Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-dependent Diabetes Mellitus (IDDM)

Richard S. Spielman,* Ralph E. McGinnis,* and Warren J. Ewens†

* Department of Genetics, University of Pennsylvania School of Medicine, and †Department of Biology, University of Pennsylvania, Philadelphia

Summary

A population association has consistently been observed between insulin-dependent diabetes mellitus (IDDM) and the "class 1" alleles of the region of tandem-repeat DNA (5′ flanking polymorphism [5′FP]) adjacent to the insulin gene on chromosome 11p. This finding suggests that the insulin gene region contains a gene or genes contributing to IDDM susceptibility. However, several studies that have sought to show linkage with IDDM by testing for cosegregation in affected sib pairs have failed to find evidence for linkage. As means for identifying genes for complex diseases, both the association and the affected-sib-pairs approaches have limitations. It is well known that population association between a disease and a genetic marker can arise as an artifact of population structure, even in the absence of linkage. On the other hand, linkage studies with modest numbers of affected sib pairs may fail to detect linkage, especially if there is linkage heterogeneity. We consider an alternative method to test for linkage with a genetic marker when population association has been found. Using data from families with at least one affected child, we evaluate the transmission of the associated marker allele from a heterozygous parent to an affected offspring. This approach has been used by several investigators, but the statistical properties of the method as a test for linkage have not been investigated. In the present paper we describe the statistical basis for this "transmission test for linkage disequilibrium" (transmission/disequilibrium test [TDT]). We then show the relationship of this test to tests of cosegregation that are based on the proportion of haplotypes or genes identical by descent in affected sibs. The TDT provides strong evidence for linkage between the 5′FP and susceptibility to IDDM. The conclusions from this analysis apply in general to the study of disease associations, where genetic markers are usually closely linked to candidate genes. When a disease is found to be associated with such a marker, the TDT may detect linkage even when haplotype-sharing tests do not.

Introduction

A crucial first step in finding gene loci that contribute to a genetic disease is to demonstrate linkage with a gene or DNA sequence of known location (a "marker," usually a DNA polymorphism). A number of investigators have used this approach in the study of diabetes mellitus. Bell et al. (1984) described a population association between insulin-dependent diabetes mellitus (IDDM) and the 5′ flanking polymorphism (5′FP), an RFLP adjacent to the insulin gene on chromosome 11p. Although it is not clear that insulin or the insulin gene itself plays a role in the pathogenesis of IDDM, the association has been found consistently in population studies (for a summary, see Cox et al. 1988). In unaffected controls, the frequency of the smaller, or "class 1," alleles is approximately .70–.75, while in IDDM patients the frequency is somewhat higher: .80–.85. This finding provides *indirect* evidence for linkage between the insulin gene region and genes that influence

Received July 13, 1992; revision received November 10, 1992.
Address for correspondence and reprints: Dr. Richard Spielman, Department of Genetics, University of Pennsylvania School of Medicine, 422 Curie Boulevard, Philadelphia, PA 19104-6145.

0002-9297/93/5203-0008$02.00

susceptibility to IDDM, since an association between disease and marker may be due to disequilibrium between linked loci. However, the problem with inferring linkage from population association is that association can occur in the absence of linkage—for example, as a result of population stratification. Thus it is not valid to use the presence of association as a test for linkage if population stratification is a possibility.

For this reason, tests of linkage that do *not* depend on association were carried out by various investigators. In most of these studies, there was no direct evidence for linkage (Hitman et al. 1985; Ferns et al. 1986). In larger samples, the distribution of 5′FP alleles in 33 affected sib pairs (ASPs) with IDDM (Cox et al. 1988) or in the 95 ASPs studied in Genetic Analysis Workshop 5 (GAW5) (Cox and Spielman 1989; Spielman et al. 1989) failed entirely to provide evidence for linkage. Thus the absence of cosegregation within families suggested that the association was due to population stratification rather than to disequilibrium with a linked locus.

However, other approaches have suggested that the association is not due solely to stratification. Using the method of Field et al. (1986), Thomson et al. (1989) analyzed the GAW5 family data by the following method. In each family, the four parental 5′FP alleles were assigned to one of two categories: (1) transmitted to at least one diabetic offspring ("diseased") and (2) not transmitted to any affected offspring ("control"). This method has been termed "AFBAC," for "affected family-based controls" (Thomson 1988). As tested by a conventional χ^2, the frequency of 5′FP class 1 alleles in the diseased category (.83) was significantly higher than that in the controls (.69) ($p < .01$). Since the control and disease samples are obtained from the same individuals, the contribution of stratification to the association is reduced or eliminated. However, the comparison does not provide a *direct* test for linkage.

In the present paper we describe a procedure which tests directly for linkage between a disease and marker locus which shows population association; this test is not affected by the presence of stratification. The data for the test are from families with one or more affected offspring and at least one parent who is *heterozygous* for a marker allele (e.g., 5′FP class 1) associated with the disease. The test procedure compares (*a*) the number of times that such heterozygous parents transmit the associated marker to an affected offspring with (*b*) the number of times that they transmit the alternate marker allele. Because of this focus on alleles transmitted to affected offspring, the test shares some features with the concept of haplotype relative risk (HRR; Falk and Rubinstein 1987) and with the AFBAC test of association (Field et al. 1986; Thomson et al. 1989; Field 1991) described above. However, because our emphasis is on testing for linkage, the actual tests are different. Since GAW5 (Spielman et al. 1989), the principle underlying this linkage test has been used explicitly (McGinnis et al. 1991) or implicitly (Owerbach et al. 1990; Julier et al. 1991) in other investigations, to provide additional evidence that determinants of IDDM are located in the insulin gene region.

