Further Scoring Types in Sequential Linkage Tests, With a Critical Review of Autosomal and Partial Sex Linkage in Man

NEWTON E. MORTON[1]

University of Wisconsin Medical School

INTRODUCTION

THIS IS the fourth and last in a series of papers on the application of probability ratio (lod) scores to human linkage. The first (Morton, 1955a) considered the properties of sequential linkage tests and the scoring types which arise when the parental genotypes are known and there are only two alleles at each locus. The second (Morton, 1956) treated pedigree data and used lods to obtain likelihood ratio tests of homogeneity and maximum likelihood estimates of linkage. The third (Steinberg and Morton, 1956) included applications to multiple allelic test loci. The present communication extends the scores to multiple alleles at both loci, pseudoalleles, and partial sex linkage, and concludes with a review of possible linkages in man.

INTERCROSSES AND MULTIPLE ALLELES. PARENTAL GENOTYPES KNOWN, BOTH PARENTS TESTED. COMPLETE PENETRANCE, NO NATURAL SELECTION

In the absence of epistasis, all single and double backcrosses can be reduced to diallelic scoring types. For example, if T_1 and T_2 are codominant and t is recessive to T_1, so that the mating $T_1t \times T_1T_2$ produces progeny of phenotypes T_1, T_1T_2, and T_2t, then the mating Gg $T_1t \times$ gg T_1T_2 is of type 1 (Morton, 1955a) with the T_1 progeny not informative, and is scored as $z_1 + c_1$ or $z_1 + e_1$ with a(GT_1T_2), b(GT_2t), c(gT_1T_2), d(gT_2t), while the mating Gg $T_1T_2 \times$ Gg tt is of type 11 and is scored as z_2 with a(GT_1), b(gT_1), c(GT_2t), and d(gT_2t).

Multiple allelism introduces several new double intercrosses. The most important intercross types may be denoted by their segregation ratios as follows (Smith, 1954):

Type	Mating	Examples: ABO	Examples: MNS
31	$Tt \times Tt$	AO × AO	MS/Ms × MS/Ms
121	$T_1T_2 \times T_1T_2$	AB × AB	Ms/Ns × Ms/Ns
211	$T_1t \times T_1T_2$	AO × AB	MS/Ms × MS/Ns
1111	$T_1T_2 \times T_3T_4$ ($T_1T_2 \times T_2t$, etc.)	$A_2B \times A_1O$	MS/Ns × Ms/NS

These four intercrosses give eleven essentially different double intercrosses (table 1), of which the first five have been treated previously (Morton, 1955a; Steinberg and

Received July 9, 1956.

[1] This study was conducted during the tenure of a National Cancer Institute Postdoctorate Research Fellowship. Department of Genetics Paper No. 644.

TABLE 1. DOUBLE INTERCROSSES

Mating class	Parental genotypes				Progeny					
(31)(31)	Gg	Tt	× Gg	Tt		T	t	Sum		
					G	a	b	s_1		
					g	c	d	s_2		
(31)(121)	Gg	T_1T_2	× Gg	T_1T_2		T_1	T_2	T_1T_2		
					G	a	c	e		
					g	b	d	f		
(121)(121)	G_1G_2	T_1T_2	× G_1G_2	T_1T_2		T_1	T_2	T_1T_2		
					G_1	a	b	g	u = a+d	
					G_2	c	d	h	v = b+c	
					G_1G_2	e	f	i	w = e+f+g+h	
(31)(211)	Gg	T_1t	× Gg	T_1T_2		T_1	T_2t	T_1T_2	Sum	
					G	a	c	e	s_1	
					g	b	d	f	s_2	
(31)(1111)	Gg	T_1T_2	× Gg	T_3T_4		T_1T_3	T_2T_3	T_1T_4	T_2T_4	Sum
					G	a	c	e	g	s_1
					g	b	d	f	h	s_2
(121)(211)	G_1G_2	T_1t	× G_1G_2	T_1T_2		T_1	T_2t	T_1T_2	Sum	
					G_1	a	d	g	s_1	
					G_1G_2	—	e	h	s_2	
					G_2	c	f	i	s_3	
(121) (1111)	G_1G_2	T_1T_2	× G_1G_2	T_3T_4		T_1T_3	T_2T_3	T_1T_4	T_2T_4	Sum
					G_1	a	d	g	j	s_1
					G_1G_2	b	e	h	k	s_2
					G_2	c	f	i	l	s_3
(211)(211) cis	G_1g	T_1t	× G_1G_2	T_1T_2		T_1	T_2t	T_1T_2		
					G_1	a	d	g		
					G_2g	b	e	h		
					G_1G_2	c	f	i		
(211)(211) trans	G_1g	T_1T_2	× G_1G_2	T_1t		T_1	T_2t	T_1T_2		
					G_1	—	d	g		
					G_2g	b	e	h		
					G_1G_2	c	f	i		
(211) (1111)	G_1g	T_1T_2	× G_1G_2	T_3T_4		T_1T_3	T_2T_3	T_1T_4	T_2T_4	
					G_1	a	d	g	j	
					G_2g	b	e	h	k	
					G_1G_2	c	f	i	l	
(1111) (1111)	G_1G_2	T_1T_2	× G_3G_4	T_3T_4		T_1T_3	T_2T_3	T_1T_4	T_2T_4	
					G_1G_3	a	e	i	m	
					G_2G_3	b	f	j	n	
					G_1G_4	c	g	k	o	
					G_2G_4	d	h	l	p	

Morton, 1956). Several of these types are likely to be selected through the progeny for both loci, and this necessitates a score correction. The main locus G may be subject to truncate or arbitrary selection and the test locus T to truncate selection. In the 211 mating $T_1t \times T_1T_2$, selection will usually require at least one T_2t child ($T_2t > 0$). For example, in the absence of s-antisera or pedigree information, the mating MS/Ms × MS/Ns can be recognized only if there is at least one MNs child, and $A_1A_2 \times A_1B$ can be recognized only if there is at least one A_2B child. In the 1111

intercross type ($T_1T_2 \times T_3T_4$), the condition for selection may be $T_1T_4 + T_2T_4 > 0$ (for example, at least one A_2 or A_2B child from the mating $A_2B \times A_1A_2$, or at least one Ms or MNs child from Ms/Ns × Ms/NS). Another common condition for selection is $T_1T_4 + T_2T_4 > 0$, $T_2T_3 + T_2T_4 > 0$ (as at least one A_2 or A_2B child and

