta}+\hat{\gamma}+\hat{\delta})]$, which is equivalent to the MLE of p_D under the assumption of random mating and to the 3-sib pair estimate (see equation (4)).

For the additive model the haplotype sharing distribution for sib trios has the form

$$\alpha = \tfrac{1}{4}(1+3p_D)/A, \quad \beta = \tfrac{3}{4}(1+5p_D+2p_D^2)/A, \quad \gamma = \tfrac{3}{2}p_D(1+p_D)/A, \quad \delta = 3p_D(1+p_D)/A, \quad (8)$$

where

$$A = 1+3p_D(3+2p_D).$$

Some representative values of this distribution are given in Table 2. By the method of moments the estimate of p_D can be obtained by solving

$$\hat{p}_D^2(12\hat{\mu}-10)+\hat{p}_D(18\hat{\mu}-19)+2\hat{\mu}-3 = 0. \quad (9)$$

Note that $\delta = 2\gamma$ for both the recessive and additive models, but this is not the case for other models.

For the general model, the observed values of the four haplotype sharing classes can be used

Table 4. *Goodness of fit of IDDM sib data to a recessive model and the susceptibility allele frequency estimates* ($\hat{p}_D$)

2-sibpair	$\hat{X}$	$\hat{Y}$	$\hat{Z}$	P value	$\hat{p}_D$	
I.T. (obs.)	373	283	55	0·907	0.382	
(exp.)	372·3	284·4	54·3			
I.S. (obs.)	327·8	239·5	44·7	0.913	0.367	
(exp.)	327·5	240·4	44·1			
3-sibpair						
I.T. (obs.)	187	168	47	> 0·3	0.319	
(exp.)	182·69	176·62	42·69			
I.S. (obs.)	139·3	126·8	30·9	> 0·7	0·303	
(exp.)	138·34	128·72	29·94			
sib trio	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$	P value	$\hat{p}_D$
I.T. (obs.)	32	69	14	19	> 0·5	0·338
(exp.)	33	67	11·3	22·7		
I.S. (obs.)	24·6	52·6	9·1	12·7	> 0.7	0·315
(exp.)	26·2	49·5	7·8	15·5		

to obtain a unique solution for p_D, λ_1, and λ_0, using the Newton–Raphson procedure, if the observed data fall within the bounds of the expectations.

The expected haplotype sharing distributions for sib quartets (4 affected sibs) for recessive and additive models are given in Appendix II.

IDDM Data Analysis

We have collected the published data on the HLA haplotype sharing distribution in Caucasian sibships with insulin dependent diabetes mellitus (IDDM) (see Table 3). We have analysed the data in the form of 2-sib pair, 3-sib pair, and sib trio distributions. We tested each data set against recessive and additive modes of inheritance using a chi-square goodness-of-fit test. (The dominant expectations in each case fall so close to the additive expectations that the two cases are not distinguishable.) For cases in which the hypothesized mode of inheritance was not rejected, an estimate of the 'disease' allele frequency was obtained. In all our analyses we considered two ascertainment methods: incomplete truncate selection (I.T.) (all families, independent of size and number of affected sibs are equally likely to be included in the sample) and incomplete single selection (I.S.) (families are ascertained with probabilities proportional to the number of affected sibs in the family).

An additive model was rejected by both the 2-sib pair and sib trio data sets (P values $< 10^{-4}$ and $< 0{\cdot}05$ respectively). The 3-sib pair distribution, however, is not significantly different from the expectations for an additive model for both ascertainment procedures (P value $< 0{\cdot}08$) and the estimated value of p_D is 0·035 for incomplete single selection. (It is always more difficult to distinguish between recessive and additive hypotheses using the 3-sib pair distribution than using the 2-sib pair distribution due to the shifting of the 3-sib pair distribution towards the random expectations.)

All three distributions are compatible with a recessive model. Table 4 shows the observed

and the expected haplotype sharing values ($\hat{X}$, $\hat{Y}$, $\hat{Z}$), the goodness-of-fit P values and the 'disease' allele frequencies that were estimated for the three distributions.

The 2-sib pair and 3-sib pair and sib trio data points fall outside the bounds of their respective expected haplotype sharing spaces. Therefore, it was not possible to obtain solutions for an intermediate or a general model in these cases.