In GAW5, Ott presented the formal theory which is necessary for any test of a hypothesis based on a comparison of frequencies of marker alleles transmitted or not transmitted to affected offspring. His analysis showed that the probabilities of the various possible combinations of transmitted and nontransmitted marker locus alleles are determined by the association (disequilibrium) parameter δ and the recombination fraction θ between the loci. However, we show below that the χ^2 procedure used as a test of association (i.e., AFBAC) is not, in general, valid as a test of linkage, and we derive a procedure which is valid. We also show (1) that our testing procedure also provides a test for association between the two loci (indeed, the test can detect linkage only if association exists); (2) the relationship of this test to tests based on sharing of haplotypes or genes (identical by descent) in ASPs, affected sib trios, etc.; and (3) the result of applying this test to data on the 5′FP in IDDM.

The Transmission Test for Linkage Disequilibrium

The transmission/disequilibrium test (TDT) considers parents who are heterozygous for an allele associated with disease and evaluates the frequency with which that allele or its alternate is transmitted to affected offspring. Compared with conventional tests for linkage, the TDT has the advantage that it does not require data either on multiple affected family members or on unaffected sibs. However, the TDT has the disadvantage that it can detect linkage between the marker locus and the disease locus only if association (due to linkage *disequilibrium*) is present.

In the following sections we describe the properties of the TDT as a test of significance for linkage. We then discuss the relationship of the TDT to tests of linkage that are based on shared haplotypes in ASPs.

We assume a disease locus D, with disease allele D_1

Table 1

Marker Alleles M_1 and M_2 among the $2n$ Transmitted and $2n$ Nontransmitted Alleles in n Families Whose Single Child Is Affected

	M_1	M_2	Total
Transmitted	w	$2n-w$	$2n$
Nontransmitted	y	$2n-y$	$2n$
Total	$w+y$	$4n-w-y$	$4n$

and a normal allele D_2, and a marker locus with codominant alleles M_1 and M_2. No assumptions are made about dominance at the D locus. In the Discussion, we consider marker loci with more than two alleles.

One Affected Child per Family

It simplifies the discussion to consider first those families with one child only, that child being affected. (For such families, of course, haplotype-sharing tests, which require data from ASPs, affected sib trios, etc., cannot be used.)

Suppose that we have a sample of n such single-child families. In these families there will be, at the marker locus M, a total of $4n$ parental alleles, $2n$ of which are transmitted and $2n$ of which are not transmitted. The data on marker alleles in the affected children can be set up as in table 1. If no restrictions concerning genotype (see below) were placed on the parents contributing data for the test, the customary χ^2 test of significance for these data (e.g., see Falk and Rubinstein 1987; Thomson et al. 1989) would be carried out by using the standard statistic for a 2×2 table, equivalent to:

$$4n(w-y)^2/[(w+y)(4n-w-y)]\,, \qquad (1)$$

with 1 df.

There are three hypotheses for which one might consider using expression (1) as a test. These are (i) no association between marker and disease, $\delta = 0$ (i.e., the hypothesis tested by AFBAC); (ii) no linkage between marker and disease, $\theta = ½$ (i.e., $1-2\theta = 0$); and (iii) either hypothesis (i) or hypothesis (ii) or both, $\delta(1-2\theta) = 0$. (This last situation corresponds to the hypothesis of HRR = 1 [Ott 1989].) Drawing on the analysis of Ott (1989), we demonstrate below that expression (1) is a valid test statistic for only the first of these hypotheses. To show why this is so, we rewrite the data of table 1 in the form of table II of Ott (1989) to give our table 2. Here we focus on the $2n$ parents (rather than on the $4n$ parental genes) and describe each parent in terms of both the marker-locus allele that he or she transmits to the affected child and the allele that he or she does not transmit. Suppose that the population frequency of M_1 is m, that the frequency of D_1 is p, that the coefficient of linkage disequilibrium [freq(M_1D_1)$-mp$] is δ, and that θ is the recombination fraction between the M and D loci. Ott (1989) showed that, for a recessive disease, the probabilities corresponding to the four cells of table 2 are as shown in our table 3.

Test statistic (1) compares the values of w ($=a+b$ in table 2) and y ($=a+c$ in table 2)—that is, the frequency of M_1 among transmitted and nontransmitted alleles. Ott (1989) commented that statistic (1) does not provide a valid χ^2 test in all circumstances. Use of this statistic assumes independence of the allelic types of transmitted and nontransmitted genes, for all parents. Such independence will hold if and only if each probability in table 3 is the product of the corresponding marginal probabilities, and algebraic manipulation shows that this requires $\theta\delta = 0$. Thus the only hypotheses for which expression (1) provides a valid test are $\delta = 0$ or $\theta = 0$.

