Am J Hum Genet 30:516–529, 1978

Variability of Human Linkage Data

D. C. RAO,[1] B. J. B. KEATS, N. E. MORTON, S. YEE, AND R. LEW

It is customary to summarize linkage data on a given pair of loci in man by lod scores which define the maximum likelihood estimate of a recombination rate $\hat{\theta}$ [1]. Apart from sampling errors, there are two causes of variability in such data. Biological variability subsumes sex, race, age, and other factors which determine recombination in a doubly heterozygous individual. Technical variability includes errors of parentage, phenotyping, recording, calculation of lods, and interpolation to estimate $\hat{\theta}$ as well as departure of the corresponding measure of goodness of fit from its assumed distribution.

Here we examine linkage data for 1,665 pairs of loci in 1,699 papers published up to the end of 1976. The purpose is to determine the frequency and cause of significant variability preparatory to construction of linkage maps. Preliminary steps have been described elsewhere [2, 3]. Errors in somatic cell assignments due to multiple isolation of the same clone, misclassification of chromosomes, and premature publication will not be considered.

LIKELIHOOD THEORY

If $\hat{z}_i$ is the maximum value of the lod in the ith of n sets of data and $\hat{z}$ is the overall maximum, then in large samples the quantity $2(\ln_e 10)\ (\Sigma\hat{z}_i - \hat{z})$ has a χ^2 distribution with $n - 1$ degrees of freedom if the n sets are homogeneous [4]. We shall be concerned about deviation from this assumed distribution in small samples, since the data sets range from a single pedigree to many nuclear families.

Certain types of data cannot be combined with lods and were therefore rejected. including the inefficient scores of Penrose, Bernstein, Wiener, and Haldane, and the Bayesian reduction of lods to an approximate posterior probability [5]. This is part of the argument against Bayesian linkage tests. Recombinant counts [6], Fisher-Finney scores, and lods may in principle be converted to a standard lod table, with an error which depends on sample size and the number of data points reported. Genetic markers which provide published evidence expressible by lods are listed in table 1.

Linkage workers are inconsistent in the values of θ for which they report lods. Sometimes only two values are given, usually for $\theta = .1$ and .3 [7]. Reconstruction of

Received January 2, 1978; revised March 31, 1978.

This work, PGL paper no. 180, was supported by grants GM 23021 from the U.S. National Institutes of Health and grant 1-475 from the National Foundation.

[1] All authors: Population Genetics Laboratory, University of Hawaii, 1980 East-West Road, Honolulu, Hawaii 96822.

the likelihood is then fraught with error, which reliable interpolation formulae can minimize. After some experience, we decided to form for each data set a standard lod table at the 11 values $\theta = .001, .01, .05, .10, .15, \ldots .45$, and from this table to estimate $\hat{\theta}$, $\hat{z}$ according to the rules given in Appendix A, the adequacy of which is tested in this paper.

Most publications provide no more than three data sets for each pair of loci, corresponding to males, females, and unclassified sex of the informative parent. In a few cases where data were summarized by investigator, race, or other variable of interest, this primary classifier was retained. We attempt to include each set only once even when published repeatedly. In some instances data were rejected because of considerable but unspecified overlap with other sets. While it is likely that a few duplications were not detected, we believe that such errors are negligible.

Detected linkages should be analyzed under different recombination values for the two sexes. Where the original data on linked loci are accessible, we calculate lods by a slight modification of the LIPED program [8] which simplifies data input, but analysis is still laborious and was not complete at the time this paper was written. Therefore, many pedigrees coded for linked loci and unspecified sex will ultimately be factored (to a close approximation) as

$$z(\theta_{m+f}) \doteq z(\theta_m; \theta_f^*) + z(\theta_f; \theta_m^*),$$

$$z(\theta_m; \theta_f^*) = \log \frac{f(\theta_m; \theta_f^*)}{f(1/2; \theta_f^*)}, \text{ and}$$

$$z(\theta_f; \theta_m^*) = \log \frac{f(\theta_f; \theta_m^*)}{f(1/2; \theta_m^*)},$$

where the subscripts m, f, and $m + f$ denote males, females, and sexes pooled respectively, θ_m^* and θ_f^* are the simultaneous maximum likelihood estimates, and $f(\cdot)$ denotes likelihood of the pedigree. This conserves the joint estimates θ_m^*, θ_f^* but neglects their covariance in the (generally small) information from double intercrosses and untested generations in which recombinants cannot with certainty be assigned to sex. We do not think that the results of the present paper are substantially affected by deferring this analysis until each linkage group is mapped in subsequent papers.