Although the sib pair data are very close to single recessive susceptibility allele expectations (see Table 4), the estimated allele frequencies are too high and incompatible with the observed population values (see, for example, Louis *et al.* 1983). For example, if $p_D = 0{\cdot}3$ (the estimated allele frequency for the sib trio data under incomplete single selection) and $f_2 = 0{\cdot}18$ (MZ concordance rate (Spielman *et al.* 1980)), then the expected population prevalence would be 0·016 which is much higher than the observed value of 0·004 (Spielman *et al.* 1980). Unless an unrealistically high recombination frequency is assumed, recombination cannot account for the discrepancies (Payami *et al.* 1984). IDDM may confer considerable selective disadvantage on the patients. If so, then the estimated 'disease' allele frequencies obtained by sib methods are lower than the true frequency (Risch, 1982; Payami *et al.* 1984). Therefore, even larger discrepancies may exist between the observed population values and the estimates obtained for the single recessive susceptibility allele model.

DISCUSSION

The classical sib pair method has been used for a number of HLA associated diseases to test recessive and additive modes of inheritance, and to obtain solutions for an intermediate model. In all intermediate models the penetrance of the non-disease allele homozygote is assumed to be zero. If all three genotypes were assumed to be disease susceptible, that is, a general model, then the haplotype sharing distribution in sib pairs would generate an infinite number of solutions. The affected sib trio method, however, can theoretically provide a unique solution for a general model, as well as for intermediate models.

The IDDM sib pair and sib trio data points fall outside their respective expected haplotype sharing spaces. Therefore, it is not possible to obtain any intermediate or general solution. The observed values are significantly different from additive (dominant) expectations, but are very close to the recessive expectations. The estimates of the disease allele frequency p_D are similar for the sib pair and sib trio data, although larger estimates are obtained from the sib pair data. The 'disease' allele frequencies obtained for the recessive model, however, are too high and incompatible with the population data. These results do not support a single susceptibility allele model for IDDM. It has been suggested that IDDM may be genetically heterogeneous (Svejgaard *et al.* 1975; Rotter & Rimoin, 1978; Barbosa *et al.* 1980; Rotter *et al.* 1983; Platz *et al.* 1981; Thomson, 1983; Anderson *et al.* 1983). The observed associations of IDDM with DR3 and DR4 and non-HLA markers such as the insulin gene (Bell & Karam, 1983; Karam, pers. comm.) suggest predisposition by at least two alleles at the HLA-linked locus, as well as non-HLA linked loci.

To investigate the genetic heterogeneity of diseases such as IDDM, we have derived the expected haplotype sharing distributions in sib pairs and sib trios under two hypotheses: (1) Two alleles at the same HLA locus predispose to the disease (allowing for negative complementation of susceptibility alleles); (2) Two alleles at different loci, only one of which is HLA-linked,

predispose to the disease (Louis *et al.* 1984). The 2 susceptibility allele expectations for sib pairs remain within the limits of the single susceptibility allele model. If the penetrance of the heterozygote genotype with the two susceptibility alleles is greater than 1·5 times the penetrance of the homozygote susceptible genotypes, then the haplotype sharing distribution in sib pairs is close to the recessive expectations under the single susceptibility allele model. The sib trio distribution under the negative complementation model, however, falls outside the bounds of single susceptibility allele expectations. Therefore, the two models may be distinguished using the sib trio haplotype sharing distribution.

For the two locus model the expected haplotype sharing distributions in sib pairs and sib trios fall within the respective spaces for the single susceptibility allele model. Therefore, these two models are indistinguishable.

The IDDM sib pair data are compatible with all three models, namely: the single susceptibility allele model with recessive inheritance, the negative complementation model, and the two locus model. The IDDM sib trio data, however, fall in a region that is expected under the negative complementation model but not the single susceptibility allele model. The single susceptibility allele model, however, cannot be statistically rejected. The possibility of the existence of a second non-HLA linked susceptibility locus is not negated by this result.

Research supported by NIH grant HD12731 and by an NSF graduate fellowship to E.J.L.

REFERENCES

ANDERSON, C.A., HODGE, S.E., RUBIN, R., ROTTER, J.J., TERASAKI, P.J., IRVINE, W.J. & RIMOIN, D.L. (1983). A search for heterogeneity in insulin dependent diabetes mellitus (IDDM): HLA and autoimmune studies in simplex, multiplex and multigenerational families. *Metabolism* **32**, no. 5, 471–477.