The hypothesis of interest to us is $\theta = ½$, and thus expression (1) does not provide a valid test. (Nor is it a valid test for HRR = 1.) We can construct a valid test from expression (1) by adding appropriate covariance terms, and this leads, after some algebra, to our test statistic (3) below. It is more straightforward, however, to argue directly. Table 3 shows that the only data values in table 2 that bear on θ are b and c. This observation implies that only the data from heterozygous M_1M_2 parents should be used in the test, as might be expected from conventional tests of linkage. Table 3 shows that, when $\theta = ½$, we have $E(b) = E(c)$, whatever the values of m, p, and δ. Now any χ^2 with 1 df must be of the form $\chi^2 = (u-v)^2/\mathrm{Var}(u-v)$, where the expected values $E(u)$ and $E(v)$ are equal under the hypothesis be-

Table 2

Combinations of Transmitted and Nontransmitted Marker Alleles M_1 and M_2 among $2n$ Parents of n Affected Children

	Nontransmitted Allele		
Transmitted Allele	M_1	M_2	Total
M_1	a	b	$a+b$
M_2	c	d	$c+d$
Total	$a+c$	$b+d$	$2n$

Table 3

Probabilities of Combinations of Transmitted and Nontransmitted Marker Alleles M_1 and M_2 among 2*n* Parents of Affected Children

Transmitted Allele	Nontransmitted Allele		
	M_1	M_2	Total
M_1	$m^2+(m\delta/p)$	$m(1-m)+[(1-\theta-m)\delta/p]$	$m+[(1-\theta)\delta/p]$
M_2	$m(1-m)+[(\theta-m)\delta/p]$	$(1-m)^2-[(1-m)\delta/p]$	$1-m-[(1-\theta)\delta/p]$
Total	$m+(\theta\delta/p)$	$1-m-(\theta\delta/p)$	1

ing tested, and (often) the denominator is a variance *estimate* rather than a known variance. Thus the appropriate χ^2 test statistic is of the form

$$(b-c)^2/[\text{estimated variance of } (b-c)]\,. \qquad (2)$$

In calculating the denominator in χ^2 test statistic (2), we note that the contributions from two heterozygous parents are independent (when $\theta = 1/2$). Thus the χ^2 test statistic (2) is the standard approximation to a binomial test of the equality of two proportions, when data from all heterozygous parents are used. This test is sometimes referred to as "McNemar's test" (Sokal and Rohlf 1969, p. 612), and in the present case it takes the form of the χ^2 statistic

$$\chi^2 = (b-c)^2/(b+c). \qquad (3)$$

Note that this is equivalent to a more familiar form of χ^2—namely, $\Sigma[(O-E)^2/E]$, with O for "observed" and with $E = [b+c]/2$ for "expected." An exact binomial test can be used, if desired, instead of expression (3).

We call χ^2 statistic (3) the "transmission/disequilibrium χ^2," or "TDT," and denote it as "χ^2_{td}." We use χ^2_{td} as a test for linkage between the D and M loci, but table 3 shows that χ^2_{td} depends on both linkage and linkage disequilibrium ($\delta > 0$), so it is useful only when there is disequilibrium between these loci. This χ^2 will reappear below in cases where more than one child in a family is affected.

We now make several observations about the use of χ^2_{td} as a test statistic for $\theta = 1/2$. First, we have carried out our calculations by assuming a recessive disease. It is easy to show that expression (3) provides an appropriate χ^2 test of linkage whatever the penetrance values and ascertainment procedure, implying that in all cases only heterozygous M_1M_2 parents should be used in the test. We will show later that χ^2_{td} is appropriate also when there are several affected children in a family.

Second, use of statistic (3) implicitly assumes that there is no segregation distortion at the M locus. If the possibility of segregation distortion exists, an appropriate alternative to using statistic (3) is to use data from unaffected children of M_1M_2 parents as well as from affected children and to use a standard 2 × 2-table χ^2 test statistic of equal frequency of transmission of M_1 to affected and unaffected (Owerbach et al. 1990; Parsian et al. 1991). This approach is illustrated below in Results.

Third, a further hypothesis which one might wish to test is that the HRR of Falk and Rubinstein (1987) is 1. This hypothesis is equivalent to $\delta(1-2\theta) = 0$, and it can be shown that the appropriate test statistic is also expression (3).

Fourth, we have used implicitly, in expression (3), the fact that, when $\theta = 1/2$, the contributions from male and female heterozygote parents to affected children are independent. It is straightforward to show that this procedure is justified.

Fifth, we note that we have shown that there are two reasons why expression (1) is not a valid χ^2 for testing the hypothesis $\theta = 1/2$. The first reason is that data from homozygous parents must not be used. The second reason is that, even if data from heterozygous parents only are used, expression (1) is still invalid as a test for $\theta = 1/2$. The covariances have not been accounted for in expression (1), and since with heterozygous parents there is a correlation of -1 between the allelic types of the transmitted and nontransmitted genes, use of expression (1) leads to a χ^2 exactly twice the correct value as given by expression (3).

Sixth, the probabilities in table 3 show that, in general, statistic (3) increases as δ increases, so that linkage is detected more readily. This is in accord with a similar observation by Clerget-Darpoux (1982), who noted that in a lod score analysis, the lod score also increases as δ increases.

Two Affected Children per Family

It is instructive to consider the case of families with two children, both of whom are affected. This will allow not only a straightforward extension of expression (3) but also a comparison to be made of χ^2_{td} and a conventional test for linkage between D and M loci, the χ^2 for haplotype sharing by affected sibs, which we denote as "χ^2_{hs}."