HETEROGENEITY

Two types of heterogeneity are analyzed here for each pair of loci: among studies within each sex (called heterogeneity within sexes) and heterogeneity between sexes. Let $\hat{z}_{ij}$, $\hat{z}_i$, and $\hat{z}$ denote, respectively, the maximum lod scores in the jth study for the ith sex, for the ith sex over all studies, and over all sexes and studies. Then heterogeneity within sexes is tested by $\chi^2 = 2(1n_e10)\,(\Sigma_j \hat{z}_{ij} - \hat{z}_i)$ for each observed sex i, and the heterogeneity between sexes is tested by $\chi^2 = 2(1n_e10)\,(\Sigma_i \hat{z}_i - \hat{z})$. Observed distributions of the tail probabilities of these χ^2's are presented in table 2. This table is subdivided depending on the overall recombination fraction $\hat{\theta}$ ($= .5$ or $< .5$) and the linkage status of the two loci involved. Tentative assignments of loci to chromosomes are shown in table 1. If two are assigned to the same chromosome, they are considered

TABLE 1

MARKERS USED IN THIS ANALYSIS

Marker	McKusick No.	Assigned to Chromosome	Marker	McKusick No.	Assigned to Chromosome	Marker	McKusick No.	Assigned to Chromosome
ABO	11030	9	GALA	30150	X	OLI	31350	X
ACP1	17150	2	GC	13920	E	OPA	31110	X
ADA	10270	20	GE	11075	···	ORO	13860	···
AED	30510	X	GLO1	13875	6	OVA	16690	···
AG	15200	(21)	GM	14710	(12)	P	11140	6
AIP	17600	···	GN	11160	···	PB	16875	···
AK1	10300	9	GPT1	13820	G	PCK	17390	(12)
ALB	10360	E	GPUT	23040	3	PEPA	16980	18
ALD	30010	X	G6PD	30590	X	PEPB	16990	12
ALL	20920	···	HA5	14274	···	PEPC	17000	1
ALZ	10430	···	HA9	···	···	PEPD	17010	(19)
AL1	20310	K	HBA	14180	(2)	PG	16970	6
AMY(1+2)	10465|70	1	HBB	14190	(4)	PGD	17220	1
AN	11035	···	HBD	14210	B	PGK	31180	X
ANH1	30130	X	HBM	30980	X	PGM1	17190	1
ANR2	10620	(1)	HC	14440	H	PGM2	17200	(4)
AOD	10900	(1)	HCH	14310	···	PGM3	17210	6
ATX1	16440	6	HEMA	30670	X	PH	16940	N
ATX2	22930	···	HEMB	30690	X	PHI	17240	19
ATX3	21320	K	HG	30823	X	PI	10740	(12)
ATX4	16460	···	HGTH	21285	P	PI1	17510	(1)
AU	11040	···	HLA(1+2)	14280|83	6	PKU	26160	···
BCNS	10940	···	HPA	14010	16	PMD	30920	X
BF	13847	6	HPRT	30800	X	PR	16878	···
BP	11150	···	HYD1	30700	X	PTC	17120	F
BY	11150	···	IB	11080	M	PTCD	30310	X
CAE1	11620	(1)	ICH1	30810	X	RD	11150	···
CAE2	11680	···	IDH1	14770	2	RES	31230	X
CAT1	30220	X	IR	14685	(6)	RH	11170	1
CAT2	21250	M	ISF	···	···	RMI	15795	(6)
CBD	30380	X	JK	11100	(7)	RP1	31260	X
CBP	30390	X	JR	11160	···	RP2	18010	···
CF	21970	···	K	11090	F	RS	31270	X
CHI	11043	6	KM	14720	···	SC	11175	1

TABLE 1 (continued)