BARBOSA, J., CHERN, M. M., ANDERSON, V. E., NOREEN, H., JOHNSON, S., REINSMOEN, N., MCCARTHY, R., KING, R. & GREENBERG, L. (1980). Linkage analysis between the major histocompatibility system and insulin dependent diabetes in families with patients in two consecutive generations. *J. Clin. Invest.* **65**, 592–601.

BARBOSA, J., KING, R. & NOREEN, H. (1977). The histocompatibility system in juvenile, insulin dependent diabetic multiplex kindreds. *J. Clin. Invest.* **60**, 989–998.

BELL, G. J. & KARAM, J. H. (1983). The polymorphic locus flanking the human insulin gene: is there an association with diabetes mellitus? Banbury Report **14**: *Recombinant DNA Applications to Human Disease.* Cold Spring Harbor Laboratory.

BENGSCH, N., KOBBERLING, J., ECKERT, G. & WILLIAMS, B. (1978). HLA-typing in juvenile diabetes with and without positive history in families with one and two diabetic siblings. *Diabetologia* **15**, 447–451.

BERGER, W., CALABI, F., GAROTTA, G., NERI, T. M. & CEPPELLINI, R. (1980). HLA and insulin dependent diabetes mellitus: high risk for DRW3/DRW4 heterozygous and segregation in a *B/DR* recombinant family. *Histocompatibility Testing 1980*, 948.

BERTRAMS, J., BAUR, M. P., CLERGET-DARPOUX, F., KING, M. C., KOLLMAR, C., NEUGEBAUR, M. & PAYAMI, H. (1984). *Histocompatibility Testing 1984.* (In the Press.)

BODMER, W. F. (1978). HLA: A super supergene. In *The Harvey Lectures*, Series 72, pp. 91–138. New York: Academic Press.

CAVALLI-SFORZA, L. L. & BODMER, W. F. (1971). *The Genetics of Human Populations.* San Francisco: W. H. Freeman and Company.

CLERGET-DARPOUX, F., BONAÏTI-PELLIÉ, C., DESCHAMPS, I., HORS, J. & FEINGOLD, N. (1981). Juvenile insulin-dependent diabetes: a possible susceptibility gene in interaction with HLA. *Ann. Hum. Genet.* **45**, 199–206.

CONTU, L., DESCHAMPS, I., LESTRADET, H., HORS, J., SCHMID, M., BUSSON, M., BENJAMIN, A., MARCELLI-BARGE, A. & DAUSSET, J. (1982). HLA haplotype study of 53 juvenile insulin dependent diabetic (I.D.D.) families. *Tissue Antigens* **20**, 123–140.

CONTU, L., PULIGHEDDER, A., MURA, C. & GABBAS, A. (1977). HLA antigens in Sardinia patients with diabetes mellitus. *Diabetes and Metabolism* **3**, 263–267.

CUDWORTH, A. G. & WOODROW, J. C. (1975). HL-A system and diabetes mellitus. *Diabetes* **24**, 345–349.

DAUSSET, J. & SVEJGAARD, A. (*Eds.*) (1977). In *HLA and Disease*. Copenhagen: Munksgaard.

DAY, N. E. & SIMONS, M. J. (1976). Disease susceptibility genes – their identification by multiple case family studies. *Tissue Antigens* **11**, 471–474.

EWENS, W. J. & CLARKE, C. P. (1984). Maximum likelihood estimation of genetic parameters of HLA linked disease using data from families of various sizes. *Am. J. Hum. Genet.* (In the Press.)

GREEN, J. R., LOW, H. C. & WOODROW, J. C. (1983). Inference on inheritance of disease using repetitions of HLA haplotypes in affected siblings. *Ann. Hum. Genet.* **47**, 73–82.

HSU, H. S., HSU, H. T., CLASE, G. A. & BIAS, W. B. (1977). Further data on HLA recombination in juvenile onset diabetes. *Transpl. Proc.* **IX.4**, 1855–1858.

JAWARSKI, M. A., COLLE, E., GUTTMA, R. D., BELMONTE, M. M., TAYLOR, B., CREPEAU, M. P., WILKINS, J. & POIRIER, R. (1980). Insulin dependent diabetes: a comparison of families with single and multiple affected siblings. *Diabetologia* **19**, 97–100.