In line with the comments made above, only heterozygous M_1M_2 parents are informative. Suppose that, in the families being considered, there are h such parents. If the mother and father in a family are both M_1M_2, then we simply count them separately, since their contributions are independent (Ott 1989). (In a subset of these families, usually a small proportion, both children will also be heterozygous, and these families pose a problem for haplotype sharing but not for the TDT. Unless some additional typing allows discrimination between paternal and maternal alleles that are "alike in state," sharing will be ambiguous, and such families must be ignored for the analysis of allele [haplotype] sharing.)

Consider then the alleles transmitted to the affected children from each of these h heterozygous parents. The information obtained can be summarized by defining the following three categories:

$$\begin{aligned} i &= \text{number of parents who transmit } M_1 \text{ to both children ;} \\ h-i-j &= \text{number of parents who transmit } M_1 \text{ to one child and } M_2 \text{ to the other ;} \\ j &= \text{number of parents who transmit } M_2 \text{ to both children .} \end{aligned} \tag{4}$$

We have noted that the only relevant data in table 2 are b and c. In terms of the data in definitions (4) above, we can write the quantities in table 2 as

$$\begin{aligned} b &= 2i+(h-i-j) = h+i-j\,, \\ c &= 2j+(h-i-j) = h-i+j\,, \end{aligned} \tag{5}$$

so that $b-c = 2i-2j$ and $b+c = 2h$. Thus, when the format of expression (3) is used, the transmission/disequilibrium χ^2 is

$$\chi^2_{td} = 2(i-j)^2/h\,. \tag{6}$$

This can be used immediately as a test statistic for linkage between the D and M loci. (This remark uses the fact that, for two heterozygous parents, the contributions to any one affected child are independent when $\theta = 1/2$, as are their contributions to all their children, both affected and unaffected.) If the possibility of segregation distortion at the M locus exists, a valid test of $\theta = 1/2$ is obtained by comparing the frequency of transmission of M_1 from heterozygous parents to affected and unaffected children, as described above.

For families with two or more affected sibs, one standard approach to testing for linkage uses the total number of haplotypes identical by descent ("mean haplotype sharing" of Blackwelder and Elston 1985). This measure ignores the allelic state of the marker allele and simply compares (*a*) the number of times a parent transmits the identical marker allele to both members of a pair of affected children with (*b*) the number of times that parent transmits different alleles. Thus the χ^2 for haplotype sharing compares $i+j$ with $h-i-j$, and the χ^2 value for this comparison is

$$\begin{aligned} \chi^2_{hs} &= [i+j-(h-i-j)]^2/h \\ &= (2i+2j-h)^2/h\,. \end{aligned} \tag{7}$$

What is the relation between this χ^2 and that in equation (6), both of which can be used to test for linkage between D and M loci? When the null hypothesis (D and M loci are unlinked) is true, the three categories defined above have respective probabilities 1/4, 1/2, and 1/4. Thus the "total" χ^2 for these three categories, with 2 df, is

$$\begin{aligned} \chi^2_{Total} &= \frac{[i-h/4]^2}{h/4} + \frac{[h-i-j-h/2]^2}{h/2} \\ &\quad + \frac{[j-h/4]^2}{h/4}\,. \end{aligned} \tag{8}$$

Algebra shows that this is simply the sum of the two χ^2's (6) and (7). In other words, expressions (6) and (7) use the total data in two mutually exclusive and independent ways to test the same hypothesis (D and M loci unlinked). Thus we can think of the χ^2_{td} as using one part of the data in definitions (4) and can think of χ^2_{hs} as using the rest. They are thus complementary tests, and we may use one or the other or both (see Discussion).

It is interesting to consider the relation between the three χ^2 statistics (6)–(8) and the t_2 and Y statistics described by Blackwelder and Elston (1985) for testing for linkage between the D and M loci. Blackwelder and Elston are concerned with general tests for linkage and therefore do not restrict attention, as we do, to parents

Table 4

Allelic Identity of Sibs Who Are Offspring of Fully Informative Parents

	Second Sib			
First Sib	M_1M_1	M_1M_2	M_2M_1	M_2M_2
M_1M_1	n_{11}	n_{12}	n_{13}	n_{14}
M_1M_2	n_{21}	n_{22}	n_{23}	n_{24}
M_2M_1	n_{31}	n_{32}	n_{33}	n_{34}
M_2M_2	n_{41}	n_{42}	n_{43}	n_{44}

Note.—The first allele in each genotype is paternally derived, and the second is maternally derived.

heterozygous for an allele that shows a population association. However, their χ^2 statistics are completely applicable to our situation. We consider the $n = h/2$ families in the sample where sharing by affected offspring of an M_1M_2 parent is unambiguous, and we make the subdivision shown in table 4. The relation between the values in table 4 and those in definitions (4) above is

$$i = 2n_{11}+n_{12}+n_{13}+n_{21}+n_{22}+n_{31}+n_{33} \,, \qquad (9)$$

$$\begin{aligned} j &= n_{22}+n_{24}+n_{33}+n_{34}+n_{42}+n_{43}+2n_{44} \,, \\ h &= 2\Sigma\Sigma n_{ij} = 2n \,. \end{aligned} \qquad (10)$$