Marker	McKusick No.	Assigned to Chromosome	Marker	McKusick No.	Assigned to Chromosome	Marker	McKusick No.	Assigned to Chromosome
CO	11045	(7)	KO	17350	···	SDT	31340	X
CP	11770	···	KOA	14930	···	SE	18210	A
CT	11550	···	KS	24440	···	SF	11180	···
CU	12010	···	LCAT	24590	16	SOD1	14745	21
C2	12060	6	LD	15210	···	SP	31285	X
C3	12070	H	LDC	12220	···	SPH1	18290	(12)
C5	12090	···	LDHA	15000	11	SPP	31290	X
C8	12095	6	LE	11110	H	STY	18160	(2)
DEAF	30450	X	LP	15220	···	SW	11150	···
DEF1	30440	X	LU	11120	A	TBG	31420	X
DHDR	12660	···	LW	···	···	TDO	13080	···
DI	11050	···	MDU	15900	N	TF	19000	D
DIA1	25080	···	MD1	31020	X	TFM	31370	X
DM	16090	A	MD2	31010	X	THA	27350	···
DO	11060	(1)	MD4	25360	···	THB	27350	···
DYS	12760	···	MLC1	15785	(6)	THH	18730	···
EBS1	13195	G	MNS	11130	(2)	TR	11150	···
EBS2	13170	···	MPS2	30990	X	UL	11200	···
EBS3	22660	···	MS	15470	···	UMPK	19173	1
ECN	10690	···	MSS	24880	P	VEL	11160	···
EDS2	30520	X	NA	16285	···	W1	19350	(9)
EL1	13050	1	NB	16286	···	WB	11150	···
EL2	13060	···	ND	31060	X	WR	11150	···
EM	13370	···	NDI	30480	X	XG	31470	X
ENO1	17245	1	NPP	16405	14	XM	31490	X
ESD	13328	13	NPS	16120	9	XPE	27870	(9)
E1	17740	D	NS	12553	···	YE	19445	···
E2	17750	(1)	NYS	31070	X	YT	11210	···
FY	11070	1	OA	30050	X			

NOTE.—A maximum of four characters are used for gene markers. A fifth character will be introduced later if necessary. Smaller symbols are ambiguous: for example, PG is used in reference 0007 [21] and reference 1129 [22] to denote PGM_1, not pepsinogen, which is conventionally PG. For comparability with computer output, the character set is restricted to numbers and capital letters; subscripts and superscripts are not permitted. Loci with no lod score evidence on linkage are omitted. Provisional or tentative assignment of a marker to a chromosome [23] is in parentheses. Linkage groups which have not yet been assigned to a chromosome are denoted by letters A–P.

TABLE 2

HETEROGENEITY WITHIN AND BETWEEN SEXES

SOURCE	NOMINAL PROBABILITY LEVEL P_e FOR χ^2								TOTAL
	1–.95	.95–.9	.9–.8	.8–.2	.2–.1	.1–.05	.05–.01	.01–.001	<.001	
A.) Within Sexes										
$\hat{\theta}$ = .5:										
Linked loci	15	1	2	6	...	...	...	...	...	24
Unlinked loci	216	14	21	80	5	1	1	...	...	338
Probably unlinked	42	3	3	15	1	...	...	...	...	64
$\hat{\theta}$ < .5:										
Linked loci	12	5	13	49	10	5	3	1	...	98
Unlinked loci	32	17	27	158	20	7	10	1	...	272
Probably unlinked	15	2	4	46	5	5	...	...	...	77
B.) Between Sexes										
$\hat{\theta}$ = .5:										
Linked loci	7	1	2	7	...	...	...	...	...	17
Unlinked loci	157	18	22	55	6	...	2	...	...	260
Probably unlinked	111	5	11	40	1	...	...	...	...	168
$\hat{\theta}$ < .5:										
Linked loci	5	1	3	17	6	5	4	2	5	48
Unlinked loci	22	9	18	129	17	7	4	2	...	208
Probably unlinked	36	8	13	60	10	5	1	...	...	133

to be linked; if they are on different chromosomes, they are unlinked; otherwise, they are taken as probably unlinked.

As anticipated by Haldane [9], the mean and variance of χ^2 within sexes is less than in large samples. This discrepancy is most striking when $\hat{\theta} = .5$, in which case the data sets commonly have the same maximum and $\chi^2 = 0$. Reduction in mean and variance is also apparent for heterogeneity between sexes (table 2).

There is greater heterogeneity for linked loci with $\hat{\theta} < .5$ (table 2): the nominal significance level $P_e = .05$ is transgressed four times within sexes ($P_o = 4/98 = .041$) and 11 times between sexes ($P_o = 11/48 = .229$). Evidently sex is the major cause of variable recombination for linked loci.

Pairs of loci with significant heterogeneity ($P_e < .05$) within or between sexes were reexamined. In most cases heterogeneity within sexes was not accompanied by significant heterogeneity between studies (pooled over sexes), and heterogeneity between sexes was consistent with linkage at different recombination values in males and females.

