Am. J. Hum. Genet. 66:1168–1172, 2000

Report

Gamete-Competition Models

Janet S. Sinsheimer,[1,2,3] John Blangero,[4] and Kenneth Lange[2]

Departments of [1]Human Genetics, [2]Biomathematics, and [3]Biostatistics, The University of California, Los Angeles; and [4]Department of Genetics, Southwest Biomedical Foundation, San Antonio

The gamete-competition model is an application of the Bradley-Terry model for ranking of sports teams. If allele *i* of a marker locus is assigned parameter $\tau_i > 0$, then the probability that a parent with heterozygous genotype *i/j* transmits allele *i* is $\Pr(i/j \to i) = \tau_i/(\tau_i + \tau_j)$. Mendelian segregation corresponds to the choice $\tau_i = 1$ for all *i*. To test whether Mendelian segregation is true, one can estimate the τ_i from pedigree data and perform a likelihood-ratio test under the constraint that one τ_i equals 1. Although this procedure generates an interesting method for performance of segregation analysis with a marker locus, its real promise lies in generalization of the transmission/disequilibrium test. Quantitative as well as qualitative outcomes can be considered. The gamete-competition model uses full pedigree data and gives an estimate of the strength of transmission distortion to affected children for each allele. Covariates are incorporated by rewriting of $\tau_i = \exp(\beta^t x_k)$, where β is a parameter vector and x_k is a covariate vector for the *k*th transmitted gamete. Examples of covariates include disease-severity indicators for the child, sex of the child, or repeat number for tandem-repeat alleles.

The transmission/disequilibrium test (TDT) has had a profound impact on the design and execution of studies in genetic epidemiology (Terwilliger and Ott 1992; Spielman et al. 1993). In this report, we propose a generalization of the TDT that allows the use of full pedigree data, covariate information, and quantitative as well as qualitative disease traits. Our generalization is based on the Bradley-Terry model of paired comparisons (Bradley and Terry 1952; Keener 1993). The Bradley-Terry model provides a flexible modeling technique that can be implemented in the software package MENDEL (Lange et al. 1988).

One of the primary uses of the Bradley-Terry model is to rank sports teams (Keener 1993). If team *i* from a league of *n* teams is assigned a parameter $\tau_i > 0$, then, under the assumed model, the probability that team *i* beats team *j* equals the ratio $\lambda_{ij} = \tau_i/(\tau_i + \tau_j)$. Once the τs are estimated by maximum likelihood, the teams can be ranked by the ordered estimates $\hat{\tau}_i$. Because multiplication of all τ_i by the same constant results in all ratios λ_{ij} being invariant, we arbitrarily constrain $\tau_k = 1$ for some team *k*. If team *i* beats team *j* a total of y_{ij} times during a season of play, then the log likelihood of the league amounts to

$$L(\tau) = \sum_{i,j} y_{ij}\{\ln \tau_i - \ln(\tau_i + \tau_j)\} \ .$$

It turns out that $L(\tau)$ can be maximized by the iteration

$$\tau_i^{n+1} = \frac{\sum_{j \neq i} y_{ij}}{\sum_{j \neq i} (y_{ij} + y_{ji})/(\tau_i^n + \tau_j^n)} \ . \qquad (1)$$

The likelihood can be maximized, provided that all teams lose at least once and that every pair of teams meet. The latter condition can be relaxed—for example, if both members of a pair play a third team. Algorithm (1) is an example of a minorize/maximize algorithm (Lange et al., in press).

Lange et al. (1988) and Jin et al. (1994) independently advocated the Bradley-Terry model for testing of Mendelian transmission of the alleles of an ordinary, nondisease marker. The *i*th allele is analogous to the *i*th team in the sports model. Allele *i* is assigned the segregation parameter $\tau_i > 0$, and the probability that a heterozygous parent *i/j* transmits allele *i* is expressed as $\Pr(i/j \to i) = \tau_i/(\tau_i + \tau_j)$. We arbitrarily set $\tau_k = 1$, where *k* denotes the

Received October 20, 1999; accepted December 8, 1999; electronically published February 25, 2000.