Blackwelder and Elston focus attention on statistics calculated from the number r_k of families where the two affected sibs share k ($k = 0,1,2$) parental marker genes. In terms of the values in table 4, these statistics are

$$\begin{aligned} r_0 &= n_{14}+n_{23}+n_{32}+n_{41} \,, \\ r_1 &= n_{12}+n_{13}+n_{21}+n_{24}+n_{31}+n_{34}+n_{42}+n_{43} \,, \\ r_2 &= n_{11}+n_{22}+n_{33}+n_{44} \,. \end{aligned}$$

The Blackwelder and Elston statistic t_2 is defined as

$$\begin{aligned} t_2 &= (r_1+2r_2-n)(2/n)^{1/2} \\ &= (r_2-r_0)(2/n)^{1/2} \,, \end{aligned} \qquad (11)$$

and this can be written as

$$(n_{11}+n_{22}+n_{33}+n_{44}-n_{14}-n_{23}-n_{32}-n_{41})(2/n)^{1/2} \,.$$

Using equations (6), (10), and (11), one can show that the square of this quantity is χ^2_{hs}; in other words, use of t_2 is equivalent to use of χ^2_{hs}.

The statistic Y—which Blackwelder and Elston also consider as a test of linkage—is defined, for ASPs, as

$$Y = 4[(r_2-n/4)^2+\tfrac{1}{2}(r_1-n/2)^2+(r_0-n/4)^2]/n \,. \qquad (12)$$

Although this statistic has the same general form as the total χ^2 statistic (8), the two statistics are different. Both statistics have χ^2_{hs} as a component, but the total χ^2 (8) has χ^2_{td} as its other component and therefore takes into account *which* allele is preferentially found in affected offspring, while the statistic Y (12) does not have this property.

More than Two Affected Children in a Family

The principles outlined above also hold for families with more than two affected children. Consider, for example, families with three affected children. We consider only heterozygous M_1M_2 parents, as above. Each parent may be classified into one of four categories (see below); we suppose that $s+t+u+v = w$. The probabilities under the null hypothesis (no linkage) are also shown:

s = number of parents who transmit M_1 to all three children (probability 1/8) ;

t = number of parents who transmit M_1 to two children and M_2 to one child (probability 3/8) ;

u = number of parents who transmit M_1 to one child and M_2 to two children (probability 3/8) ;

v = number of parents who transmit M_2 to all three children (probability 1/8) .

The total χ^2, analogous to χ^2 statistic (8), is

$$\frac{8}{w}\left[(s-w/8)^2+\frac{1}{3}(t-3w/8)^2 + \frac{1}{3}(u-3w/8)^2+(v-w/8)^2\right] . \qquad (13)$$

The transmission/disequilibrium χ^2 (a component of the total) is $\chi^2_{\mathrm{td}} = (3s+t-u-3v)^2/3w$. The haplotype-sharing χ^2 is $\chi^2_{\mathrm{hs}} = (3s-t-u+3v)^2/3w$. There is also a residual χ^2: $\chi^2_{\mathrm{r}} = 3(s-t+u-v)^2/3w$. The sum of these three χ^2's is the total χ^2. Since the residual χ^2 has little obvious interpretation, the total χ^2 should not be used as a test statistic for linkage, since 1 df is used in it for no apparent purpose. The transmission/disequilibrium and haplotype-sharing χ^2's can be used, separately or together, to test for linkage. These considerations also generalize to sibships with four or more affected.

We have shown above how a χ^2 statistic to test transmission/disequilibrium can be calculated for data in which all families have the same number of affected children. In any real set of data, we can expect to observe families with varying numbers of affected children. In such a case we recommend simply combining all affected children in the data, irrespective of number of affected in the family, in one overall transmission/disequilibrium χ^2 statistic of the form $(B-C)^2/(B+C)$, where B is the total number of transmissions of M_1 to affected children and C is the total number of transmissions of M_2. In the case where segregation distortion at the M locus is a possibility, an aggregate 2×2-table χ^2 is appropriate, corresponding to that discussed above for the case of one affected child per family. We use such a χ^2 procedure below in the Results subsection.

Data and Results

Data

The data for this study were assembled for GAW5 from 94 families with two or more IDDM children (Baur et al. 1989; Spielman et al. 1989). For GAW5, Southern blots of genomic DNA digested with *Pvu*II were hybridized with phins 310 (Bell et al. 1984), a probe specific for the 5′FP, and alleles were assigned by eye to one of three classes corresponding to fragment size. (Class 1 is smallest, and class 3 is largest.) Gel positions of genomic bands and markers were also recorded; for the present reanalysis we assigned restriction fragments to allele class 1 if they were smaller than 1 kb, to class 2 if they were 1–2 kb, and to class 3 if they were larger than 2 kb. Since our analysis focuses on the role of class 1 alleles, class 2 and class 3 alleles were grouped together as class X. Among the 94 families, there were 53 in which at least one parent was heterozygous for class 1 and class X alleles.

Results

In order to demonstrate the usefulness of the TDT, we review the findings with respect to population association and haplotype sharing. The family data obtained for GAW5 do not lend themselves to a conventional association study, which would include unrelated controls. However, when just unrelated diabetics (the oldest affected sib in each family) are considered, the frequency of class 1 alleles in the present (GAW5) data is 138/162 = .85. This value is similar to those reported for "random" diabetics and is higher than that found in unrelated controls, as has been observed elsewhere (Cox et al. 1988).