Preliminary examination revealed reference 0586* [10] as the source of apparent heterogeneity for several pairs (table 3). A quotation from that paper may be relevant: "The analysis was carried out by a computer program known as the MARK II on a Univac 418. This program is known to be capable of giving erroneous results, due to cumulative rounding errors, where close linkage is present between loci, at least one of which is rare." Presumably the clause "where close linkage is present between loci" should read "for small values of θ," since rounding errors may depend on the assumed value of linkage, but cannot depend on its unknown true value. Apart from numerical errors, the calculation from only two values of θ makes interpolation error-prone.

Problems with this reference are not exclusively numerical. ABO (on chromosome 9) gives a lod of 3.0 with the centromere of chromosome 1 at $\theta = .1$ ($\hat{z} = 3.27$ at $\hat{\theta} = .02$). This is the only instance known to us where such strong evidence of linkage is spurious. The authors note that "The high lod score . . . is derived largely from sibling data and does not appear in two- and three-generation families," which is contrary to the general experience that sibling data are relatively uninformative. Other studies give no suggestion of linkage (heterogeneity $\chi^2_1 = 9.27$; table 3). Not only does this paper give a spurious linkage, but it claims unusually frequent recombination between the centromere of chromosome 1 and the linked FY locus ($\chi^2_1 = 8.51$), whereas recombination with GC is surprisingly rare ($\chi^2_1 = 5.03$).

On the above evidence we have rejected the data in reference 0586 [10]. Keats et al. [11] give reasons to omit two other bodies of data: reference 0045 [12] because of inextricable duplication with other sources, and the JK tests from references using the Brazilian data [13] because of serological unreliability. The above exclusions were made before the analysis presented here (tables 1–5).

From other instances of apparent heterogeneity, we selected the most striking for further analysis (table 3). Some appear to be genuine type I errors, illustrating that $z > 2$ is not enough evidence for linkage, but at least two may be systematic. In reference

* Each linkage paper in our data is assigned a unique four-digit number.

TABLE 3

PROBABILITY RATIO TESTS FOR HETEROGENEITY

First locus	Second Locus	Source	$\hat{z}$	$\hat{\theta}$	χ^2 among sources	df	P_e	P_o
A.) Reference 0586 [10] vs. Others								
01Q12	ABO	0586 [10]	3.266	.022	9.27	1	.002	.001
		Others	.097	.366				
		Total	1.349	.283				
01Q12	FY	0586 [10]	.017	.463	8.51	1	.004	.001
		Others	11.049	.136				
		Total	9.218	.176				
01Q12	GC	0586 [10]	1.197	.022	5.03	1	.025	.011
		Others	0	.5				
	Total		.104	.394				
B.) Selected Pairs of Loci								
ACP1	LE	0955 [24]	2.227	.145	8.63	1	.004	.001
		Others	0	.500				
		Total	.354	.365				
ADA	GLO1	0494 [25]	1.660	.013	6.42	1	.011	.004
		Others	0	.500				
		Total	.265	.237				
FY	K	0624 [26]	2.513	.192	(based on U & K scores)			
		Others	0	.500	8.29	1	.004	.001
		Total	.713	.340				
FY	LP	1129 [22]	2.006	.029	6.23	1	.013	.005
		Others	0	.5				
		Total	.654	.216				
GC	MNS	0007 [21]	.518	.422	5.00	1	.025	.011
		Others	2.518	.331				
		Total	1.950	.394				
GM	PI	Pi^Z [15]	9.935	.162	5.78	1	.016	.007
		Others	7.152	.267				
		Total	15.832	.240				
PGD	RH	Blacks [14]	2.794	.191	1.02	1	.313	.245
		Whites	4.549	.263				
		Total	7.122	.247				
PGM1	RH	Blacks [14]	2.285	.230	3.30	1	.069	.039
		Whites	.405	.362				
		Total	1.974	.330				
AMY	FY	Blacks [14]	0	.5	5.69	1	.017	.007
		Whites	2.140	.102				
		Total	.904	.240				

0998, Weitkamp [14] suggested that linkage between loci on chromosome 1 may be variable among populations. In two of three instances this is supported by our test of heterogeneity between races ($P_o < .05$). Gedde-Dahl et al. ([15], reference 0337) proposed that the Pi^Z allele may undergo recombination with GM less frequently than other alleles, as if it were associated with a duplication, deficiency, or inversion. This too is supported by our analysis ($\chi^2_1 = 5.78$).