Address for correspondence and reprints: Dr. Janet Sinsheimer, Department of Biomathematics, UCLA School of Medicine, Los Angeles, CA 90095-1766. E-mail: janet@sunlab.ph.ucla.edu

0002-9297/2000/6603-0039$02.00

Table 1

Segregation at the PGM Locus

Allele	Frequency	$\hat{\tau}_i$	SD
1−	.6200	1.000	Fixed
1+	.1700	.7877	.2911
2−	.1400	.8395	.2418
2+	.0700	1.258	.5862

most frequent allele. This gamete-competition model has the advantage of being very parsimonious, compared with the classical procedure of constructing a Punnet square for each mating type. Low offspring counts and missing mating types plague the Punnet square procedure.

In our proposed gamete-competition model, all τs are estimated by maximum likelihood. Under the null hypothesis of Mendelian segregation, $\tau_i = 1$ for all i. The alternative hypothesis permits all but one of the τ_i to range over the interval $(0, \infty)$. The null hypothesis can be tested by computation of the likelihood-ratio statistic, $2 \times \ln[L(\hat{\tau})/L(1)]$, which asymptotically follows a χ^2 distribution with $n - 1$ df, n being the number of alleles. In principle, one can replace this large-sample approximation by comparing the observed statistic with simulated values of the statistic under random assignment of alleles to pedigree founders followed by Mendelian transmission from parents to children.

We can also use the gamete-competition model to test for biased transmission of marker alleles to affected individuals, reserving Mendelian transmission for transmission to normal individuals and to individuals of unknown phenotype. The resulting model is a generalization of the TDT model of Sham and Curtis (1995). This TDT version of the gamete-competition model deals with missing data in a straightforward manner and accommodates pedigrees of arbitrary size and complexity. The null hypothesis of $\tau_i = 1$ for all i is equivalent to no linkage and no association. When pedigrees are restricted to trios consisting of two genotyped parents and an affected child, the gamete-competition model reduces to the Sham and Curtis (1995) model. Regardless of the type of pedigrees analyzed, the τs can be estimated by maximum likelihood and the null hypothesis tested by a likelihood-ratio statistic.

Table 1 illustrates the testing of ordinary Mendelian segregation by the gamete-competition model. The data consist of 86 individuals spread over five families and typed at the four alleles of the PGM1 locus (Lewis et al. 1980). The likelihood-ratio statistic is 1.884, with 3 df (asymptotic $P = .597$), so the null hypothesis of Mendelian segregation cannot be rejected.

Our second example examines transmission distortion of a microsatellite marker within the sulfonyl urea receptor–1 (SUR) gene (Goksle et al. 1998) in offspring affected with non–insulin-dependent diabetes in 27 Mexican American families (74 genotyped affected offspring). For these disease data, the τs are estimated only from the outcomes of transmission to affected children. Mendelian transmission is assumed for unaffected children and children of unknown disease status. To avoid violations of the large-sample assumption implicit in the likelihood-ratio test, we combine the rare allele 1 with its neighboring allele 2. The maximum-likelihood estimates displayed in table 2 produce a likelihood-ratio statistic of 9.133 and lend support to the alternative hypothesis of transmission distortion to affected children (asymptotic $P = .043$, 9 df).

Our third example involves a 287-bp insertion/deletion polymorphism in the angiotensin-1–converting enzyme (ACE) gene. The deletion allele appears to be associated with high plasma ACE activity (Keavney et al. 1998). In the current sample, ACE activity was determined in 404 people in 69 families. For the purpose of this example, any person with a sex-adjusted ACE level in the top 20th percentile of the unselected population (65/315 genotyped offspring in the sample) is considered to be affected. Table 3 presents the allele frequencies and estimated transmission parameters. The likelihood-ratio statistic of 17.50 (asymptotic $P = 1.56 \times 10^{-5}$, 1 df) confirms the strong association of the deletion allele with high ACE activity.