TABLE 2. DOUBLE INTERCROSS SCORES FOR THE PARENTS SEPARATELY

Mating class	Genotype		Selection		Score
			G	T	
(31)(211)	Gg	T_1t	$g > 0$	$T_2t > 0$	$z_2(c, d, e, f) + c_2(c + d + e + f)$
(31)(211)	Gg	T_1t	arbitrary	$T_2t > 0$	$z_2(c, d, e, f) + e_2(c + e, d + f)$
(31)(211)	Gg	T_1T_2	arbitrary	complete	$z_2(a, b, c + e, d + f)$
(31)(1111)	Gg	T_1T_2	arbitrary	complete	$z_2(a + e, b + f, c + g, d + h)$
(31)(1111)	Gg	T_1T_2	arbitrary	$T_2 > 0$	$z_2(a + e, b + f, c + g, d + h) + e_2$
(31)(1111)	Gg	T_1T_2	$g > 0$	$T_2 > 0$	$z_2(a + e, b + f, c + g, d + h) + c_2$
(31)(1111)	Gg	T_3T_4	arbitrary	complete	$z_2(a + c, b + d, e + g, f + h)$
(31)(1111)	Gg	T_3T_4	arbitrary	$T_4 > 0$	$z_2(a + c, b + d, e + g, f + h) + e_2$
(121)(211)	G_1G_2	T_1t	complete	arbitrary	$z_1(d + i, f + g)$
(121)(211)	G_1G_2	T_1t	arbitrary	$T_2t > 0$	$z_1(d + i, f + g) + e_1(d + g, f + i)$
(121)(211)	G_1G_2	T_1T_2	arbitrary	complete	$z_1(a + f + i, c + d + g)$
(121)(1111)	G_1G_2	T_1T_2	complete	arbitrary	$z_1(a + f + g + l, c + d + i + j)$
(121)(1111)	G_1G_2	T_1T_2	arbitrary	complete	$z_1(a + f + g + l, c + d + i + j)$
(121)(1111)	G_1G_2	T_1T_2	arbitrary	$T_2 > 0$	$z_1(a + f + g + l, c + d + i + j) + e_1(s_1, s_3)$
(121)(1111)	G_1G_2	T_3T_4	arbitrary	complete	$z_1(a + d + i + l, c + f + g + j)$
(121)(1111)	G_1G_2	T_3T_4	arbitrary	$T_4 > 0$	$z_1(a + d + i + l, c + f + g + j) + e_1(s_1, s_3)$
(211)(211) cis	G_1g	T_1t	$G_2g > 0$	$T_2t > 0$	$z_1(e + i, f + h) + c_1(e + f + h + i)$
(211)(211) cis	G_1G_2	T_1T_2	arbitrary	complete	$z_1(a + e + f + h + i, b + c + d + g)$
(211)(211) trans	G_1g	T_1T_2	$G_2g > 0$	$T_2t > 0$	$z_1(b + f + i, c + e + h) + c_1(b + c + e + f + h + i)$
(211)(211) trans	G_1G_2	T_1t	complete	arbitrary	$z_1(d + h + i, e + f + g)$
(211)(1111)	G_1g	T_1T_2	arbitrary	complete	$z_1(b + f + h + l, c + e + i + k)$
(211)(1111)	G_1g	T_1T_2	$G_2g > 0$	$T_2 > 0$	$z_1(b + f + h + l, c + e + i + k) + c_1(b + c + e + f + h + i + k + l)$
(211)(1111)	G_1G_2	T_3T_4	complete	arbitrary	$z_1(a + d + h + i + k + l, b + c + e + f + g + j)$
(1111)(1111)	G_1G_2	T_1T_2	complete	arbitrary	$z_1(a + c + f + h + i + k + n + p, b + d + e + g + j + l + m + o)$
(1111)(1111)	G_1G_2	T_1T_2	arbitrary	complete	$z_1(a + c + f + h + i + k + n + p, b + d + e + g + j + l + m + o)$
(1111)(1111)	G_1G_2	T_1T_2	$G_2 > 0$	$T_2 > 0$	$z_1(a + c + f + h + i + k + n + p, b + d + e + g + j + l + m + o) + c_1$
(1111)(1111)	G_3G_4	T_3T_4	complete	arbitrary	$z_1(a + b + e + f + k + l + o + p, c + d + g + h + i + j + m + n)$
(1111)(1111)	G_3G_4	T_3T_4	arbitrary	complete	$z_1(a + b + e + f + k + l + o + p, c + d + g + h + i + j + m + n)$
(1111)(1111)	G_3G_4	T_3T_4	$G_4 > 0$	$T_4 > 0$	$z_1(a + b + e + f + k + l + o + p, c + d + g + h + i + j + m + n) + c_1$

at least one A_1 or A_2 child from the mating $A_1A_2 \times B0$). Finally, the condition for selection may be $T_2T_4 > 0$ (as at least one MNs child in the mating Ms/NS × MS/Ns). For brevity the three types of selection for 1111 test factors may be denoted by $T_4 > 0$; T_2 , $T_4 > 0$; and $T_2T_4 > 0$.

Smith (1954) has developed maximum likelihood u scores for separation of the parental recombination values in double intercrosses. This facilitates a test of homogeneity of recombination in mothers and fathers and in many cases simplifies the scoring procedure. The parental scores are confounded if both factors are of type 31 or 121. This approach is also applicable to lods, and results in a considerable simplification of the scores. It is necessary to fix the segregation of the first parent in obtaining the score for the second. The order in which the parents are considered is arbitrary, but the scoring is simpler if the parent with the more restrictive selection of the test factor is taken first. Selection of a family through children not scorable for linkage is equivalent to complete selection for the children scored. With the exception of types (31)(31), (31)(121), and (121)(121), all families of known parental genotype may be scored for each parent separately, and apart from a negligible dependence in T_2T_4 selection the scores are independent (table 2). With T_2T_4 selection, the scores for $T_2 > 0$ and $T_4 > 0$ are formally correct and strictly independent in the absence of linkage. For examples of the procedure see Steinberg and Morton, 1956.

This very simple analysis is all that most data require, and since most families

TABLE 3. DOUBLE INTERCROSS SCORES FOR THE PARENTS JOINTLY

Mating class	Selection		Score	Notes
	G	T		
(31)(31)	arbitrary	$t > 0$	$z_3(a, b + c, d) + e_3$	See tables 12, 13, 17 (Morton, 1955a)
	$g > 0$	$t > 0$	$z_3(a, b + c, d) + c_3$	
(31)(121)	$g > 0$	complete	z_4	For preliminary analysis use $z_3(a + e, b + c + f, d)$
(121)(121)	arbitrary	complete	z_5	For preliminary analysis use $z_3(a + c + g + i, b + c + f + h, d)$
(31)(211)	$g > 0$	$T_2t > 0$	$z_6 + c_3$	
	arbitrary	$T_2t > 0$	$z_6 + e_3$	
(31)(1111)	$g > 0$	$T_4 > 0$	$z_7 + c_2$	
	$g > 0$	$T_2, T_4 > 0$	$z_7 + c_7$	
	$g > 0$	$T_2T_4 > 0$	$z_7 + c_3$	
	arbitrary	$T_4 > 0$	$z_7 + e_2$	
	arbitrary	$T_2, T_4 > 0$	$z_7 + e_7$	For preliminary analysis score parents separately (table 2)
	arbitrary	$T_2T_4 > 0$	$z_7 + e_3$	
(121)(211)	complete	$T_2t > 0$	z_8	
	arbitrary	$T_2t > 0$	$z_8 + e_4$	
(121)(1111)	complete	arbitrary	z_9	
	arbitrary	$T_4 > 0$	$z_9 + e_1(s_1, s_3)$	
	arbitrary	$T_2, T_4 > 0$	$z_9 + e_9$	
	arbitrary	$T_2T_4 > 0$	$z_9 + e_4$	
(211)(211), (211)(1111), (1111)(1111)				Score parents separately (table 2)

TABLE 4. LOD FACTORS AND SCORES

$K = 4\theta_1(1-\theta_1) = PQ$
$L = 4\theta_1(2-\theta_1)/3 = PR$
$M = 4(1-\theta_1^2)/3 = QS$
$N = 4(2+\theta_1-\theta_1^2)/9 = RS$
$P = 2\theta_1$
$Q = 2(1-\theta_1)$

$R = 2(2-\theta_1)/3$
$S = 2(1+\theta_1)/3$
$T = 4(1-\theta_1+\theta_1^2)/3$
$U = 2(1+2\theta_1-2\theta_1^2)/3$
$V = 2(1-2\theta_1+2\theta_1^2)$
$W = 4(3-2\theta_1+\theta_1^2)/9$
$X = 4(2+\theta_1^2)/9$

$$z_1(a+d,\ b+c) = \log \tfrac{1}{2}\{P^{a+d}Q^{b+c} + Q^{a+d}P^{b+c}\}$$

$$z_2(a,\ b,\ c,\ d) = \log \tfrac{1}{2}\{R^aP^bS^cQ^d + S^aQ^bR^cP^d\}$$

$$z_3(a,\ b+c,\ d) = \log \tfrac{1}{4}\{W^aL^{b+c}Q^{2d} + 2N^aT^{b+c}K^d + X^aM^{b+c}P^{2d}\}$$

$$z_4 = \log \tfrac{1}{4}\{M^aP^{2b}L^cQ^{2d}T^cK^f + 2T^{a+c}K^{b+d}U^eV^f + L^aQ^{2b}M^cP^{2d}T^eK^f\}$$

$$z_5 = \log \tfrac{1}{4}\{Q^{2u+w}P^{2v+w}V^i + 2K^{u+v+i}V^w + P^{2u+w}Q^{2v+w}V^i\}$$

$$z_6 = \log \tfrac{1}{4}\{R^{a+c}P^{b+c+f}Q^{2d+f}T^e + R^{a+e}P^{b+d+e}T^cQ^{d+2f} + S^{a+e}Q^{b+d+e}T^cP^{d+2f} + S^{a+c}Q^{b+c+f}P^{2d+f}T^e\}$$

$$z_7 = \log \tfrac{1}{4}\{M^aP^{2b+d+f}T^{c+e}Q^{d+f+2h}L^g + T^{a+g}P^{b+2f+h}Q^{b+2d+h}L^cM^e + T^{a+g}Q^{b+2f+h}\cdot P^{b+2d+h}M^cL^e + L^aQ^{2b+d+f}T^{c+e}P^{d+f+2h}M^g\}$$

$$z_8 = \log \tfrac{1}{4}\{P^{a+d+f+h+2i}Q^{c+d+f+2g+h}V^e + Q^{a+d+f+h+2i}P^{c+d+f+2g+h}V^e + P^{a+e+2f+g+i}Q^{c+2d+e+g+i}V^h + Q^{a+e+2f+g+i}P^{c+2d+e+g+i}V^h\}$$

$$z_9 = \log \tfrac{1}{4}\{P^{a+c+2d+e+h+2i+j+l}Q^{a+c+e+2f+2g+h+j+l}V^{b+k} + Q^{a+c+2d+e+h+2i+j+l}\cdot P^{a+c+e+2f+2g+h+j+l}V^{b+k} + P^{2a+b+d+f+g+i+k+2l}Q^{b+2c+d+f+g+i+2j+k}V^{e+h} + Q^{2a+b+d+f+g+i+k+2l}P^{b+2c+d+f+g+i+2j+k}V^{e+h}\}$$

c_1, c_2, c_6. See table 13 (Morton, 1955a).