TABLE 4

CUMULATIVE DISTRIBUTION OF NOMINAL AND EMPIRICAL SIGNIFICANCE LEVELS WITHIN AND AMONG SEX, EXCLUDING LINKED LOCI ($\hat{\theta} < .5$)

Nominal P_e	Empirical P_o	Fitted P_o
.95	.848	.940
.75	.668	.706
.50	.442	.432
.25	.177	.186
.15	.094	.100
.10	.061	.061
.05	.026	.026
.01	.004	.004
.001	0	.0002

NOTE.—Observed (o_i) and expected (e_i) frequencies of P values over the three intervals of P = (0, .05), (.05, .1) and (.1, 1), obtained from table 1, yield the significance test: $\chi^2 = \Sigma(o_i - e_i)^2/e_i = 12.26$ on 2 df.

Whatever the proportion of true instances of biological heterogeneity may be, the empirical significance level is conservative. Letting P_e denote the nominal (tabular) significance level and P_o be the empirical significance level, we obtain a relationship between P_e and P_o, based on table 2, as follows. Consider the observed distributions of P_e values of table 2 for "unlinked" and "probably unlinked" pairs with $\hat{\theta} < .5$. Summing these four distributions gives the overall distribution of 690 values. These 690 P values would be expected to be uniformly distributed in large samples under the null hypothesis of homogeneity. However, the observed distribution shows significantly reduced tail areas (table 4). Regression of $\ln P_o$ on $\ln P_e$ through the origin for the nine values of table 4 gave the regression coefficient as 1.212. Since some cases of apparent heterogeneity may be true, $P_0 \leqslant P_e^{1.212}$ gives the empirical significance corresponding to any tabular value. For example, Weitkamp's racial difference for PGM1 and RH yielded $\chi^2_1 = 3.30$ ($P_e = .069$) for which $P_o \leqslant (.069)^{1.212} = .039$ which is significant. The empirical relation between P_o and P_e will be useful in assessing the significance of other instances of apparent heterogeneity in human linkage data.

TABLE 5

POWER AND RELIABILITY OF THE LINKAGE TEST AMONG 1,665 PAIRS OF LOCI

	$\log_{10} A$							
CHARACTERISTIC	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Number with $\hat{z} > \log_{10} A$	228	118	74	53	46	41	32	31
$P(\hat{Z} > \log_{10} A)$	.137	.071	.044	.032	.028	.025	.019	.019
Power, $1 - \beta$, at $\phi = .054$	.388	.477	.530	.502	.485	.448	.354	.344
Reliability, ρ, at $\phi = .054$	.153	.363	.644	.851	.948	.983	.994	.998
Conservative significance, $1/A$	.316	.100	.032	.010	.003	.001	.0003	.0001
Power at $\phi = .095$	.359	.385	.354	.306	.283	.257	.202	.196
Reliability at $\phi = .095$	.249	.512	.768	.913	.971	.991	.997	.999

POWER AND RELIABILITY

A simple test for linkage sums the lods for a pair of loci over sexes and sources to consider the maximum value $\hat{z}$, asserting linkage if $\hat{z} > \log_{10} A$, where $A \doteq (1 - \beta)/\alpha$ [16]. Here α is the probability of falsely asserting linkage (a type I error), and β is the probability of failing to detect true linkage (a type II error). If we take pairs of loci at random, the frequency of significant tests is

$$P(\hat{z} > \log_{10} A) = \phi(1 - \beta) + (1 - \phi)\,\alpha = (1 - \beta)\,(A\phi + 1 - \phi)/A,$$

where ϕ is the prior probability of linkage. Among significant tests the posterior probability of linkage is

$$\rho = \frac{\phi(1 - \beta)}{P(\hat{z} > \log_{10} A)} = \frac{\phi A}{A\phi + 1 - \phi}.$$

Morton [1] called ρ the *reliability of* the test. From other organisms he guessed $\phi = .05$, assuming a uniform distribution of loci on the linkage map. Renwick [17] calculated $\phi = .054$ for man, and this was confirmed by Elston and Lange [18] using the mitotic lengths from the Paris Conference [19]. Table 5 gives these empirical measures of power and reliability, which do not depend explicitly on the criterion used to reject linkage. It cannot be doubted that linkage is ultimately rejected after attempts to detect linkage for a particular pair of loci have failed. Therefore the linkage test is always sequential, although the decision rule is not so simple as the theory supposes. Reliability increases with A, and a value of 1,000 (i.e., $\log_{10} A = 3$) is necessary and sufficient for a strong claim of linkage as Morton [1] suggested. These empirical values from tests in progress resemble his deductions for completed sequential tests, although the data depart somewhat from his assumptions. For example, for autosomal as well as X-linked loci, 158 pairs of loci are considered linked, instead of the 90 expected under random sampling of loci: the excess may well be due to preferential testing of loci inferred from other evidence to be linked. Estimates of reliability should be increased (and of power decreased) accordingly (table 5 with $\phi = 158/1{,}665 = .095$). Even when only autosomal loci are considered, 107 out of 1,595 pairs are considered linked, instead of the 86 expected (giving $\phi = 107/1{,}595 = .067$).