The TDT examples presented thus far assume that transmission to unaffected offspring conforms to Mendelian segregation. We can relax this assumption and simultaneously estimate τs for both unaffected and affected offspring. Table 4 gives estimated transmission parameters for both affected (top 20th percentile) and unaffected (bottom 80th percentile) offspring. In this example, in which affection is directly related to an underlying quantitative trait, it is hardly surprising that the likelihood-ratio statistic of 33.70 (asymptotic $P = 4.81 \times 10^{-8}$, 2 df) computed from the estimates shown in table 4 is more significant than the likelihood-ratio statistic computed from table 3. It is possible to arrive

Table 2

Non–Insulin-Dependent Diabetes: Marker within SUR

Allele(s)	Frequency	$\hat{\tau}_i$	SD
1 and 2	.054	.288	.215
3	.210	1.00	Fixed
4	.190	.810	.447
5	.048	1.40	.985
6	.047	.697	.681
7	.108	.383	.204
8	.140	.556	.288
9	.091	.567	.322
10	.071	.499	.509
11	.042	.082	.104

Table 3

High Plasma ACE Activity

Allele	Frequency	$\hat{\tau}_i$	SD
Insertion	.512	1.00	Fixed
Deletion	.488	3.74	1.30

at the same conclusion more parsimoniously by adoption of a single set of transmission parameters for both affected and unaffected offspring. For transmission to unaffected offspring, we reverse the roles of $\tau_{\text{insertion}}$ and τ_{deletion}. The resulting model yields the estimate $\hat{\tau}_{\text{deletion}} = 2.028 \pm 0.265$ and a likelihood-test statistic of 30.83 (asymptotic $P = 2.88 \times 10^{-8}$, 1 df).

Diseases often involve severity levels. One can, accordingly, postulate different gamete-transmission parameters for each severity level. Suppose that we label, in the ACE data, the top 10th percentile as severely affected, the 10th–20th percentile as mildly affected, and the remaining lowest 80th percentile as unaffected. Table 5 shows the resulting parameter estimates. A likelihood-ratio statistic of 21.14 (asymptotic $P = 2.57 \times 10^{-5}$, 2 df) strongly suggests that τ_{severe} and τ_{mild} differ from 1. A likelihood-ratio statistic of 3.64 (asymptotic $p = .056$, 1 df) weakly rejects the null hypothesis that τ_{severe} equals τ_{mild} for the deletion allele.

The gamete-competition model can also be extended to quantitative traits. For example, we can replace τ_i in the transmission of allele i to person k by $\exp(\omega_i x_k)$, where x_k is the standardized trait value for person k and ω_i is a parameter attached to allele i. In this notation, a parent with heterozygous genotype i/j transmits allele i to k with probability $\Pr(i/j \rightarrow i|x_k) = 1/\{\exp[(\omega_j - \omega_i) \times x_k] + 1\}$. This parameterization highlights the correspondence of the TDT to conditional logistic regression (Sham and Curtis 1995; Collins and Morton 1998). Environmental covariates z_k can also be included—for example, by introduction of another parameter β and replacement of τ_i by

$$\exp[\omega_i(x_k + \beta z_k)] \qquad (2)$$

for transmission to person k. More generally, one can estimate a separate β_i for each allele i. Note that under the null hypothesis of Mendelian transmission, all $\omega_i = 0$ and β is undefined. The use of covariates is not restricted to quantitative traits. Equation (2) still applies to qualitative traits, provided that we take $x_k = 1$ for all affected k.

Table 6 treats these two models, using ACE activity levels and the insertion-deletion polymorphism with and without sex as a covariate. Although an increasing ACE level is highly correlated with transmission of the deletion allele (likelihood-ratio statistic of 82.76, asymptotic $P = 9.8 \times 10^{-20}$, 1 df), the hypothesis of a sex difference in transmission cannot be accepted (likelihood-ratio statistic of 1.72, asymptotic $P = .190$, 1 df).