$$c_7 = \log \frac{4^{s+1}(4^s - 3^s - 2^{s+1} + 1) + 8(6^s) - 4(3^s)}{4^{s+1}(4^s - 3^s - 2^{s+1} + 1) + 4(6^s)(R^s + S^s) - 3^s(M^s + 2T^s + L^s)}$$

$$d_2(s_1, s_2) = \log \frac{2^{2s+1} - 2(3^s)}{2^{2s+1} - 3^s(R^{s_1}S^{s_2} + S^{s_1}R^{s_2})}$$

$$e_1(s_1, s_2) = \log \frac{2^{s+1} - 2}{2^{s+1} - (P^{s_1}Q^{s_2} + Q^{s_1}P^{s_2})}$$

$$e_2(s_1, s_2) = \log \frac{2^{s+1} - 2}{2^{s+1} - (R^{s_1}P^{s_2} + S^{s_1}Q^{s_2})}$$

$$e_3(s_1, s_2) = \log \frac{4^{s+1} - 4(3^s)}{4^{s+1} - 3^s(W^{s_1}L^{s_2} + 2N^{s_1}T^{s_2} + X^{s_1}M^{s_2})}$$

$$e_4(s_1, s_2, s_3) = \log \frac{4^{s+1} - 4(3^s)}{4^{s+1} - 3^s(M^{s_1}T^{s_2}L^{s_3} + 2T^{s_1+s_3}U^{s_2} + L^{s_1}T^{s_2}M^{s_3})}$$

$$e_7(s_1, s_2) = \log \frac{4^{s+1} - 2^{s+3} + 4}{4^{s+1} - 2^{s+2}(R^{s_1}P^{s_2} + S^{s_1}Q^{s_2}) + (M^{s_1}P^{2s_2} + 2T^{s_1}K^{s_2} + L^{s_1}Q^{2s_2})}$$

$$e_9 = \log \frac{4^{s+1} - 2^{s+3} + 4}{4^{s+1} - 2^{s+2}(P^{s_1}Q^{s_3} + Q^{s_1}P^{s_3}) + (P^{2s_1}K^{s_2}Q^{2s_3} + 2K^{s_1+s_3}V^{s_2} + Q^{2s_1}K^{s_2}P^{2s_3})}$$

TABLE 5. POWERS OF LOD FACTORS

$\Theta_1 = .05$

x	K^x	L^x	M^x	N^x	P^x	Q^x
1	.1900	.1300	1.330	.9100	10^{-x}	1.900
2	.03610	.01690	1.769	.8281		3.610
3	$.0^{2}6859$	$.0^{2}2197$	2.353	.7536		6.859
4	$.0^{2}1303$	$.0^{3}2856$	3.129	.6857		13.03
5	$.0^{3}2476$	$.0^{4}3713$	4.162	.6240		24.76
6	$.0^{4}4705$	$.0^{5}4827$	5.535	.5679		47.05
7	$.0^{5}8939$	$.0^{6}6275$	7.361	.5168		89.39
8	$.0^{5}1698$	$.0^{7}8157$	9.791	.4703		169.8
9	$.0^{6}3227$	$.0^{7}1060$	13.02	.4279		322.7
10	$.0^{7}6131$	$.0^{8}1379$	17.32	.3894		613.1

	R^x	S^x	T^x	U^x	V^x	W^x	X^x
1	1.300	.7000	1.270	.7300	1.810	1.290	.8900
2	1.690	.4900	1.613	.5329	3.276	1.664	.7921
3	2.197	.3430	2.048	.3890	5.930	2.147	.7050
4	2.856	.2401	2.601	.2840	10.73	2.769	.6274
5	3.713	.1681	3.304	.2073	19.43	3.572	.5584
6	4.827	.1176	4.196	.1513	35.16	4.608	.4970
7	6.275	.08235	5.329	.1105	63.64	5.945	.4423
8	8.157	.05765	6.768	.08065	115.2	7.669	.3937
9	10.60	.04035	8.595	.05887	208.5	9.893	.3504
10	13.79	.02825	10.92	.04298	377.4	12.76	.3118

$\Theta_1 = .10$

x	K^x	L^x	M^x	N^x	P^x	Q^x
1	.3600	.2533	1.320	.9289	.2000	1.800
2	.1296	.06418	1.742	.8628	.04000	3.240
3	.04666	.01626	2.300	.8015	$.0^{2}8000$	5.832
4	.01680	$.0^{2}4119$	3.036	.7445	$.0^{2}1600$	10.50
5	$.0^{2}6047$	$.0^{2}1043$	4.007	.6915	$.0^{3}3200$	18.90
6	$.0^{2}2177$	$.0^{3}2643$	5.290	.6424	$.0^{4}6400$	34.01
7	$.0^{3}7836$	$.0^{4}6696$	6.983	.5967	$.0^{4}1280$	61.22
8	$.0^{3}2821$	$.0^{4}1696$	9.217	.5543	$.0^{5}2560$	110.2
9	$.0^{3}1016$	$.0^{5}4298$	12.17	.5148	$.0^{6}5120$	198.4
10	$.0^{4}3656$	$.0^{5}1089$	16.06	.4782	$.0^{6}1024$	357.0

	R^x	S^x	T^x	U^x	V^x	W^x	X^x
1	1.267	.7333	1.213	.7867	1.640	1.249	.8933
2	1.604	.5378	1.472	.6188	2.690	1.560	.7980
3	2.032	.3944	1.786	.4868	4.411	1.948	.7129
4	2.574	.2892	2.167	.3830	7.234	2.433	.6369
5	3.261	.2121	2.630	.3013	11.86	3.038	.5689
6	4.130	.1555	3.191	.2370	19.46	3.794	.5083
7	5.232	.1141	3.871	.1864	31.91	4.739	.4540
8	6.627	.08364	4.697	.1467	52.33	5.918	.4056
9	8.394	.06134	5.699	.1154	85.82	7.391	.3623
10	10.63	.04498	6.915	.09076	140.7	9.231	.3237

TABLE 5—*Continued*

$\theta_1 = .20$

x	K^x	L^x	M^x	N^x	P^x	Q^x
1	.6400	.4800	1.280	.9600	.4000	1.600
2	.4096	.2304	1.638	.9216	.1600	2.560
3	.2621	.1106	2.097	.8847	.06400	4.096
4	.1678	.05308	2.684	.8493	.02560	6.554
5	.1074	.02548	3.436	.8154	.01024	10.49
6	.06872	.01223	4.398	.7828	$.0^{3}4096$	16.78
7	.04398	$.0^{2}5871$	5.629	.7514	$.0^{3}1638$	26.84
8	.02815	$.0^{2}2818$	7.206	.7214	$.0^{3}6554$	42.95
9	.01801	$.0^{2}1353$	9.223	.6925	$.0^{3}2621$	68.72
10	.01153	$.0^{2}6493$	11.81	.6648	$.0^{3}1049$	110.0

	R^x	S^x	T^x	U^x	V^x	W^x	X^x
1	1.200	.8000	1.120	.8800	1.360	1.173	.9067
2	1.440	.6400	1.254	.7744	1.850	1.377	.8220
3	1.728	.5120	1.405	.6815	2.515	1.615	.7453
4	2.074	.4096	1.574	.5997	3.421	1.895	.6758
5	2.488	.3277	1.762	.5277	4.653	2.224	.6127
6	2.986	.2621	1.974	.4644	6.328	2.609	.5555
7	3.583	.2097	2.211	.4087	8.605	3.062	.5037
8	4.300	.1678	2.476	.3596	11.70	3.592	.4566
9	5.160	.1342	2.773	.3165	15.92	4.215	.4140
10	6.192	.1074	3.106	.2785	21.65	4.946	.3754

$\theta_1 = .30$

x	K^x	L^x	M^x	N^x	P^x	Q^x
1	.8400	.6800	1.213	.9822	.6000	1.400
2	.7056	.4624	1.472	.9648	.3600	1.960
3	.5927	.3144	1.786	.9476	.2160	2.744
4	.4979	.2138	2.167	.9308	.1296	3.842
5	.4182	.1454	2.630	.9142	.07776	5.378
6	.3513	.09887	3.191	.8980	.04666	7.530
7	.2951	.06723	3.871	.8820	.02799	10.54
8	.2479	.04572	4.697	.8663	.01680	14.76
9	.2082	.03109	5.699	.8509	.01008	20.66
10	.1749	.02114	6.915	.8358	$.0^{2}6047$	28.93