While the empirical power and reliability must be interpreted cautiously, it appears that most linkages suggested by $\hat{z} > 1$ are false, whereas most linkages supported by $\hat{z} > 2$ are true. Keats et al. [20] give lods for eight pairs of loci with $\hat{z} > 1.5$ but not known to be linked. Of these, at least several are expected to be true.

SUMMARY

Linkage scores (lods) are reduced to a standard table for 1,665 pairs of loci. The likelihood ratio test of homogeneity is shown to be conservative in these small samples; for $\hat{\theta} < .5$ the nominal significance level of .05 corresponds to a true significance level of .026. Sex is the major cause of variability for linked loci. Biological and technical sources of residual heterogeneity are discussed and illustrated by published examples. Empirical power and reliability are in good agreement with earlier predictions.

APPENDIX A
CONVERSION OF DATA TO STANDARD LOD TABLE

For pedigrees analyzed by us, our modified LIPED program [8] gives standard lod tables directly for each sex. Where there is evidence of linkage and data are adequately reported, we plan to do such analysis. Otherwise we use an algorithm described in this appendix to construct standard lod tables if the data are acceptably reported and reject them otherwise. In the future, data should be reported as a standard lod table for each sex separately and for recombination values in the other sex close to the maximum likelihood estimate when there is evidence for linkage.

We follow an elaborate algorithm to construct the standard table of lod scores. The main features are outlined below. A few numerical examples are presented in table A1.

Case 1: Recombination Counts

The counting method of Smith [6] gives estimates of the number of recombinants A and the number of informative progeny N, such that $\hat{\theta} = A/N$ even though A is not in general binomial nor N the sample size. These may be converted into equivalent values of U and K (Fisher-Finney U scores) which can then be converted into lods (case 2):

Letting $T = 1 - 4\theta(1 - \theta)$,

$$K = \left(-\frac{\delta^2 \ln L}{\delta\theta^2}\right) \Big/ \left(\frac{\delta T}{\delta\theta}\right)^2 \Bigg|_{\theta = \hat{\theta}},$$

where

$$-\frac{\delta^2 \ln L}{\delta\theta^2}\Bigg|_{\theta = \hat{\theta}} = N/\hat{\theta}(1 - \hat{\theta}) = \frac{N^3}{A(N - A)},$$

and

$$\frac{\delta T}{\delta\theta} = -4(1 - 2\theta) = -\frac{4(N - 2A)}{N}.$$

Therefore, $K = N^5/16A(N - A)(N - 2A)^2$.

Since $\hat{\theta} \doteq (1 - \sqrt{U/K})/2$, a recombinant count A in a sample of size N corresponds to $U = N^3/16A(N - A)$. If A is actually binomial, it may be converted directly to lods.

Case 2: Fisher-Finney U Scores

Given U,K directly or by conversion of recombination counts, we estimate

$$\hat{\theta} = \begin{cases} .5 & \text{if } U \leq 0 \\ (1 - \sqrt{U/K})/2 & \text{if } 0 < U < K \\ 0 & \text{if } U \geq K \end{cases}$$

$$z \doteq (\log_{10} e)(TU - T^2K/2),$$

where $T = 1 - 4\theta(1 - \theta)$.