Our final example illustrates the flexibility of the gamete-competition model in the testing of nonstandard models of gametic transmission. The histidine-rich glycoprotein (HRG) gene on chromosome 3q28-q29 has a dinucleotide-repeat polymorphism in its last intron (Hennis et al. 1994, 1995). The gamete-competition model of this last example with estimated ωs and no β shows a significant association between the 12 HRG alleles and serum concentration of HRG (likelihood-ratio statistic of 26.7, asymptotic $P = 3.7 \times 10^4$, 7 df after combination of neighboring rare alleles). Examination of the estimated ωs suggests that high repeat number leads to high serum concentration. This feature of the data is borne out by the significance ($P < .0001$) of the t statistic for the slope coefficient in the estimated regression line E(serum HRG) = $81.1 + 3.86 \times r$, where r denotes the sum of the maternal and paternal allele numbers of a person. To validate this trend via the gamete-competition model, we considered the linear gamete-competition model $\omega_i = i \times \alpha$ involving a single parameter α. This parsimonious model fits the data, as well as the full-gamete-competition model, with a separate ω_i for each allele or combined allele (asymptotic $P = .0718$). The estimated α is $.195 \pm .059$. A more complicated model with a separate α for females and males leads to no improvement.

In summary, the gamete-competition model is a very attractive generalization of the TDT. The null hypotheses differ slightly between the TDT and the gamete-competition model. The null hypothesis for the TDT applied to parent-offspring trios is no association or no linkage; the null hypothesis for the gamete-competition model is no association and no linkage. To the credit of the gamete-competition model, it accommodates quantitative traits and covariates. It also applies to arbitrary pedigrees, whereas the TDT applies only to parent-offspring trios or sibling sets. To the credit of the TDT, it controls for ethnic association, whereas the gamete-competition model fills in missing genotypes according to their pop-

Table 4

Estimation of Transmission in Affected and Unaffected Individuals

Disease Status and Allele	$\hat{\tau}_i$	SD
Unaffected:		
Insertion	1.00	Fixed
Deletion	.555	.082
Affected:		
Insertion	1.00	Fixed
Deletion	3.41	1.20

ulation frequencies. But to the extent that full typing prevails, the gamete-competition model is also insensitive to ethnic association. Because it may be applied to multiple affected individuals per pedigree, the gamete-competition model may lead to significant results in the absence of association when the marker and disease loci are linked. If the data consist of a handful of large pedigrees, the danger exists that particular marker alleles will be assigned causal effects, when, in fact, they are merely the marker alleles linked to the disease allele in a few relevant founders. Permutation versions of the TDT enjoy the advantage of not relying on large-sample approximations. In applying the gamete-competition model in the presence of rare alleles, one must exercise caution. Although there is no obvious permutation version of the gamete-competition model, one can, in principle, approximate the distribution of the likelihood-ratio test by dropping genes.

Finally, it is worth mentioning that both the TDT and the gamete-competition model extend to multiple markers. Lazzeroni and Lange (1998) discuss the TDT case. In the case of the gamete-competition model, consider two markers on different chromosomes, with alleles i and j at the first marker and alleles k and l at the second marker, where $i \neq j$ and $k \neq l$. One can parameterize the gamete-competition model by introducing transmission parameters $\tau_{i,k}$, $\tau_{i,l}$, $\tau_{j,k}$, and $\tau_{j,l}$ and writing

$$\Pr(i/j,k/l \rightarrow i,k) = \frac{\tau_{i,k}}{\tau_{i,k} + \tau_{i,l} + \tau_{j,k} + \tau_{j,l}} . \qquad (3)$$

If $i = j$, then

$$\Pr(i/i,k/l \rightarrow i,k) = \frac{\tau_{i,k}}{\tau_{i,k} + \tau_{i,l}} . \qquad (4)$$

Under the null hypothesis of Mendelian transmission, all τs equal 1. This model allows for interaction in the transmission of disease-associated alleles. It can be contrasted with the independent-transmission model parameterized by $\tau_{i,k} = \tau_i\tau_k$.

If the markers are linked with recombination fraction θ, then we must consider two parental haplotypes, (i,k) and (j,l), where, again, $i \neq j$ and $k \neq l$. The natural extension of the gamete transmission probability (3) is

$$\Pr[(i,k)/(j,l) \rightarrow (i,k)] = \frac{\tau_{i,k}(1-\theta)}{\tau_{i,k}(1-\theta) + \tau_{i,l}\theta + \tau_{j,k}\theta + \tau_{j,l}(1-\theta)} . \qquad (5)$$

If $i = j$, then transmission probability (4) applies. When computing likelihoods, using equation (5) and similar equations for the other transmission probabilities, one should take into account any linkage disequilibrium between the markers.