KRAWISZ, J. E., PALUMBO, P. J., TASWELL, H. F. & ELVEBACK, L. R. (1978). HLA antigens in patients with juvenile diabetes and their first degree relatives. *Mayo Clin. Proc.* **53**, 782–787.

LOUIS, E. J., THOMSON, G. & PAYAMI, H. (1983). The affected sib method. II. The intermediate model. *Ann. Hum. Genet.* **47**, 225–243.

LOUIS, E. J., THOMSON, G. & PAYAMI, H. (1984). Affected sib methods. *Histocompatibility Testing 1984*. 662–663.

LUDWIG, H., SCHERNTHANER, G. & MAYR, W. R. (1977). The importance of HLA genes to susceptibility in the development of juvenile diabetes mellitus. A study of 93 patients and 68 first degree blood relations. *Diabetes and Metabolism* **3**, 43–48.

MOTRO, U. & THOMSON, G. (1985). The affected sib methods: statistical features of the affected sib pair method. *Genetics* **10**, 525–538.

PAYAMI, H., THOMSON, G. & LOUIS, E. J. (1984). The affected sib method. III. Selection and recombination. *Am. J. Hum. Genet.* **36**, 352–362.

PLATZ, P., JAKOBSEN, B. K., MORLING, N., RYDER, L. P., SVEJGAARD, A., THOMSEN, M., CHRISTY, M., KROMANN, H., BENN, J., NERUP, J., GREEN, A. & HANGE, M. (1981). HLA-D and -DR antigens in genetic analysis of insulin dependent diabetes mellitus. *Diabetologia* **21**, 108–115.

RISCH, N. (1982). Affected sib pair marker allele sharing: the effect of reduced fertility, variable family size, and a second unlinked locus. *Am. J. Hum. Genet.* **34**, 191A.

ROTTER, J. J., ANDERSON, C. E., RUBIN, R., CONGLETON, J. E., TERASAKI, P. J. & RIMOIN, D. L. (1983). HLA genotypic study of insulin dependent diabetes. The excess of DR3/DR4 heterozygotes allows rejection of the recessive hypothesis. *Diabetes* **32**, 169–174.

ROTTER, J. J. & RIMOIN, D. L. (1978). Heterogeneity in diabetes mellitus. *Diabetes* **27**, 5, 599–608.

RUBINSTEIN, P., GINSBERG-FELLNER, F. & FALK, C. (1981). Genetics of type I diabetes mellitus: a single, recessive predisposition gene mapping between *HLA-B* and *GLO*. *Am. J. Hum. Genet.* **33**, 865–882.

RUBINSTEIN, P., NICHOLSON, J. F. & SUCIO-FOCA, N. (1977). Intra-HLA recombinations in juvenile diabetes mellitus. *Diabetes and Metabolism* **3**, 199–204.

RUBINSTEIN, P., SUCIO-FOCA, N., NICHOLSON, J. F., FOTINO, M., MOLINAR, A., HARISIADIS, L., HARDY, M. A., REEMTSMA, K. & ALLEN, H. (1976). The HLA system in the families of patients with juvenile diabetes mellitus. *J. Exp. Med.* **143**, 1277–1283.

RYDER, L. P., CHRISTY, M., NERUP, J., PLATZ, P., SVEJGAARD, A. & THOMSEN, M. (1979). HLA studies in diabetes. *Exp. Med. Biol.* **119**, 41–50.

RYDER, L. P. & SVEJGAARD, A. (1981). Genetics of HLA disease associations. *Ann. Rev. Genet.* **15**, 169–187.

SAVI, M., NERI, T. M., ZAVARONI, I. & COSCELLI, I. (1977). HLA antigens and insulin dependent diabetes mellitus – a family study. *Diabetes and Metabolism* **3**, 259–262.

SERJEANTSON, S., KIRK, R. L., DRY, P. J., RYAN, D. P., COURT, J., ZIMMET, P. & STEPANAS, A. V. (1980). HLA studies in Australian multiple case families of juvenile onset diabetes mellitus. *Med. J. Aust.* **1**, 107–109.

SPIELMAN, R. S. (1982). Differential penetrance and dose of susceptibility genes in insulin dependent diabetes. Serono Symposium No. 47, *The Genetics of Diabetes Mellitus* (eds. J. Kobberling and R. Tattersall). London, New York: Academic Press.