	R^x	S^x	T^x	U^x	V^x	W^x	X^x
1	1.133	.8667	1.053	.9467	1.160	1.107	.9289
2	1.284	.7511	1.110	.8962	1.346	1.225	.8628
3	1.456	.6510	1.169	.8484	1.561	1.355	.8015
4	1.650	.5642	1.231	.8031	1.811	1.500	.7445
5	1.870	.4889	1.297	.7603	2.100	1.660	.6915
6	2.119	.4238	1.366	.7198	2.436	1.837	.6424
7	2.402	.3673	1.439	.6814	2.826	2.033	.5967
8	2.722	.3183	1.515	.6450	3.278	2.250	.5543
9	3.085	.2758	1.596	.6106	3.803	2.490	.5148
10	3.496	.2391	1.681	.5781	4.411	2.755	.4782

TABLE 5—*Continued*

$\theta_1 = .40$

x	K^x	L^x	M^x	N^x	P^x	Q^x
1	.9600	.8533	1.120	.9956	.8000	1.200
2	.9216	.7282	1.254	.9911	.6400	1.440
3	.8847	.6214	1.405	.9867	.5120	1.728
4	.8493	.5302	1.574	.9823	.4096	2.074
5	.8154	.4525	1.762	.9780	.3277	2.488
6	.7828	.3861	1.974	.9736	.2621	2.986
7	.7514	.3295	2.211	.9693	.2097	3.583
8	.7214	.2812	2.476	.9650	.1678	4.300
9	.6925	.2399	2.773	.9607	.1342	5.160
10	.6648	.2047	3.106	.9564	.1074	6.192

	R^x	S^x	T^x	U^x	V^x	W^x	X^x
1	1.067	.9333	1.013	.9867	1.040	1.049	.9600
2	1.138	.8711	1.027	.9735	1.082	1.100	.9216
3	1.214	.8130	1.041	.9605	1.125	1.154	.8847
4	1.295	.7588	1.054	.9477	1.170	1.210	.8493
5	1.381	.7082	1.068	.9351	1.217	1.270	.8154
6	1.473	.6610	1.083	.9226	1.265	1.332	.7828
7	1.571	.6170	1.097	.9103	1.316	1.397	.7514
8	1.676	.5758	1.112	.8982	1.369	1.465	.7214
9	1.788	.5374	1.127	.8862	1.423	1.537	.6925
10	1.907	.5016	1.142	.8744	1.480	1.612	.6648

are of tabulated scoring types (z_1, z_2, z_3), it should be satisfactory to score the (31)(121) and (121)(121) types with z_3 as (31)(31), neglecting information given by codominance in order to use tabulated scores for all matings. However, if the data suggest linkage (say if $\Sigma z > 2$), extra effort may be worthwhile to extract every bit of information. To do this, the (31)(121) and (121)(121) matings should be scored efficiently with z_4 and z_5. Some other double intercrosses give more information by considering the parents together instead of separately; this is fully efficient if the recombination value is the same in both sexes. These joint scores are given in Table 3, and may be calculated with the aid of Tables 4 and 5 which may also be used to compute scores for large families beyond the range of previous tables. Finally, families of uncertain parental genotype may be scored from the gene frequencies on the assumption of random mating. For rare dominants, the amount of information can be appreciably increased by computing probabilities of large pedigrees, including doubtful families, but the information given by such families is almost negligibly small in data from only two generations (Finney, 1940).

Incomplete penetrance and disturbed viability are considered elsewhere (Morton, 1955a, page 301).

ALLELES AND PSEUDOALLELES

One of the most important problems in linkage analysis is the detection of allelic and pseudoallelic relationships; this may be done efficiently by taking $\theta_1 = .05$. If

there is evidence for linkage of two genes, and if the data include no certain recombinant, then the hypothesis of allelism may be entertained, especially if the two genes have similar effects. However, there is always a possibility that more extensive observation will reveal recombination, and until this is demonstrated a linkage analysis can only provide a confidence interval of the type $0 \leqslant \theta < \theta'$, where θ' is small. Methods will now be given for evaluating θ'.

All of the matings usable in linkage tests give some information about allelism, but the most important sibships are of the z_1 type, with probability proportional to $\theta^w(1-\theta)^x + \theta^x(1-\theta)^w$; this includes double backcrosses (a + d = w, b + c = x) and two classes (b = w, d = x) of single backcrosses. If w or x is zero, there is no certain recombinant, but if both w and x are greater than zero, there are | w-x | certain recombinants and the hypothesis of allelism is disproved, assuming legitimacy and no errors of classification. Suppose there is no certain recombinant. If there is complete selection for one or both loci, the score for allelism is

$$z^0 = \log \frac{f(y;\theta)}{f(y;0)} = \log \{\theta^s + (1-\theta)^s\},$$

where w + x = s and z^0 is obtained from z_1 by substitution of f(y; 0) for f(y; 1/2).

If there is truncate selection for both loci,

$$\log \frac{f(y;\theta)}{f(y;0)} = z^0 + c^0, \qquad \text{where } c^0 = \log \frac{2^{s+1}-3}{2^{s+1}-4+\theta^s+(1-\theta)^s}.$$

If there is arbitrary selection at the main locus and truncate selection at the test locus,

$$\log \frac{f(y;\theta)}{f(y;0)} = z^0 + e^0.$$

If $s_1 > 0$ and $s_2 > 0$, $e^0 = -\log \{1 - \frac{1}{2}\theta^{s_1}(1-\theta)^{s_2} - \frac{1}{2}\theta^{s_2}(1-\theta)^{s_1}\}$, and if $s_1 = 0$ or $s_2 = 0$, $e^0 = -\log \{2 - \theta^s - (1-\theta)^s\}$. If a certain recombinant is observed, $z^0 = \infty$.

A small-sample test of allelism may be carried out sequentially against a fixed alternative θ_1 according to the rule $\log B < \Sigma z < \log A$ ($z = z^0$, $z^0 + c^0$, or $z^0 + e^0$), or against a class of alternatives $\theta' > 0$ as follows. Let $z_i(\theta)$ be the score for the i^{th} family and $Z(\theta) = \Sigma z_i(\theta)$. If there is at least one certain recombinant, we reject allelism and terminate the test. If a certain recombinant is not observed, then the test continues and the interval $\theta < \theta'$ corresponding to $Z(\theta)' < \log \alpha$ provides a confidence interval of strength $1 - \alpha$.

Unless the sample is extremely small, this procedure may be replaced by a still simpler one. For if there is no certain recombinant,

$$Z(\theta) = \log e^{-\theta \Sigma s} + 0(\theta), \qquad (s \geqslant 2)$$

and

$$\theta < \frac{\ln (1/\alpha)}{\Sigma s} \tag{1}$$

is a confidence interval of strength about $1 - \alpha$ if Σs is large. The matings not

TABLE 6. A COMPARISON OF SMALL- AND LARGE-SAMPLE CONFIDENCE INTERVALS OF THE TYPE $\Theta < \Theta'$ WHEN THERE IS NO CERTAIN RECOMBINANT

n = the required number of double-backcross sib pairs under truncate selection at both loci
α = the probability of no certain recombinant in a sample of size n if $\Theta = \Theta'$
Θ'^* = the large sample approximation to Θ'

Θ'	Θ'^*	$\frac{n}{\alpha = .10}$	$\frac{n}{\alpha = .01}$
.10	.081	15	29
.05	.040	29	58
.01	.008	144	288
.001	.001	1439	2878

scorable as double backcrosses, (31)(31), (31)(121), and (121)(121), may be used to detect recombination, but cannot easily be incorporated into equation (1). Familes of known linkage phase contribute to Σ s without the restriction $s \geqslant 2$.

The accuracy of this large-sample interval may be assessed by determining for different values of θ' and α the number n of families containing no certain recombinant required to make a statement of the type $P\{Z(\theta') < \log \alpha\} < \alpha$ and the large-sample interval (1) that would be formed on the same amount of data. For any type of family, $n = (\log \alpha)/z(\theta')$, and the corresponding large-sample limit is $\theta'^* = -(\ln \alpha)/ns$. Using natural logarithms for $z(\theta')$, $\theta'^* = -z(\theta')/s$, which is independent of α. Table 6 compares the large- and small-sample intervals for double-backcross sib pairs under truncate selection at both loci, which because of small family size and incomplete selection is as unfavorable to large-sample theory as any body of data likely to be met in practice. The absolute error is nearly proportional to θ' and of smaller magnitude. If the data are numerous enough to establish linkage and there is no certain recombinant, the large sample confidence interval (1) will clearly be accurate.