Table 5

TDT for Alleles 1 and X of 5′FP in IDDM: Data for 1/X Parents of All Affected Children

	No. of Alleles Transmitted				
	1	X	Total	χ^2_{td}	Significance (p)
Observed	78	46	124	8.26	.004
Expected	62	62			

An analysis of haplotype sharing in the GAW5 family data was previously carried out by Cox and Spielman (1989). Using the χ^2 test statistic of equation (12) ("*Y*" of Blackwelder and Elston [1985], applied strictly to ASPs), Cox and Spielman (1989) did not find even modest departures from random sharing. This result also held when they considered only families with at least one parent heterozygous for class 1/class 3 at the 5′FP. (For the corresponding test by equation [7] or t_2 of Blackwelder and Elston [1985], see table 7 below.) Thus there is population association but no evidence for linkage, by conventional tests.

However, when linkage is tested by the TDT, a different conclusion emerges (table 5). There were 57 parents heterozygous for alleles 1 and X of the 5′FP; these parents transmitted 124 alleles (78 class 1 alleles and 46 class X alleles) to their diabetic offspring. Under the hypothesis of no linkage, the expected number of transmissions of 1 and X is equal (i.e., 62). When equation (5) is used, the difference observed is highly significant; $\chi^2_{td} = (78-46)^2/124 = 8.26$, $p = .004$.

As explained above, the difference found with the TDT could be due to an "artifact" of meiotic segregation distortion, which would be expected to apply to both affected and unaffected offspring, if unrelated to disease. For this reason, we compared affected and unaffected offspring with respect to transmitted class 1 and class X alleles. The results are shown in table 6. Among affected offspring, 78 (63%) of 124 alleles received from heterozygous parents were class 1. The corresponding figure for unaffected offspring was 42 (40%) of 104; the difference is highly significant ($\chi^2_1 = 11.5, p < .001$). This result confirms the finding of linkage; there is no evidence for segregation distortion.

Table 6

Comparison of Alleles I and X of 5′FP Transmitted to IDDM-affected Offspring and Unaffected Sibs

	No. of Alleles Transmitted				
	1	X	Total	χ^2_{td}	Significance (p)
Affected	78	46	124	11.5	<.001
Unaffected	42	62	104		

Note.—Data for 1/X parents.

The strong evidence for linkage, based on the TDT, stands in striking contrast to conclusions obtained, in earlier studies, from the Y statistic for haplotype sharing. However, the TDT (above) and the Y statistic were based on overlapping but not identical sets of families. This discrepancy arose because families with only one parent heterozygous 1/X could be used for the TDT but not for the Y statistic. Furthermore, parents with two distinguishable class 1 alleles were used for Y but not for the TDT.

These differences in the data led us to ask the following question: Is the failure to find linkage with the Y statistic due entirely to the difference between the samples used, or are linkage tests based on haplotype sharing inherently less sensitive than the TDT for the present data? To answer this question, we applied the transmission/disequilibrium (χ^2_{td}) and haplotype-sharing (χ^2_{hs}) tests to exactly the same data. Not all the data from table 5 can be used, because some are from simplex families or from sibships with more than two affected sibs. Accordingly, we used just those families with at least one 1/X parent and exactly two affected sibs, as appropriate for equations (4)–(8). Table 7 shows the data in the form of definitions (4).

For the TDT we compare i with j, by equation (6), and obtain $\chi^2_{td} = 3.60$ ($p = .058$). Unlike the corresponding test above ($\chi^2_{td} = 8.26$), the present comparison is not "quite" significant. Although the proportion of class 1 alleles transmitted here ($54/90 = .60$) is almost the same as that in table 5 ($78/124 = .63$), the χ^2_{td} is smaller, and the significance level is less striking, because of the smaller sample size.

For the haplotype-sharing test, we compare (*a*) the number of parents ($i+j = 21$) who transmitted the same allele (1 or X) to both affected children with (*b*) the number ($h-i-j = 24$) who transmitted different alleles. This is equivalent to comparing the number of ASPs who received the same allele ("shared") with the number who received different alleles ("unshared"). The resulting χ^2_{hs} (0.20) is not significant, and the difference is in the *opposite* direction of that predicted by linkage, presumably reflecting random variation. There is not even a "trend" toward increased sharing.

Thus, in the present analysis of a single body of data, we see the discrepancy identified in earlier reports. There is a population association between IDDM and the class 1 allele of the 5′FP, but sharing of alleles by affected sibs (cosegregation) provides no evidence for linkage. Nevertheless, there is highly significant evidence of linkage in the TDT.