SPIELMAN, R. S., BAKER, L. & ZMIJEWSKI, C. M. (1979). Inheritance of susceptibility to juvenile onset diabetes. In *Genetic Analysis of Common Diseases: Application to Predictive Factors in Coronary Disease* (eds. C. F. Sing and M. Skolnick), pp. 567–585. New York: Alan R. Liss.

SPIELMAN, R. S., BAKER, L. & ZMIJEWSKI, C. M. (1980). Gene dosage and susceptibility to insulin-dependent diabetes. *Ann. Hum. Genet.* **44**, 135–150.

SUAREZ, B. K. (1978). The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. *Tissue Antigens* **12**, 87–93.

SUAREZ, B. K., CROUSE, J. & VAN EERDEWEGH, P. (1983). HLA sharing in multiplex sibships. *Ann. Hum. Genet.* **47**, 153–159.

SUAREZ, B., O'ROURKE, D. & VAN EERDEWEGH, P. (1982). Power of the affected sib pair method to detect disease susceptibility loci of small effect: an application to multiple sclerosis. *Am. J. Med. Genet.* **12**, 309–326.

SUAREZ, B. K., RICE, J. & REICH, T. (1978). The generalized sib pair IBD distribution: its use in the detection of linkage. *Ann. Hum. Genet.* **42**, 87–94.

SUCIO-FOCA, N., RUBINSTEIN, P., NICHOLSON, J., SUSINNO, E., WEINER, J., GODFREY, M., HARDY, M., RAYFIELD, E. & REEMTSMA, K. (1979). Juvenile diabetes mellitus and the HLA system. *Transpl. Proc.* **XI.2**, 1309–1313.

SVEJGAARD, A., PLATZ, P., RYDER, L., STAUB NIELSON, L. & THOMSEN, M. (1975). HLA and disease associations – a survey. *Transpl. Rev.* **22**, 3–43.

THOMSEN, M., PLATZ, P., ORTVED ANDERSEN, O., CHRISTY, M., LYNGSOE, J., NERUP, J., RASMUSSEN, K., RYDER, L. P., STAUB NIELSON, L. & SVEJGAARD, A. (1975). MLC typing in juvenile diabetes mellitus and idiopathic Addison's disease. *Transpl. Rev.* **22**, 125–147.

THOMSON, G. (1981). A review of theoretical aspects of HLA and disease associations. *Theor. Pop. Biol.* **20** (2), 168–208.

THOMSON, G. & BODMER, W. F. (1977*a*). The genetics of HLA and disease associations. In *Measuring Selection in Natural Populations* (eds. F. B. Christiansen, T. Fenchel and O. Barndorff-Nielson). Berlin, New York: Springer-Verlag.

THOMSON, G. & BODMER, W. F. (1977*b*). The genetic analysis of HLA and disease associations. In *HLA and Disease* (eds. J. Dausset and A. Svejgaard). Copenhagen: Munksgaard.

THOMSON, G. (1983). Investigation of the mode of inheritance of HLA associated diseases by the antigen genotypic frequencies amongst diseased method. *Tissue Antigens* **21**, 81–104.

TORFS, C. (1984). Genetic susceptibility in the epidemiology of autoimmune diseases. Ph.D. Thesis. University of California, Berkeley.

WALKER, A. & CUDWORTH, A. G. (1980). Type 1 (insulin dependent) diabetic multiplex families. Mode of genetic transmission. *Diabetes* **29**, 1036–1039.

WEITKAMP, L. (1981). HLA and disease: predictions for *HLA* haplotype sharing in families. *Am. J. Hum. Genet.* **33**, 776–784.

WOLF E., SPENCER, K. M. & CUDWORTH, A. G. (1983). The genetic susceptibility to type 1 (insulin dependent) diabetes: analysis of the HLA–Dr association. *Diabetologia* **24**, 224–230.

APPENDIX I

Derivation of the sib trio expected haplotype sharing values for a general model

Mating type Expected haplotype sharing distribution in 3 *affected offspring*

$DD \times DD \quad \alpha_1 = \frac{1}{16} p_D^4 f_2^3, \quad \beta_1 = \frac{3}{8} p_D^4 f_2^3, \quad \gamma_1 = \frac{3}{16} p_D^4 f_2^3, \quad \delta_1 = \frac{3}{8} p_D^4 f_2^3.$