PARTIAL SEX LINKAGE

Homologous loci in the X and Y chromosomes are known in some insects and fish (Haldane, 1936), and reports of X-Y chiasmata and postreduction suggest that partial sex linkage may occur in other organisms (Matthey, 1951). In man, the material basis of partial sex linkage is in dispute. Cytologists have variously asserted that the sex chromosomes form pachytene chiasmata and are sometimes postreduced (Koller, 1937), that they never conjugate in meiosis and are always prereduced (Sachs, 1954), and that there is no Y chromosome (Oguma, 1930). The genetic evidence for partial sex linkage is as controversial, and the claim of Haldane (1936) to have detected this condition is not universally accepted (Woolf, 1953; Neel and Schull, 1954). Objections center about the alternative of sex-biased manifestation (viability, penetrance, and ascertainment), as discussed by Harris (1948). The "indirect" method of Haldane (1936), applicable when linkage phase is unknown, does not discriminate between sex-biased manifestation and linkage, and we must rely on cases where the parental phase can be inferred.

Consider first rare recessive traits when the propositi are offspring of unaffected but closely related parents. It is of course possible that the two recessive genes which

come together in the homozygous progeny are of independent mutational origin or are derived from a more remote, possibly unrecorded, common ancestor. However, if p is the gene frequency and F_i the contribution of the i^{th} common ancestor to the inbreeding coefficient of the progeny, the probability of an affected child is $p^2 + p(1 - p)\Sigma F_i$, so that if the largest term in ΣF_i is considerably greater than p or the other F_i, we shall not often be wrong if we assume that homozygosity is by descent from the nearest common ancestor. Haldane used the fact that, on the hypothesis of partial sex linkage, a gene in the father must be on his X chromosome if inherited through his maternal line and on his Y chromosome if inherited through his paternal line. If there are *a* normal and *b* affected children of the same sex as the paternal grandparent who is related to the mother, and *c* normal and *d* affected children of the opposite sex, and if the sex ratio is unity and penetrance complete, the probability of the sibship will be proportional to $(1 + \theta)^a(1 - \theta)^b(2 - \theta)^c\theta^d$, regardless of viability or ascertainment of abnormals.

If we now consider the possibility of sex-biased manifestation, it is obvious that in many cases (about one-half, if the bias is symmetrical) the association between affection and sex of the child will be in the same direction as expected from partial sex linkage. However, if the parents and immediate ancestors of the affected children are normal, there should be no bias with respect to the way in which the husband is related to his wife. This suggests the following test. For a given sibship, fix the numbers of males and females, and of abnormals and normals, and consider the set of two events formed by interchanging *a* with *c* and *b* with *d*. This corresponds to two determinants with the same absolute magnitudes but opposite signs, viz. $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ and $\begin{vmatrix} c & d \\ a & b \end{vmatrix}$. Under symmetrical sex-biased manifestation, these two events are equally frequent, but if there is partial sex linkage the conditional frequency of the observed type is

$$f(y;\theta) = \frac{(1 + \theta)^a(1 - \theta)^b(2 - \theta)^c\theta^d}{(1 + \theta)^a(1 - \theta)^b(2 - \theta)^c\theta^d + (2 - \theta)^a\theta^b(1 + \theta)^c(1 - \theta)^d}$$

and the lod score for partial sex linkage is $w_2 - z_2$, where

$$w_2 = a \log \{2(1 + \theta_1)/3\} + b \log \{2(1 - \theta_1)\} + c \log \{2(2 - \theta_1)/3\} + d \log \{2\theta_1\},$$

and z_2 is the score for a single backcross of unknown phase. The same linkage score is obtained under asymmetrical sex-bias, if relationships through the husband's male and female lines are equiprobable in the general population of consanguineous marriages. The linkage test is carried out as usual, according to the rule

$$-2 < \Sigma (w - z) < 3.$$

This test is not affected by symmetrical sex-bias regardless of the frequencies of different modes of relationship, nor by asymmetrical sex-bias if relationships through the husband's male and female lines are equally frequent. Unequal frequencies of the two modes of relationship (Morton, 1955b) can simulate linkage only if the same sex preponderates in relationship and affection. Fortunately the recessive factors claimed

TABLE 7. TESTS OF RECESSIVE PARTIAL SEX LINKAGE

Disease	Source	Σ (w − z)			
		θ_1			
		.1	.2	.3	.4
Retinitis pigmentosa without deafness	Haldane, 1936 Hanhart, 1939	−1.223	.334	.961	.921
Xeroderma pigmentosum	Macklin, 1952 Koller, 1948	−4.991	−2.296	−.816	−.040
Xeroderma pigmentosum	incl. Nerger, 1906	−11.187	−6.039	−3.049	−1.063
Achromatopsia	Haldane, 1936	.871	.809	.696	.474
Oguchi's disease	Haldane, 1936	.252	.400	.415	.298
Spastic paraplegia and ataxia	Haldane, 1941	1.971	1.768	1.448	.912

to be partially sex-linked do not give disturbed sex ratios except for Oguchi's disease, where there is an excess of affected males and of relationship through the husband's male line. None of these factors gives significant evidence for partial sex linkage (Table 7).

The information on which these scores are based is too small to detect or exclude loose linkage, so that nonsignificance of the scores cannot be considered conclusive evidence against partial sex linkage until more data are collected. However, the test is reasonably efficient, with average sample numbers of 22 pairs of affected progeny if $\theta = \theta_1 = .2$, or 45 pairs if $\theta = \theta_1 = .3$. For xeroderma pigmentosum, there is convincing evidence against a value of θ as small as .18, which Fisher (1936b) estimated by the indirect method. Curiously enough, the family of Barckmann (Nerger, 1906), which is contrary to partial sex linkage by the direct test, was the largest contributor to the indirect test. This family was omitted from Haldane's consanguineous matings, apparently because of a misunderstanding of the mode of relationship (Cockayne, 1933, erroneously lists the parents as second cousins). Even without this family, there is no suggestion that xeroderma pigmentosum is partially sex-linked.

An analogous procedure may be followed for rare dominant traits. Only sibships which inherited the gene through the father are informative. If the pedigree contains a normal and b affected children of the same sex as the paternal grandparent who transmitted the gene, and c normal and d affected children of the opposite sex, the score for partial sex linkage is $w_1 - z_1$, where

$$w_1 = (a + d) \log \{2\theta_1\} + (b + c) \log \{2(1 - \theta_1)\},$$

and z_1 is the usual score for a double backcross family of unknown phase. Unfortunately, the method is sensitive to sex-biased manifestation, which also tends to make the sex of affected children concordant with the sex of the affected grandparent. The most that can be expected of this scoring procedure is that it will attenuate the sex bias, since it is not much influenced by exceptional pedigrees, even large ones, the maximum values of $w_1 - z_1$ being log 2 for a single pedigree. However, if the test is significant, it is necessary to apply other methods before the hypothesis of partial sex linkage can be accepted. *Ex hypothesi*, there should be no relation between the

sexes of affected children and affected parents or maternal grandparents, while a positive relation is expected under sex-biased manifestation. Because dominant traits permit examination of affected relatives and no single method is conclusive, tests of partial sex linkage are more eclectic than with recessive traits.

Haldane (1936) proposed that dominant retinitis pigmentosa is partially sex linked in about 40 per cent of the 14 pedigrees he examined. Rywlin (1951) added 11 pedigrees, which showed no evidence for partial sex linkage. Applying my method to the 25 pedigrees, I obtain the following scores

θ	.1	.2	.3	.4
Z	−23.561	−12.489	−5.720	−1.087

The maximum score is .066 at $\hat{\theta} = .481$, which gives no evidence of linkage. The nonsignificance of this test, which does not attach as much weight to large pedigrees as Haldane's test, suggests that the peculiarity of his data is restricted to one or two large pedigrees. His totals clearly indicate the pedigrees responsible. As he pointed out, there is an asymmetrical difference between the progenies of men who inherited retinitis pigmentosa from their mothers and fathers, respectively (Table 8). Men who inherited the condition from their mothers give an excessive number of normal male children with no excess of affected females, and men whose inheritance is from their fathers show an excess of affected male children with no excess of normal females. Moreover, the excesses are largely due to two pedigrees, Bell 5 and 6, which can only be described as extraordinary under any simple genetic hypothesis. The first gives 28 affected children to 3 normal from informative backcross matings, and the second, 25 normal to 2 affected. The extreme asymmetry of the distributions does not in the least suggest partial sex linkage. With these two pedigrees omitted, Haldane's data are no longer significantly in favor of partial sex linkage, and when pooled with Rywlin's pedigrees give an estimated recombination fraction of 164/343 = .478, nearly identical to the value obtained by the w-z scores with Bell 5 and 6 included.