TABLE A1

$\hat{z}$, $\hat{\theta}$ AND STANDARD TABLE OF LOD SCORES FOR A FEW EXAMPLES

OUR CODE*	LOCUS-1	LOCUS-2	$\hat{z}$	$\hat{\theta}$	LOD SCORES AT θ = .001	.01	.05	.1	.15	.2	.25	.3	.35	.4	.45
$U = 13$ and $K = 21$ yield:															
1441201 [27]	LU	SE	1.75	.11	1.10	1.22	1.58	1.75	1.67	1.44	1.13	.79	.47	.22	.06
$U = -.41$ and $K = 7.4$ yield:															
1440206 [28]	FY	PTC	0	.5	− 1.77	− 1.65	− 1.20	−.77	− .47	− .27	− .14	− .07	−.03	−.01	−.00
0045468 [12]	GM	SE	0	.5	−75.16	−43.94	−22.60	−13.90†	−8.29	−4.59	−2.31	−1.02†	−.37	−.10	−.02
0045467 [12]	GM	SE	.15	.36	−39.06	−22.84	−11.75	− 7.23†	−3.81	−1.70	− .54	− .01†	.14	.11	.04
0007573 [21]	E1	TF	.53	.10	.04	.34	.51	.53†	.51	.48	.42	.36†	.29	.20	.11
0007200 [21]	GM	SE	.62	.34	−37.75	−22.07	−11.35	−6.98†	−2.94	- .57†	.12	.51†	.59	.42†	.23
0568205 [29]	HLA	MNS	2.77	.22	− 8.37	− 4.89	− 2.52†	.90†	2.14	2.70†	2.54	2.08†	1.34	.76†	.32

* First four digits correspond to the reference number in our list of publications on linkage: the last three digits correspond to the serial number of the data set within; that reference number (serial numbers start at 200).

† These lod scores are reported in the literature.

Case 3: Two Values of z

A mooted convention gives only two lods, z_1 and z_2, for $\theta_1 = .1$ and $\theta_2 = .3$ [7]. If θ_i with larger z_i is small, we use binomial theory corresponding to a count of A recombinants and B nonrecombinants. Then $z = A \log(2\theta) + B \log[2(1-\theta)]$, where

$$B = \frac{z_1 \log(2\theta_2) - z_2 \log(2\theta_1)}{\log(2\theta_2)\log[2(1-\theta_1)] - \log(2\theta_1)\log[2(1-\theta_2)]},$$

$$A = \frac{z_1 - B\log[2(1-\theta_1)]}{\log(2\theta_1)}$$

$$\hat{\theta} = \begin{cases} .001 & \text{if } A < 0, B > 0 \\ .5 & \text{if } A > 0, B < 0 \\ A/(A+B) & \text{otherwise.} \end{cases}$$

If θ_i with larger z_i is large, we prefer the Fisher-Finney equation, $z = AT + BT^2$, where

$$B = \frac{z_1 T_2 - z_2 T_1}{T_1 T_2 (T_1 - T_2)}$$

$$A = z_1/T_1 - BT_1$$

$$\theta = \begin{cases} .5 & \text{if } -A/2B \leq 0 \\ (1 - \sqrt{-A/2B})/2 & \text{if } 0 < -A/2B < 1 \\ 0 & \text{if } -A/2B \geq 1. \end{cases}$$

Case 4: Three or More Lod Scores

Judging from the given lod scores, if the maximum is expected to be around $\theta = .5$, we use Fisher-Finney method; if the maximum is around $\theta = 0$, we use the binomial method outlined in case 3. If the maximum occurs for an intermediate value of θ, we fit a quadratic equation of the type $z = a + b\theta + c\theta^2$ to three points (θ_1, z_1), (θ_2, z_2), and (θ_3, z_3), where z_2 is the largest given lod score, z_1 is the largest given lod score for $\theta < \theta_2$, and z_3 is the maximum given lod for $\theta > \theta_2$ [4]:

$$c = \frac{(z_1 - z_3)(\theta_2 - \theta_3) - (z_2 - z_3)(\theta_1 - \theta_3)}{(\theta_1^2 - \theta_3^2)(\theta_2 - \theta_3) - (\theta_2^2 - \theta_3^2)(\theta_1 - \theta_3)}$$

$$b = \frac{z_1 - z_3}{\theta_1 - \theta_3} - c(\theta_1 + \theta_3), \text{ and}$$

$$a = z_1 - b\theta_1 - c\theta_1^2,$$

to estimate

$$\hat{\theta} = \begin{cases} .5 & \text{if } -b/2c \geq .5 \\ -b/2c & \text{if } 0 < -b/2c < .5 \\ 0 & \text{if } -b/2c \leq 0. \end{cases}$$