$DD \times Dd \quad \alpha_2 = \frac{1}{8} p_D^3 p_d (f_2^3 + f_1^3), \quad \beta_2 = \frac{3}{8} p_D^3 p_d (f_2^3 + f_1^3 + f_2^2 f_1 + f_2 f_1^2),$
$\gamma_2 = \frac{3}{8} p_D^3 p_d (f_2^2 f_1 + f_2 f_1^2), \quad \delta_2 = \frac{3}{4} p_D^3 p_d (f_2^2 f_1 + f_2 f_1^2).$

$Dd \times Dd \quad \alpha_3 = \frac{1}{16} p_D^2 p_d^2 (f_2^3 + 2f_1^3 + f_0^3), \quad \beta_3 = \frac{3}{8} p_D^2 p_d^2 (f_2^2 f_1 + f_2 f_1^2 + f_1^2 f_0 + f_1 f_0^2),$
$\gamma_3 = \frac{3}{16} p_D^2 p_d^2 (2f_1^3 + f_2^2 f_0 + f_2 f_0^2), \quad \delta_3 = \frac{3}{8} p_D^2 p_d^2 (f_2 f_1^2 + f_1^2 f_0 + 2f_2 f_1 f_0).$

$Dd \times dd \quad \alpha_4 = \frac{1}{8} p_D p_d^3 (f_1^3 + f_0^3), \quad \beta_4 = \frac{3}{8} p_D p_d^3 (f_1^3 + f_0^3 + f_1^2 f_0 + f_1 f_0^2),$
$\gamma_4 = \frac{3}{8} p_D p_d^3 (f_1^2 f_0 + f_1 f_0^2), \quad \delta_4 = \frac{3}{4} p_D p_d^3 (f_1^2 f_0 + f_1 f_0^2).$

$DD \times dd \quad \delta_5 = \frac{1}{8} p_D^2 p_d^2 f_1^3, \quad \beta_5 = \frac{3}{4} p_D^2 p_d^2 f_1^3, \quad \gamma_5 = \frac{3}{8} p_D^2 p_d^2 f_1^3, \quad \delta_5 = \frac{3}{4} p_D^2 p_d^2 f_1^3.$

$dd \times dd \quad \alpha_6 = \frac{1}{16} p_d^4 f_0^3, \quad \beta_6 = \frac{3}{8} p_d^4 f_0^3, \quad \gamma_6 = \frac{3}{16} p_d^4 f_0^3, \quad \delta_6 = \frac{3}{8} p_d^4 f_0^3.$

Then, the expected population frequency of, for example, sharing both haplotypes by all 3 affected sibs (α) is the sum of $\alpha_1 - \alpha_6$ divided by the total, that is, $\alpha = \Sigma_{i=1}^{6} \alpha_i / \Sigma_{i=1}^{6} \alpha_i + \beta_i + \gamma_i + \delta_i$. Likewise, β, γ and δ are the sum of their respective frequencies for all mating types divided by the total.

APPENDIX II

The distributions of expected haplotype sharing values for sib quartets for (a) recessive and (b) additive models of inheritance

The numbers in parenthesis represent the number of sib pairs in each quartet that share two, one and zero haplotypes, respectively. For example, (2,4,0) is the case in which 2 sib pairs share two, 4 share one, and none share zero haplotypes.

(*a*) Recessive

$(6,0,0) = 1/A \quad (2,0,4) = 3P_D^2/A$
$(3,3,0) = 8p_D/A \quad (1,4,1) = 12p_D^2/A$
$(3,0,3) = 4p_D^2/A \quad (1,3,2) = 24p_D^2/A$
$(2,4,0) = 6p_D/A \quad (0,4,2) = 6p_D^2/A$

where $A = (1+7p_D)^2$.

(*b*) additive

$(6,0,0) = \frac{1}{8}(1+7p_D)/B$
$(3,3,0) = \frac{1}{2}(1+9p_D+6p_D^2/B$
$(3,0,3) = p_D(1+3P_D)/B$
$(2,4,0) = \frac{3}{8}(1+7p_D+8p_D^2)/B$
$(2,0,4) = \frac{3}{4}p_D(1+2p_D+p_D^2)/B$
$(1,4,1) = \frac{3}{2}p_D(3+4p_D+p_D^2)/B$
$(1,3,2) = 3p_D(2+5p_D+p_D^2)/B$
$(0,4,2) = \frac{3}{4}p_D(1+6p_D+P_D^2)/B$

where $B = 1+3p_D(2p_D^2+12p_D+7)$.