In addition to the nonsignificance of the w-z test and the asymmetry of the two exceptional pedigrees, there is another reason for dismissing evidence based on two peculiar pedigrees out of 25. "It is, of course, characteristic of these linkage studies that large families ··· supply far more information than smaller families. In the absence of methods of collection directed, critically, with a view to the requirements of genetic problems, this constitutes a serious weakness in the available data; for

TABLE 8. DATA ON DOMINANT RETINITIS PIGMENTOSA

Sex of children	Status of children	Haldane (1936)				Rywlin (1951)	Probability under partial sex linkage
		Grandfather affected		Grandmother affected			
		Bell 6	Others	Bell 5	Others		
Same as father's affected parent	affected	19	31	2	29	24	$(1 - \theta)/2$
	normal	1	26	6	31	33	$\theta/2$
Opposite	affected	9	21	0	30	23	$\theta/2$
	normal	2	24	19	38	33	$(1 - \theta)/2$

linkage ought not to be postulated on the strength of any single family, the recorders of which may have attached no high importance to the correct statement of the sex of the children. We have, in fact, adequate grounds for asserting linkage only when it appears to be indicated by the concurrent testimony of a number of accurately reported families" (Fisher, 1936b).

Partial sex linkage has been suggested for some conditions on data which do not approach significance. Macklin (1952) refuted an earlier claim (Snyder and Palmer, 1943) that idiopathic convulsive disorder gives some indication of partial sex linkage. One pedigree of dominant ataxia and paraplegia (Schut, 1951) showed 6 recombinants and 13 nonrecombinants with the X chromosome, but Haldane (1941) found 34 recombinants among 67 informative children in pedigrees of dominant spastic paraplegia, and Bell's monograph (1948) gives no hint of partial sex linkage in pedigrees of dominant ataxia. The excess of affected males in zygodactyly is more suggestive of sex-biased manifestation than of partial sex linkage (Pipkin and Pipkin, 1945).

Stephens, Perkoff, Dolowitz, and Tyler (1951) traced the inheritance of a semidominant gene for susceptibility to acute pyelonephritis with nerve deafness in adult males and mild renal infection (pyuria, cylindruria, and positive urine cultures with occasional dysuria) in females and young males. Several females and one male transmitted but did not manifest the trait. The inheritance is suggestive of complete sex linkage, except that among the 39 children of affected males there were two sons who at age 2 and 5, respectively, gave evidence of renal infection, and four normal daughters. Among 83 children of affected women there were 20 affected sons and 23 normal daughters. From this the authors concluded that the trait is partially sex linked, with a crossover value of 15.4 ± 5.8 per cent. An alternative is complete sex linkage, with the exceptional females asymptomatic carriers, and the exceptional males sporadics of different etiology who will not develop the characteristic deafness and acute pyelonephritis or transmit the gene. Several kindred with hereditary hematuria and nerve deafness (Sturtz and Burke, 1956) contain only one child from an affected father, a normal son. Children from affected mothers give the ratios expected from an autosomal or sex-linked dominant, with males more severely affected than females. Without a follow-up study, the evidence of this pedigree is not sufficient to establish partial sex linkage.

In a brief note Kaliss and Schweitzer (1943) suggested partial sex linkage of dominant "hemorrhagic diathesis". This trait is a mixture of genetic entities, with both sex-linked and autosomal sex-biased manifestation (for bibliography see Gates, 1946). The evidence they present is inadequate to discriminate partial sex linkage.

AUTOSOMAL LINKAGE

Usually a linkage test in man requires several investigations to detect or exclude moderate or even close linkage. Under these conditions a sequential test is appropriate and highly efficient. The choice of θ_1 is arbitrary, like the sample number in fixed-sample-size tests, but reasonable values of θ_1 are suggested by the average sample numbers for various types of data (Morton, 1955a). If we distinguish common test factors (ABO, Rh, MNS), less common test factors like Duffy, P, and Lutheran, rare "dominants" like elliptocytosis, and rare recessives like cystic fibrosis of the

pancreas, then the following choices for θ_1 will on the average lead to a decision about linkage within a few investigations:

	θ_1
common × less common test factors	.3
two less common test factors	.2
rare dominant × common test factor	.2
rare recessive × common test factor	.1
rare dominant × less common test factor	.1
rare recessive × less common test factor	.05
two rare factors	.05

Although the usual sequential test defined by $-2 < \Sigma z < 3$ is peculiarly appropriate when the data are collected by stages, a nonsequential test may be preferable if information on a particular pair of loci is virtually nonrecurrent, e.g. elliptocytosis × telangiectasia (Roberts, 1945), or for a review of linkage, or to truncate a sequential test if further collection of data does not seem worthwhile. The appropriate small-sample procedure is to accept linkage at a significance level of .001 or less if $\Sigma z > 3$ and to reject linkage or defer a decision if $\Sigma z < 3$. This significance level holds only for preassigned θ_1 ; clearly the data are nonsignificant if the maximum value of Σz is less than 3, but not conversely. Only in rather vast bodies of human linkage data will a large-sample test be entirely trustworthy.

A heterogeneity test should be performed as soon as linkage has been detected and differentiated from other phenomena with which it might be confounded. Small-sample theory has not been developed, but the likelihood ratio test is usually fairly accurate, and seems to be more reliable than the maximum likelihood test (Morton, 1956). If the data are homogeneous, a maximum likelihood estimate of linkage can be obtained from three neighboring points by simple interpolation, or from the frequency of certain recombinants if θ is small, say $<.05$. In the latter case, an accurate confidence interval may be assigned from binomial or Poisson tables. If θ is larger and the amount of data large, a reliable confidence interval is given by normal theory (Morton, 1956). Where this is not appropriate we may use a normalizing transformation, for an approximate confidence interval, or an exact formula of Haldane and Smith (1947). They showed that the inequality

$$Z' > \log \Lambda - \log A \quad (A > 1)$$

where Z' is a lod score and

$$\Lambda = 2 \int_0^{1/2} \Pi \frac{f(y; \theta)}{f(y; 1/2)} \, d\theta,$$

provides, by projection on the θ axis, a confidence interval of strength at least $1 - 1/A$, without any assumption except that θ is constant. In general Λ must be estimated by numerical integration, but if linkage is highly significant and there is at least one certain recombinant, the maximum value of Σz is

$$\hat{Z} \cong \log \left\{ \frac{\Lambda}{2\sigma\sqrt{2\pi}} \right\},$$

and

$$\log \Lambda \cong \hat{Z} + \log \sigma + .7001.$$

These methods have been used to analyze such of the reported autosomal linkages in man as can be scored by lods. For this it is necessary that both traits depend on single, regular mendelian factors. Polygenic associations, pleiotropy "possibly due to several completely linked genes", and various ill-defined genetic correlations are omitted (for these see Gates, 1954). Except where linkage is suggested ($\Sigma z > 2$) only families of known parental genotype have been scored, distinguishing in pedigree data between families of known and unknown parental phase.

Hogben and Pollack (1935) analyzed 12 families segregating for recessive Friedreich's ataxia and ABO, with no evidence of linkage by Bernstein's method. Fisher (1936a) applied his u scores, remarking that "At this early stage in the search for linkage in human genetics, even small collections of data have a certain importance in supplying experience of the kinds of difficulties and inconsistencies likely to be met with, which, if their presence is not suspected, will be liable to give rise to misleading conclusions. Caution is more than ordinarily necessary in this connection, since methods have been developed on purely theoretical considerations, and without experience of their practical reliability." He found $\chi_1^2 = 2.34$ in favor of linkage ($P = .06$) and $\chi_{11}^2 = 29.14$ for homogeneity ($P < .01$). The suggestion of linkage and heterogeneity was entirely due to one family which provided "decisive evidence either of linkage, or of the heterogeneity of the twelve families reported". However, Haldane (1946) noted that in the absence of linkage the probability of one such family among several was not small and that the u scores were apparently unreliable. This is also suggested by the probability ratio analysis. The maximum score $\hat{Z}$ is only .44 and the heterogeneity χ_{11}^2 only 5.51 ($P > .9$), giving no hint of heterogeneity or linkage. Steinberg and Morton (1956) presented another example of apparently unreliable u scores.

