A Comparison of the Two-Generation and Three-Generation Methods of Estimating Linkage Values on the X Chromosome in Man with Special Reference to the Loci Determining the Xg Blood Group and Glucose-6-Phosphate Dehydrogenase Deficiency

G. R. FRASER[1,2] AND O. MAYO[1]

Substantial discrepancies have occurred between estimates, derived by the two-generation and three-generation methods of analysis, of the recombination fraction between the loci determining the Xg blood group and G6PD deficiency. The results of the various surveys undertaken to study this point are summarized in the companion paper (Stamatoyannopoulos *et al.*, 1968). Briefly, initial studies using mainly Morton's (1955) *lod*-score method of analysis on families consisting of two generations in Greece and in Israel gave rise to an estimate of a recombination fraction of 0.26 between these loci. Later surveys in both of these countries, however, using mainly the three-generation method, in which study of the maternal grandfather permits a direct count of recombinant and nonrecombinant chromosomes, failed to confirm these indications of measurable linkage.

Siniscalco *et al.* (1966) also found no evidence that the two loci were within measurable distance of each other from the results of a survey in Sardinia in which the three-generation method was largely used. They pointed out, however, that if their results of the testing of the maternal grandfather in three-generation pedigrees from Sardinia are ignored and information only from the remaining two generations in the same families is used, the estimate of the recombination fraction is 0.32. These authors state further: "Had we continued to test mostly two-generation families we should still be confident that the loci for Xg and G6PD were within measurable distance of each other."

In an attempt to resolve these discrepancies, various populations were generated by Monte Carlo methods on a CDC 6400 computer. In this way, it has been possible to obtain multiple estimates ($\hat{\theta}$) of the recombination fraction (θ) from populations which correspond in size to those from which actual linkage data have been obtained and, in addition, from much larger populations, and to verify how closely these estimates correspond to the input values (θ_1).

Received March 11, 1968.

A portion of this work formed part of a Ph.D. dissertation prepared by one of the authors (O. M.) during the tenure of an Australian Commonwealth Postgraduate Award at the University of Adelaide.

[1] Department of Genetics, University of Adelaide, South Australia.

[2] Present address: Department of Medicine (Division of Medical Genetics) and Department of Preventive Medicine, University of Washington, Seattle, Washington 98105.

METHODS AND RESULTS OF SIMULATION EXPERIMENTS

In the first series of experiments, 1,000 grandparental matings were generated, using simulation techniques described by Mayo (1967). Briefly, the procedure consisted of choosing at random (using pseudorandom numbers) three X chromosomes (two for the grandmother and one for the grandfather). The probabilities, referred to each X chromosome, were .1 that the G6PD-deficient allele and .9 that the normal G6PD allele was selected, the alleles being represented by the pair of numbers 0 and 1. Linkage equilibrium was simulated by assigning, in the same way, independent probabilities of .6 for the Xg^a and .4 for the Xg allele. In these families, the grandfather's X chromosome was then included in the mother's genotype, and her other X chromosome was generated from those of the grandmother by first choosing at random, with probabilities of θ_1 and $1 - \theta_1$, whether this X chromosome would be recombinant or not and then continuing the random choice assigning probabilities of .5 to the two possible results in each case. Finally, this last process was repeated to generate two X chromosomes representing two sons from the mother's genotype.

The experiment was repeated once with each of the recombination fractions 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, and 0.75 as the input or "true" value (θ_1). The value of 0.75 was included because of the suggestion (J. H. Edwards, personal communication, 1966) that the discrepancies between results obtained by the two- and three-generation methods might be due to heterogeneity of the families studied and that these might be of two types, with recombination fractions of approximately 0.25 and 0.75, respectively. As is to be expected, the values of $\hat{\theta}$ when $\theta_1 = 0.75$ were indistinguishable from those when $\theta_1 = 0.25$, when the analysis was by the two-generation method, and were near 0.75 when the analysis was by the three-generation method. While the discrepancies mentioned could theoretically be due to such a phenomenon, it is a far less likely explanation than that adduced in this communication.

The output was analyzed in each case both including and excluding the grandparental mating. In the first case, a direct count was made of recombinant and nonrecombinant X chromosomes among the grandsons, while in the second the *lod* scores of Morton (1955), corrected by the factor appropriate to the ascertainment (e_1), were calculated on the computer for values of θ from 0.01–0.50 at intervals of 0.01. No a priori corrections (Renwick and Schulze, 1964) were applied.