REFERENCES

1. MORTON NE: Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277–318, 1955
2. MORTON NE, RAO DC, LINDSTEN J, HULTÉN M, YEE S: A chiasma map of man. *Hum Hered* 27:38–51, 1977
3. RAO DC, MORTON NE, LINDSTEN J, HULTÉN M, YEE S: A mapping function for man. *Hum Hered* 27:99–104, 1977
4. MORTON NE: The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am J Hum Genet* 8:80–96, 1956
5. SMITH CAB: Some comments on the statistical methods used in linkage investigations. *Am J Hum Genet* 11:289–304, 1959
6. SMITH CAB: Counting methods in genetical statistics. *Ann Hum Genet* 21:254–276, 1957
7. MCKUSICK VA, EDWARDS JH: Report of the committee on unassigned syntenic groups and theoretical considerations, in *Human Gene Mapping 2,* edited by BERGSMA D, New York, National Foundation, 1975, pp 26–28
8. OTT J: Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* 26:588–597, 1974
9. HALDANE JBS: The mean and variance of χ^2 when used as a test of homogeneity, when expectations are small. *Biometrika* 31:346–355, 1940
10. FERGUSON-SMITH MA, ELLIS PM, MUTCHINICK O, GLEN KP, CÔTÉ GB, EDWARDS JH: Centromeric linkage, in *Human Gene Mapping 2*, edited by BERGSMA D, New York, National Foundation, 1975, pp 130–137
11. KEATS BJB, MORTON NE, RAO DC: Likely linkage: Inv with Jk. *Hum Genet* 39:157–159, 1977
12. EDWARDS JH: Total lods for linkage analysed by the New York Blood Center and Birmingham Computer programs, in *Human Gene Mapping,* edited by BERGSMA D, New York, National Foundation, 1974, pp 187–202
13. MORTON NE: Genetic studies of northeastern Brazil. *Cold Spring Harbor Symp Quant Biol* 29:69–79, 1964
14. WEITKAMP LR: Population differences in meiotic recombination frequency between loci on chromosome 1, in *Human Gene Mapping,* edited by BERGSMA D, New York, National Foundation, 1974, pp 179–182
15. GEDDE-DAHL T JR, COOK PJL, FAGERHOL MK, PIERCE JA: Improved estimate of the Gm—Pi linkage. *Ann Hum Genet* 39:43–50, 1975
16. WALD A: *Sequential Analysis.* New York, Wiley, 1947
17. RENWICK JH: Progress in mapping human autosomes. *Br Med Bull* 25:65–73, 1969
18. ELSTON RC, LANGE K: The prior probability of autosomal linkage. *Ann Hum Genet* 38:341–350, 1975
19. PARIS CONFERENCE (1971): *Standardization in Human Cytogenetics.* New York, National Foundation, 1972
20. KEATS BJB, MORTON NE, RAO DC: Possible linkages (lod score over 1.5), and a tentative map of the Jk-Km linkage group, in *Human Gene Mapping 4,* New York, National Foundation. In press, 1978
21. CHAUTARD-FREIRE-MAIA EA: Linkage relationships between 22 autosomal markers. *Ann Hum Genet* 38:191–198, 1974
22. FALK CT, WALKER ME, MARTIN MD, ALLEN FH: Autosomal linkage in humans (methodology and results of computer analysis). *Ser Haematol* 8:153–237, 1975
23. MCKUSICK VA, RUDDLE FH: The status of the gene map of the human chromosomes. *Science* 196:390–405, 1977
24. VAN CONG N, MOULLEC J: Linkage probable entre les groupes de phosphatase acide des globules rouges et le systeme Lewis. *Ann Génét* 14:121–125, 1971
25. KOMPF J, BISSBORT S, RITTER H: Red cell glyoxalase I (E.C.4.4.1.5): formal genetics and linkage relations. *Humangenetik* 28:249–251, 1975

26. MOHR J: Genetics of fourteen marker systems: associations and linkage relations. *Acta Genet Stat Med (Basel)* 16:1–58, 1966
27. METAXAS MN, METAXAS-BÜHLER M, DUNSFORD I, HOLLÄNDER L: A further example of anti-Lu^b together with data in support of the Lutheran-Secretor linkage in man. *Vox Sang* 4:298–307, 1959
28. MERTON BB: Taste sensitivity to PTC in 60 Norwegian families with 176 children. Confirmation of the hypothesis of single gene inheritance. *Acta Genet Stat Med (Basel)* 8:114–128, 1958
29. MAYR WR, MAYR D: Analyse der Koppelung zwischen dem HL-A-System und anderen loci. *Humangenetik* 24:129–133, 1974

New Editor

As of July 1, 1978, all manuscripts and correspondence concerning editorial matters should be directed to the new editor, Dr. David E. Comings, Department of Medical Genetics, City of Hope Medical Center, 1500 E. Duarte Road, Duarte, California 91010.