During the last twenty years several linkages have been suggested on nonsignificant evidence. Finney (1940) found a t of 1.70, $P = .045$, for linkage between allergy and ABO, assuming a model for the inheritance of allergy that has not been generally accepted. The probability ratio analysis on the same hypothesis gives $\hat{Z} = .74$, which does not approach significance. Bernstein, Borison, and Finkel (1943) have criticized Finney's claim on other grounds. Burks and Wyandt (1941) found $t = 1.53$, $P = .06$, for linkage between elliptocytosis and ABO, but did not publish the family data. Subsequent studies on elliptocytosis have not suggested linkage with ABO. Bianco, Ceppellini, Silvestroni, and Siniscalco (1954) reported that u scores gave $t = 3.12$, $P < .001$, for linkage between thalassemia and the Lewis^a erythrocyte factor. Through the kindness of Dr. Ceppellini I have been able to apply lod scores to these data, and find $\hat{Z} = .90$. The very slight suggestion of linkage is largely due to a single family, which has probability 1/16 on the null hypothesis. There is also no evidence for linkage between thalassemia and ABO secretion. Snyder (1949) reported linkage between MN and sickling, without distinguishing homozygotes from heterozygotes. As he classified them, his twelve informative families give a χ_1^2 for linkage of 16.64, $P < .001$. Analyzed in the same way, the lod score maximum is 2.26, which is not

significant. His first two families did not segregate for sickling and his four intercrosses contribute no information if homozygotes are sublethal. Omitting these six families, the u-score test gives $\chi^2 = 11.56$, $P < .001$, but the lod score is still nonsignificant ($\hat{Z} = 1.71$). Snyder's data comprise at most 31 units of u-score information. Two later studies, which contribute nearly 300 units, give no evidence of linkage (Waller, Waller, and Hughes, 1952; Neel, Schull, and Shapiro, 1952). This provides a practical demonstration of the reliability of lod scores, the necessity for a stringent significance level, and the unreliability of large-sample linkage tests in man.

Mohr (1954) obtained a strong suggestion of linkage between the erythrocyte antigens Lewis^a and Lutheran ($t = 4.29$, $P < .0001$). Including three additional families (Holt, Thompson, Sanger, and Race, 1952; Mohr, 1956), the maximum lod score is 4.18 at $\hat{\theta} = .06$, and $\Sigma z = 3.17$ against the reasonable alternative of $\theta_1 = .20$. However, there are several reasons for caution in interpreting this result. The evidence for linkage is largely contributed by a single family, number 61, and does not approach significance with this family excluded, although the other families are not significantly discordant with it. Mohr considered Le^a to be recessive, but his families contain an unexpectedly high frequency of children with the antigen ($\chi^2_2 = 5.98$, $P \cong .05$), and the assumed mode of inheritance is probably incorrect (Andresen and Henningsen, 1951). It seems best to be somewhat cautious about this report until studies can be made with Lewis and ABO secretions (Grubb and Morgan, 1949).

Two clear-cut cases of linkage have recently been found by workers at the Galton Laboratory (Renwick and Lawler, 1955). Elliptocytosis is very closely linked to the Rh blood system in four large pedigrees but not linked in three others, which suggests that linkage analysis has subdivided elliptocytosis into two distinct genetic, and presumably phenotypic, entities (Morton, 1956). If this is true, it is the first case in man of the resolution of heterogeneity by autosomal recombination analysis, a method that has had some success for mixed autosomal and sex-linked inheritance in muscular dystrophy, gargoylism, and other traits (Walton, 1955; Herndon, 1954). In families of known parental genotype the cumulative score for $\theta_1 = .2$ exceeds two in the fifth pedigree, and becomes significant in the eighth. The total pedigree score is 7.41, leaving no doubt of the existence of linkage. For the four large pedigrees which show close linkage the estimate of recombination is 3.3 ± 2.3 per cent.

The other proven case of linkage involves the nail-patella syndrome and the ABO blood group (Renwick and Lawler, 1955). Table 9 gives the probability ratio analysis for these data. Heterogeneity is of doubtful significance, in contrast with the maximum likelihood analysis, which erroneously indicates gross heterogeneity between pedigree E and the rest of the data. Renwick and Lawler showed that the P-value for this heterogeneity is greater than .032. The maximum likelihood heterogeneity test may break down when there is no certain recombinant, but the likelihood ratio test does not. Using $\theta_1 = .05$, $\theta_2 = .10$, $\theta_3 = .15$ (Morton, 1956, equation 2), the following estimates are obtained:

$$\hat{\theta} = .110, \sigma = .031, \hat{Z} = 9.62, \log \Lambda = 8.81,$$

which are in close agreement with the values obtained by Renwick and Lawler from more elaborate calculations. A recent pedigree (Jameson, Lawler, and Renwick,

TABLE 9. LOD SCORES FOR LINKAGE OF ABO AND NAIL-PATELLA LOCI
(Renwick and Lawler, 1955)

θ	Complete pedigrees					Total	Families of known parental genotype
	A	B	C	D	E		
.05	1.432	2.912	−1.792	.958	5.301	8.812	2.833
.10	1.430	2.984	−.709	1.117	4.775	9.598	3.802
.15	1.296	2.794	−.177	1.113	4.227	9.253	
.20	1.105	2.488	.092	1.032	3.656	8.374	3.811
.30	.654	1.650	.287	.752	2.446	5.789	2.886
$\hat{\theta}$	1.46	3.01	.30	1.12	5.81	9.62	

Heterogeneity $\chi^2_4 = 9.58$, P $\backsim$ .045

1956) confirms linkage and absence of significant heterogeneity. Most of the information in this record is contributed by 20 double-backcross progeny, with only two certain recombinants. Considering the two lineages from III.4 and III.9 separately, the maximum score is 2.60. Pooling this pedigree with the other five, the corresponding estimates are:

$$\hat{\theta} = .109, \sigma = .028, \hat{Z} = 12.22, \log \Lambda = 11.36, \chi^2_5 = 9.58 \text{ (P } \sim .09).$$

There is no substantial evidence against the hypothesis that the recombination fraction is the same in all six pedigrees. In contrast with elliptocytosis, the nail-patella syndrome appears on present evidence to be controlled by the same locus in different families.

DISCUSSION

Linkage studies in man may be greatly disturbed by genetic heterogeneity, nonrandom assortment of chromosomes, chromosomal aberrations, heritable and environmental variations in recombination rate, and so forth. However, there is no strong evidence to suggest that lod scores will not give reliable answers when applied to appropriate test traits. The utility of linkage analysis is obvious, since the first established autosomal linkage apparently discriminates between two previously unrecognized genetic entities. Such studies will ultimately provide a sounder basis for estimates of single-locus mutation rates, gene frequencies, and selection effects, as well as for the classification of medical entities. Linkage maps may also be of occasional use in the detection of genetic carriers. The problem remains of deciding what markers to use, how much data to collect, and how to do it most efficiently.

A relevant datum is the probability of detecting linkage in routine tests. Let us suppose that human autosomes are about 100 centimorgans in lengths, in which case the recombination fraction θ will have a nearly uniform distribution for linked genes, with a probability of about .05 that a random pair of genes be linked (Morton, 1955a). Assume also that the average power for double backcross sib pairs is approximately correct for other types of data (Morton, 1955a, Fig. 5), that the test loci are in different chromosomes, and that θ_1 is chosen as in the previous section (which will often require that several investigations be pooled). This means that if θ is constant,

the probability of detecting linkage to a particular test factor is about (.05)(.28) = .014 for $\theta_1 = .05$, .0195 for $\theta_1 = .1$, .028 for $\theta_1 = .2$, and .0355 for $\theta_1 = .3$ (Morton, 1955a, Table 2). If there are x common test factors and n less-common test factors, the chance of detecting linkage is about D = .028x + .0195n for rare dominant main factors, and about R = .0195x + .014n for rare recessives. Thus with ABO, Rh, and MNS alone, D = .084 and R = .058, but with 8 additional factors (P, Lutheran, Kell, Duffy, Kidd, etc.), D = .24, R = .17. The chance of detecting linkage with rare recessives is not so much less than for rare dominants as might be supposed from the differences in the appropriate values of θ_1. However, heterogeneity of θ tips the balance considerably in favor of rare dominants, since large pedigrees are more likely than small families to permit recognition and an incisive analysis of linkage. For a test against $\theta_1 = .3$ the probability of detecting linkage with double backcross sib pairs is reduced from .0355 per test factor, if θ is constant, to about .028 if for each family nonlinkage and linkage with constant intensity θ are equiprobable (Morton, 1956). Loss of power under genetic heterogeneity is minimized with a large number of markers, since there will be some chance of detecting linkage with more than one of them. Of course, if heterogeneity in the main factor involves many loci, the chance of detecting linkage with any one of them must be very small, except perhaps in large pedigrees.

The probability of detecting linkage in an extensive study of a rare trait is large enough when θ is constant or bimodal so that the discovery of two linkages by the Galton Laboratory in a few tests is no cause for astonishment. It would seem that routine linkage tests could profitably be carried out in any large-scale family study of a rare recessive, and especially of a rare dominant, and that in the absence of a number of very common test factors, less common traits like Duffy and Lewis can contribute considerably to the probability of detecting linkage.

SUMMARY

Lod scores can be applied to multiple allelic intercrosses very easily when the parental sexes are kept separate, and laboriously but more efficiently by joint scores. Allelism tests and confidence intervals are obtained for the case of no certain recombinant.