The methods of ascertainment used in the studies in Sardinia, Greece, and Israel depend mainly on the presence of at least one G6PD-deficient son or grandson in the family, and, for purposes of comparison, the same procedure was adopted in the case of the computer-generated data which simulate a complete ascertainment of such G6PD-deficient individuals. In addition, an important distinction must be made in the three-generation pedigrees between models of ascertainment which include and models which disregard information obtained from the phenotype of the grandmother. This has often not been systematically tested in actual surveys in which the determination of the genotype of the maternal grandfather has been considered to be sufficient. Such a course has been followed also in the analysis of the simulated populations made here. These and other problems of ascertainment involving the three-generation method are discussed by Fraser (1968).

In the first two rows of Table 1, estimates of recombination fractions are set out from an analysis of these simulated populations, and comparisons can be made between the results of the two- and three-generation methods of estimation. In fact, the estimates are remarkably consistent between the two methods. Nevertheless, the three-generation method is considerably superior. This can be shown in various ways.

Using sequential analysis, Morton (1955) suggested that a *lod* score of +3 at a point $\theta = \theta'$ might constitute a significant indication of linkage of θ' or tighter. Even when sequential methods are not used, as in the present instance, a score of +3 is suggestive, although this does not correspond to any formal level of statistical significance, as it would in the case of a sequential test (Morton, 1957). Taking a score of +3 as a convenient yardstick of the discriminatory power of the test, Table 1 shows that in the simulated population under study, the three-generation method gives maximum

TABLE 1

RESULTS OF FIRST SERIES OF SIMULATION EXPERIMENTS USING THE TWO-GENERATION (A) AND THREE-GENERATION (B) MODELS OF ASCERTAINMENT

	Recombination Fraction							
Input (θ_1)	0.00	0.10	0.20	0.25	0.30	0.40	0.50	0.75
Output ($\hat{\theta}$):								
A	0.00	0.08	0.18	0.23	0.31	0.50	0.50	0.23
B	0.00	0.07	0.15	0.26	0.29	0.43	0.49	0.80
Maximum *lod* score $Z(\hat{\theta})$:								
A	.....	5.2178	1.7813	0.8154	0.1996	0	0	.....
B	.....	22.7701	14.2624	6.0089	4.8005	0.5687	0.0090	.....
Maximum value of θ at which $Z(\theta) < -2$:								
A	.....	0.00	0.04	0.07	0.12	0.18	0.17	.....
B	.....	0.00	0.00	0.07	0.13	0.28	0.33	.....

values greater than this for values of linkage up to 0.30, while the two-generation method is effective in this sense only at values up to 0.10. Another convenient criterion used in sequential tests is to exclude linkage at a level of θ' or tighter if the *lod* score at a point $\theta = \theta'$ is less than -2. The use of this parameter will give some indication of the range of linkage values which can be excluded by the test even though, again, in a nonsequential analysis, no formal level of statistical significance can be assigned. Table 1 shows that, for linkage values of 0.40 and 0.50 in the input, the levels at which linkage can be tentatively excluded are far higher in the three- than in the two-generation method.

In fact, in a population of this size, and it should be remembered that this is a far greater size than many actual populations from which estimates of linkage on the X chromosome have been derived, Table 1 shows that the two-generation method can be regarded as suitable only for values of θ_1 less than or equal to 0.20; at $\theta_1 = 0.25$, this method is already relatively ineffective, with a very low maximum *lod* score $Z(\hat{\theta})$. At

higher values of θ_1, such as 0.4 and 0.5, the two-generation method achieves very little discrimination, the ratio Prob($\theta = \hat{\theta}$)/Prob($\theta = 0.30$) being in both cases as little as 3. It would be hardly surprising, therefore, that in the small series actually studied in human populations, estimates of θ of the order of 0.30 should be obtained in the case of the loci determining Xg and G6PD, even if the true recombination fraction were in fact close to 0.50.

The possibility was explored further by a second series of simulated experiments. In these, sibships containing three sons were generated from doubly heterozygous mothers on whose X chromosomes the phase of the two loci was chosen at random. The genotypes of the three sons were then obtained by a repetition of the methods described in connection with the first series of experiments.

Linkage values of 0.10, 0.20, 0.25, 0.30, 0.40, 0.50, and 0.75 were introduced and the simulation was continued until a fixed number of ascertainable sibships was generated, an ascertainable sibship being one which could be used for linkage estimation in an actual population survey, that is, one in which the doubly heterozygous genotype of the mother at loci determining recessive traits could be deduced from her sons. As before, the appropriate correction factor (e_1) was applied to the z_1 score obtained from each sibship. Sibships containing three sons were used, since the amount of information added to the uncorrected *lod* score (z_1) for each son is dependent on sibship size and, in fact, increases as size increases (see Appendix 1). Three-son sibships may, therefore, be regarded as providing an average amount of information for each son when the results from these simulated populations are compared to those from actual population surveys.