Discussion

Linkage studies for so-called complex genetic diseases pose problems not found in standard linkage analysis. Because these diseases, in general, have reduced penetrance, unaffected family members usually provide much less information for linkage than do affected members. In this situation, it is essential to study families with multiple affected members and to focus on affected relatives, such as ASPs. The ASP approach has been applied with great success to unravel the role of the HLA complex in several diseases to which HLA appears to make a large contribution. For a locus that makes a modest contribution, however, the approach is severely limited. It has been shown by computer simulation (Cox and Spielman 1989) that, when ASPs are used, the power to detect linkage to such a locus is very modest and may require hundreds of ASPs. Furthermore, an additional consequence of the low penetrance

Table 7

Transmission from 45 1/X Parents of IDDM-affected Sib Pairs

No. of 1/X Parents Who Transmit			
Class 1 to Both Children	Class 1 to One Child and Class X to the Other	Class X to Both Children	Total
$i = 15$	$h-i-j = 24$	$j = 6$	$h = 45$

Note.—Data are for comparison of χ^2_{td} (TDT) and χ^2_{hs} (haplotype sharing).

is that the high-density pedigrees ideal for linkage analysis are usually rare.

There is an alternative approach that has received much attention and does not require family studies: testing for statistical or "population" association, due to linkage disequilibrium, between disease and genetic marker. Association can be demonstrated, if it exists, by comparing allele frequencies at the marker locus in random samples of unrelated patients and controls. These studies are logistically much simpler than conventional linkage studies requiring data from whole families of ASPs, because a single affected member of each family is studied. This kind of study has become popular in complex diseases (e.g., see, among many others, Hoover et al. 1986; Breslow 1988; Li et al. 1988; Cox and Bell 1989; Comings et al. 1991).

However, as indicated in the Introduction, association between marker and disease at the population level can occur without linkage—for example, as the result of stratification or intrapopulation heterogeneity in allele frequency, and this is a potentially fatal flaw in the method. Accordingly, it is desirable to use a method that combines the advantages of the linkage approach (segregation in families) and the population approach (not requiring multiple affected relatives).

The TDT described here has these properties. To our knowledge, the principle underlying the method was first proposed (independently) by Rubinstein et al. (1981) and Weitkamp (personal communication). In the form of the AFBAC test (Thomson 1988), a related method has been used as a test of association (Field et al. 1986; Falk and Rubinstein 1987; Field 1989, 1991; Thomson et al. 1989). The method of the TDT has been used as a test for linkage between IDDM and the insulin gene region (Owerbach et al. 1990; Julier et al. 1991; McGinnis et al. 1991) and in other disease-marker studies (Parsian et al. 1991). However, the statistical/genetic properties of this method *as a test for linkage* have not been investigated formally, and the underlying principle has been used in various ways by different investigators. In the present paper, we have presented the correct χ^2 test for linkage and have shown how it is related to conventional tests based on identity by descent in ASPs.

Advantages and Disadvantages of the TDT

In genetic studies of "complex" diseases, the finding of population association with a marker is usually taken to suggest involvement of a nearby gene. The next step is often to test for sharing of parental alleles (or haplotypes) in ASPs, to confirm (or rule out) linkage. Our present analysis, as well as previous attempts by others to demonstrate linkage in this and other complex diseases, show that it may be very difficult to detect linkage in ASPs, even when it is present.

Our findings suggest that there are significant advantages in using the TDT instead. First, as we show here, the TDT may be much more sensitive than haplotype-sharing tests, with the same data. Second, unlike the TDT, haplotype-sharing tests require multiplex sibships, and, even for relatively common complex diseases, multiplex families may be difficult to ascertain. In contrast, simplex families are usually easier to ascertain, and the TDT can be based on them exclusively. However, if data are also available from multiplex families, the corresponding transmission data can simply be combined with those from the simplex families for analysis by the TDT, a further advantage. Third, as an extension of the preceding, transmission data from multiplex families with different numbers of affected sibs can simply be pooled directly, without the statistical complexities that arise when this is done in haplotype-sharing tests. Finally, it has been suggested that patients in multiplex families may have alleles of disease genes (or numbers of disease genes) that make them somewhat distinct from patients in simplex families. Since the majority of patients with low-penetrance disorders are from simplex families, conclusions based on multiplex families may have restricted applicability.

However, it is important to note again that the TDT will not detect linkage between disease and marker unless there is also population association (linkage disequilibrium). This is an absolute requirement; as shown in table 3 (and by Ott 1989), even very close linkage will not be detected by the TDT if $\delta = 0$. Consequently, the usual use of the TDT will be in cases where an association has already been found, and the goal is to establish linkage.

Multiple Alleles at the Marker Locus

The TDT can be generalized to a marker locus with more than two alleles. Consider three alleles—M_1, M_2, and M_3. Suppose first that population data suggest that both M_1 and M_2 are (positively) associated with disease. In this case, we test both relevant heterozygotes by the TDT; that is, we examine transmission of M_1 to affected offspring by M_1M_3 parents and transmission of M_2 from M_2M_3, by using, in each case, statistic (3). If there is evidence that only M_1 is associated with susceptibility, then we can examine transmission of M_1 from M_1M_2 parents and from M_1M_3 parents, again by using, in each case, statistic (3). This approach can be generalized to an arbitrary number of alleles.

Power of the TDT

It may be puzzling that the TDT and the haplotype-sharing test, both testing for linkage in the same data, give such different results (table 7). In explanation we note that, as shown by the partition of the χ^2_{Total} (eq. [8] or eq. [13]), the two χ^2's are independent, so their numerical values are not necessarily correlated.