Scores which discriminate partial sex linkage from sex-biased manifestation do not confirm claims of partial sex linkage. No case is significant, and there is evidence that earlier tests on retinitis pigmentosa and xeroderma pigmentosum confounded sex-bias with partial sex linkage.

No significant evidence is found for most of the reported autosomal linkages. In at least two cases it is apparent that the lod scores are more reliable than large-sample tests. Linkage between the Lutheran antigen and the Lewis or ABO secretions is likely, but more study of the secretions is required. Two established cases of autosomal linkage are elliptocytosis with Rh, and nail-patella syndrome with ABO.

The probability of detecting linkage in random tests with 11 independent test factors may be as much as .24 for rare dominants and .17 for rare recessives. It is suggested that routine linkage tests would be profitable in extensive family studies of rare genes.

REFERENCES

ANDRESEN, P. H. AND K. HENNINGSEN. 1951. The Lewis blood group system and the X-system in 5 tables. *Acta Haemat.* 5: 123–127.

BELL, JULIA. 1948. *The Treasury of Human Inheritance, vol. IV.* Cambridge University Press.

BERNSTEIN, F., H. L. BORISON, AND S. FINKEL. 1943. A case of spurious linkage as claimed for the classical blood groups and allergy. *J. Immunol.* 46: 245–248.

BIANCO, I., R. CEPPELLINI, E. SILVESTRONI, AND M. SINISCALCO. 1954. Data for the study of linkage in man. Microcythaemia and the Lewis-Secretor characters. *Ann. Human Genet.* 19: 81–89.

BURKS, B. AND H. WYANDT. 1941. Oval blood cells in human subjects tested for linkage with taste for PTC, mid-digital hair, hair color, A-B agglutinogens and sex. *Genetics* 26: 223–233.

COCKAYNE, E. A. 1933. *Inherited Abnormalities of the Skin and its Appendages.* London: Oxford University Press.

FINNEY, D. J. 1940. The detection of linkage. *Ann. Eugen.* 10: 171–214.

FISHER, R. A. 1936a. Heterogeneity of linkage data for Friedreich's ataxia and the spontaneous antigens. *Ann. Eugen.* 7: 17–21.

FISHER, R. A. 1936b. Tests of significance applied to Haldane's data on partial sex linkage. *Ann. Eugen.* 7: 87–104.

GATES, R. R. 1946. *Human Genetics.* New York: MacMillan.

GATES, R. R. 1954. *Genetic Linkage in Man.* The Hague: Junk.

GRUBB, R. AND W. T. MORGAN. 1949. The "Lewis" blood group characters of erythrocytes and body fluids. *Brit. J. Exp. Path.* 30: 198–208.

HALDANE, J. B. S. 1936. A search for incomplete sex-linkage in man. *Ann. Eugen.* 7: 28–57.

HALDANE, J. B. S. 1941. The partial sex-linkage of recessive spastic paraplegia. *J. Genet.* 41: 141–147.

HALDANE, J. B. S. 1946. The cumulants of the distribution of Fisher's 'u_{11}' and 'u_{31}' scores used in the detection and estimation of linkage. *Ann. Eugen.* 13: 122–134.

HALDANE, J. B. S., AND C. A. B. SMITH. 1947. A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. *Ann. Eugen.* 14: 10–31.

HARRIS, H. 1948. On sex limitation in human genetics. *Eugen. Rev.* 40: 70–76.

HERNDON, C. N. 1954. Genetics of the lipidoses. *Proc. Assoc. Res. Nerv. and Ment. Dis.* 33: 239–258.

HOGBEN, L., AND R. POLLACK. 1935. A contribution to the relation of the gene loci involved in the isoagglutinin reaction, taste blindness, Friedreich's ataxia and major brachydactyly of man. *J. Genet.* 31: 353–361.

HOLT, H. A., J. S. THOMPSON, RUTH SANGER, AND R. R. RACE. 1952. Linkage relations of the blood group genes of man. *Heredity* 6: 213–216.

JAMESON, R. J., SYLVIA D. LAWLER, AND J. H. RENWICK. 1956. Nail-patella syndrome: clinical and linkage data on family G. *Ann. Human Genet.* 20: 348–361.

KALLIS, N., AND M. D. SCHWEITZER. 1943. Hereditary hemorrhagic diathesis—a case of partial sex-linkage in man? *Genetics* 28: 78–79.

KOLLER, P. C. 1937. The genetical and mechanical properties of sex chromosomes. III. Man. *Pr. Roy. Soc. Edinb.* 57: 194–214.

KOLLER, P. C. 1948. Inheritance of xeroderma and its chromosome mechanism. *Brit. J. Cancer* 2: 149.

MACKLIN, M. T. 1952. Is idiopathic convulsive disorder partially sex-linked? *Am. J. Human Genet.* 4: 356–362.

MATTHEY, R. 1951. The chromosomes of the vertebrates. *Advances in Genetics* 4: 159–180.

MOHR, J. 1954. *A study of linkage in man.* Opera ex Domo Biol. Hered. Humanae Univ. Hafn. 33.

MOHR, J. 1956. To what extent has linkage between various human blood group systems been excluded? *Act. Genet. Stat. Med.* 6: 24–34.

MORTON, N. E. 1955a. Sequential tests for the detection of linkage. *Am. J. Human Genet.* 7: 277–318.

MORTON, N. E. 1955b. Non-randomness in consanguineous marriage. *Ann. Human Genet.* 20: 116–124.

MORTON, N. E. 1956. The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am. J. Human Genet.* 8: 80–96.

NEEL, J. V., W. J. SCHULL, AND H. S. SHAPIRO. 1952. Absence of linkage between the genes responsible for the sickling phenomenon, the MN blood types, and the S-agglutinogen. *Am. J. Human Genet.* 4: 204–208.

NEEL, J. V., AND W. J. SCHULL. 1954. *Human Heredity.* Chicago: Chicago University Press.

NERGER, B. 1906. *Beiträge zur Kenntnis des Xeroderma pigmentosum.* Thesis, Kiel: Schmidt and Klaunig. pp. 23.

OGUMA, K. 1930. A further study of the human chromosomes. *Arch. de Biol.* 40: 205–226.

PIPKIN, A. C., AND SARAH B. PIPKIN. 1945. Two new pedigrees of zygodactyly. *J. Hered.* 36: 313–316.

RENWICK, J. H., AND S. D. LAWLER. 1955. Genetical linkage between the ABO and nail-patella loci. *Ann. Human Genet.* 19: 312–331.

ROBERTS, J. A. F. 1945. Genetic linkage in man, with particular reference to the usefulness of very small bodies of data. *Quart. J. Med. N. S.* 14: 27–33.

RYWLIN, A. 1951. Recherches de marqueurs de chromosomes dans une famille atteinte de rétinite pigmentaire dominante. *Acta Genet. Stat. Med.* 2: 85–100.

SACHS, L. 1954. Sex-linkage and the sex chromosomes in man. *Ann. Eugen.* 18: 255–261.

SCHUT, J. W. 1951. Hereditary ataxia; a survey of certain clinical, pathologic and genetic features with linkage data on five additional hereditary factors. *Am. J. Human Genet.* 3: 93–110.

SMITH, C. A. B. 1954. The separation of the sexes of parents in the detection of linkage in man. *Ann. Eugen.* 18: 278–301.

SNYDER, L. H., AND D. M. PALMER. 1943. An idiopathic convulsive disorder with deterioration apparently dependent upon an incompletely sex-linked gene. *J. Hered.* 34: 207–212.

SNYDER, L. H. 1949. The linkage relationships of sickle cell anemia. *Proc. Eighth Internat. Congr. Genet., Hereditas, suppl.*, 446–450.

STEINBERG, A. G., AND N. E. MORTON. 1956. Sequential test for linkage between cystic fibrosis of the pancreas and the MNS locus. *Am. J. Human Genet.* 8: 177–189.

STEPHENS, F. E., G. T. PERKOFF, D. A. DOLOWITZ, AND F. H. TYLER. 1951. Partially sex-linked dominant inheritance of interstitial pyelonephritis. *Am. J. Human Genet.* 3: 303–313.

STURTZ, G. S., AND E. C. BURKE. 1956. Hereditary hematuria, nephropathy, and deafness: preliminary report. *N. England J. M.* 254: 1123–1126.

WALLER, MARION, R. K. WALLER, AND R. D. HUGHES. 1952. Linkage studies of genes for sickling and MNS blood groups: negative findings. *Proc. Soc. Exp. Biol. N. Y.* 80: 479–481.

WALTON, J. N. 1955. On the inheritance of muscular dystrophy. *Ann. Human Genet.* 20: 1–12.

WOOLF, B. 1953. in SMITH, C. A. B. The detection of linkage in human genetics. *J. Roy. Stat. Soc. B* 15: 153–192.