In conclusion, we hope that our practical examples and our recitation of the merits of the gamete-competition model will encourage its adoption as part of the standard repertoire of genetic epidemiologists. At the very least, we believe that further theoretical study is warranted.

Table 5

Very High and High ACE Activity

Level and Allele	$\hat{\tau}_i$	SD
Mild:		
Insertion	1.00	Fixed
Deletion	1.97	1.72
Severe:		
Insertion	1.00	Fixed
Deletion	7.32	4.00

Table 6

ACE as a Quantitative Trait

Parameter	Model	MLE	SD
$\omega_{\text{insertion}}$	I	.000	Fixed
	II	.000	Fixed
ω_{deletion}	I	1.31	.169
	II	1.29	.169
β	I	.000	Fixed
	II	−.206	.161

NOTE.—MLE = maximum-likelihood estimate.

Acknowledgments

The authors thank R. Cantor for her helpful suggestions; G. M. Lathrop, B. Keavney, and C. McKenzie for the ACE data; M. Stern and P. O'Connell for the diabetes data; and B. C. Hennis for the HRG data. The research was supported in part by USPHS grants CA16042 (to J.S.S.), DK47482 (to J.B.), DK42273 (to J.B.), MH59490 (to J.B.), and GM53275 (to K.L.).

References

Bradley RA, Terry ME (1952) Rank analysis of incomplete block designs. Biometrika 39:324–345

Collins A, Morton NE (1998) Mapping a disease locus by allelic association. Proc Natl Acad Sci USA 95:1741–1745

Goksel, DL, Fischbach K, Duggirala R, Mitchell BD, Aguilar-Bryan L, Blangero J, Stern MP, et al (1998) Variant in sulfonylurea receptor-1 gene is associated with high insulin concentrations in non-diabetic Mexican Americans: SUR-1 gene variant and hyperinsulinemia. Hum Genet 103:280–285

Hennis BC, Frants RR, Bakker E, Vossen RHAM, Van der Poort EW, Blonden LA , Cox S, et al (1994) Evidence for

the absence of intron H of the histidine rich glycoprotein gene: genetic mapping and in situ localization of HRG chromosome 3q28-29. Genomics 19:195–197

Hennis BC, Van Boheemen PA, Koeleman BPC, Boomsma DI, Engesser L, Van Weeb AGM, Novakova I, et al (1995) A specific allele of the histidine-rich glycoprotein (HRG) locus is linked with elevated plasma levels of HRG in a Dutch family with thrombosis. Br J Haematol 89:845–852

Jin K, Speed TP, Klitz W, Thomson G (1994) Testing for segregation distortion in the HLA complex. Biometrics 50: 1189–1198

Keavney B, McKenzie CA, Connell JM, Julier C, Ratcliffe PJ, Sobel E, Lathrop M, et al (1998) Measured haplotype analysis of the angiotensin-1 converting enzyme gene. Hum Mol Genet 11:1745–1751

Keener JP (1993) The Perron-Frobenius theorem and the ranking of football team. SIAM Rev 35:80–93

Lange K, Boehnke M, Weeks DE (1988) Programs for pedigree analysis: MENDEL, FISHER, and dGENE. Genet Epidemiol 5:471–472

Lange K, Hunter D, Yang I. Optimization transfer algorithms in statistics. J Computational Graphical Stat (in press)

Lazzeroni LC, Lange K (1998) A conditional inference framework for extending the transmission/disequilibrium test. Hum Hered 48:67–81

Lewis M, Kaita H, Philipps S, Giblet ER, Anderson JE, McAlpine PJ, Nickel B (1980) The position of the Radin blood group in relation to other chromosome 1 loci. Ann Hum Genet 44:179–184

Sham PC, Curtis D (1995) An extended transmission/disequilibrium test (TDT) for multi-allelic marker loci. Ann Hum Genet 59:323–336

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

Terwilliger JD, Ott J (1992) A haplotype-based "haplotype relative risk" approach to detecting allelic associations. Hum Hered 42:337–346