This observation, however, reveals nothing about the relative *magnitudes* of the two χ^2's. For example, in the data presented here, the TDT detects an effect that is not even suggested by the haplotype-sharing test. Under what circumstances will χ^2_{td} be appreciably larger than χ^2_{hs}? We have begun to investigate the relative power of the two tests in a more general way. We have found (R. E. McGinnis, W. J. Ewens, and R. S. Spielman, in preparation) that it is possible to rewrite the χ^2's for transmission/disequilibrium (eq. [6]) and haplotype sharing (eq. [7]) and compare their expected magnitudes under various genetic assumptions. The results of these analyses indicate that the TDT will provide a powerful test for linkage in a wide variety of diseases that show association with a genetic marker.

Acknowledgments

We thank Susan Hodge and Glenys Thomson for helpful comments on the manuscript. This work was supported by NIH grants DK35047 and GM21135.

References

Baur MP, Fimmers R, Fritsche C, Hümmelink M, Neugebauer M, Acton RT, Hodge TW, et al (1989) Genetic analysis of IDDM: the GAW5 multiplex family dataset. Genet Epidemiol 6:15–20

Bell GI, Horita S, Karam JH (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 33:176–183

Blackwelder WC, Elston RC (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. Genet Epidemiol 2:85–97

Breslow JL (1988) Apolipoprotein genetic variation and human disease. Physiol Rev 68:85–132

Clerget-Darpoux F (1982) Bias of the estimated recombination fraction and lod score due to an association between a disease gene and a marker gene. Ann Hum Genet 46:363–372

Comings DE, Comings BG, Muhleman D, Dietz G, Shahbahrami B, Tast D, Knell E, et al (1991) The dopamine D_2 receptor locus as a modifying gene in neuropsychiatric disorders. JAMA 266:1793–1800

Cox NJ, Baker L, Spielman RS (1988) Insulin-gene sharing in sib pairs with insulin-dependent diabetes mellitus: no evidence for linkage. Am J Hum Genet 42:167–172

Cox NJ, Bell GI (1989) Disease associations: chance, artifact or susceptibility genes? Diabetes 38:947–950

Cox NJ, Spielman RS (1989) The insulin gene and susceptibility to IDDM. Genet Epidemiol 6:65–69

Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 51:227–233

Ferns GAA, Hitman GA, Trembath R, Williams L, Tarn A, Gale EA, Galton DJ (1986) DNA polymorphic haplotypes on the short arm of chromosome 11 and the inheritance of type 1 diabetes mellitus. J Med Genet 23:210–216

Field LL (1989) Genes predisposing to insulin-dependent diabetes in multiplex families. Genet Epidemiol 6:101–106

——— (1991) Non-HLA region genes in insulin dependent diabetes mellitus. Ballière's Clin Endocrinol Metab 5:413–438

Field LL, Fothergill-Payne C, Bertrams J, Baur MP (1986) HLA-DR effects in a large German IDDM dataset. Genet Epidemiol Suppl 1:323–328

Hitman GA, Tarn AC, Winter RM, Drummond V, Williams LG, Jowett NI, Bottazzo GF, et al (1985) Type I (insulin-dependent) diabetes and a highly variable locus close to the insulin gene on chromosome 11. Diabetologia 28:218–222

Hoover ML, Angelini G, Ball E, Stastny P, Marks J, Rosenstock J, Raskin P, et al (1986) HLA-DQ and T-cell receptor genes in insulin-dependent diabetes mellitus. Cold Spring Harb Symp Quant Biol 51:803–809

Julier C, Hyer RN, Davies J, Merlin F, Soularue P, Briant L, Cathelineau G, et al (1991) Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. Nature 354:155–159

Li SR, Baroni MG, Oelbaum RS, Stock J, Galton DJ (1988) Association of genetic variant of the glucose transporter with non-insulin-dependent diabetes mellitus. Lancet 2:368–370

McGinnis RE, Spielman RS, Ewens WJ (1991) Linkage between the insulin gene (IG) region and susceptibility to insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet Suppl 49:A476

Ott J (1989) Statistical properties of the haplotype relative risk. Genet Epidemiol 6:127–130

Owerbach D, Gunn S, Gabbay KH (1990) Multigenic basis for type I diabetes: association of HRAS1 polymorphism with HLA-DR3, DQw2/DR4, DQw8. Diabetes 39:1504–1509

Parsian A, Todd RD, Devor EJ, O'Malley KL, Suarez BK, Reich T, Cloninger CR (1991) Alcoholism and alleles of the human D_2 dopamine receptor locus. Arch Gen Psychiatry 48:655–663

Rubinstein P, Walker M, Carpenter C, Carrier C, Krassner J, Falk C, Ginsberg F (1981) Genetics of HLA disease associations: the use of the haplotype relative risk (HRR) and the "haplo-delta" (Dh) estimates in juvenile diabetes from three racial groups. Hum Immunol 3:384

Sokal RR, Rohlf FJ (1969) Biometry. WH Freeman, San Francisco

Spielman RS, Baur MP, Clerget-Darpoux F (1989) Genetic analysis of IDDM: summary of GAW5-IDDM results. Genet Epidemiol 6:43–58

Thomson G (1988) HLA disease associations: models for insulin dependent diabetes mellitus and the study of complex human genetic disorders. Annu Rev Genet 22:31–50

Thomson G, Robinson WP, Kuhner MK, Joe S (1989) HLA, insulin gene, and Gm associations with IDDM. Genet Epidemiol 6:155–160