Table 2 shows the results of these experiments for values of θ_1 of 0.10, 0.20, 0.25 (and 0.75), 0.30, 0.40, and 0.50. For each size of total, the number of computer runs is set out, together with the maximum values of $\hat{\theta}$ at which $Z(\hat{\theta}) > +3$ and also the maximum values of θ at which $Z(\theta) < -2$ when $\hat{\theta} = 0.5$. As explained above, these values can give a rough yardstick, without any presumption about levels of statistical significance, of the order of size of samples at which it can be hoped to indicate or exclude linkage at various values. These sample sizes can in fact also be derived algebraically (Appendix 2).

In addition, the range of the estimates of the recombination fraction obtained with each size of trial is also set out in Table 2, and in Table 3 the complete distributions of the runs obtained with an input value (θ_1) of 0.50 are presented.

A study of Tables 2 and 3 confirms the weaknesses of the two-generation method in both detection and estimation of linkage. Thus, although in no case when $\theta_1 = 0.5$ has a false linkage been demonstrated at the level of $Z(\hat{\theta}) > +3$, in some cases, especially in the trials with smaller numbers of sibships, extreme results near this level have been obtained (Table 3). It is striking that an unequivocal indication of the absence of measurable linkage ($\hat{\theta} = 0.50$) was obtained in only just over half of all trials with an input value (θ_1) of 0.50, including every sibship size (79 out of 151). Even the very large trials involving 1,000 sibships are inadequate to demonstrate loose linkage convincingly when $\theta_1 = 0.40$, or to exclude loose linkage when $\theta_1 = 0.5$.

Some actual surveys concerning linkage between X-chromosomal loci are of a size

TABLE 2

RESULTS OF SECOND SERIES OF SIMULATION EXPERIMENTS

Number of three-son sibships...	10	20	40	100	200	400	1000	2000	5000
Total number of runs.....	60	40	15	11	9	7	7	1	1
Maximum value of $\hat{\theta}$ at which $Z(\hat{\theta})>+3$.....	0.04	0.13	0.19	0.26	0.30	0.32	0.32	0.33	0.40
Maximum value of $\hat{\theta}$ at which $Z(\theta)<-2$ when $\hat{\theta}=0.50$.....	0.20	0.25	0.26	0.34	0.35	0.37	0.40	0.41	0.42
Range of $\hat{\theta}$ when $\theta_1=0.10$.....	0.01–0.31	0.03–0.24	0.06–0.13	0.08–0.13	0.08–0.12	0.09–0.12	0.10–0.12	0.10	0.10
Range of $\hat{\theta}$ when $\theta_1=0.20$.....	0.07–0.50	0.11–0.50	0.12–0.34	0.13–0.23	0.17–0.25	0.18–0.22	0.19–0.21	0.20	0.20
Range of $\hat{\theta}$ when $\theta_1=0.25$.....	0.03–0.50	0.11–0.50	0.20–0.27	0.20–0.31	0.18–0.26	0.23–0.26	0.23–0.26	0.25	0.24
Range of $\hat{\theta}$ when $\theta_1=0.75$.....	0.08–0.50	0.09–0.50	0.19–0.36	0.23–0.30	0.24–0.26	0.22–0.27	0.23–0.25	0.25	0.24
Range of $\hat{\theta}$ when $\theta_1=0.30$.....	0.11–0.50	0.20–0.50	0.25–0.50	0.23–0.37	0.25–0.35	0.31–0.37	0.28–0.32	0.33	0.30
Range of $\hat{\theta}$ when $\theta_1=0.40$.....	0.11–0.50	0.21–0.50	0.28–0.50	0.29–0.50	0.35–0.50	0.36–0.50	0.39–0.43	0.40	0.40
Range of $\hat{\theta}$ when $\theta_1=0.50$.....	0.08–0.50	0.23–0.50	0.31–0.50	0.33–0.50	0.38–0.50	0.41–0.50	0.43–0.50	0.50	0.50

θ_1 is the input value of the recombination fraction θ.
$\hat{\theta}$ is the output estimate of the recombination fraction θ.
$Z(\theta)$ is the *lod* score.

TABLE 3

SIMULATION OF SIBSHIPS OF THREE SONS WHEN $\theta_1 = 0.50$

$\hat{\theta}$	Number of Sibships								
	10	20	40	100	200	400	1000	2000	5000
0.50	33	18	11	5	4	3	3	1	1
0.49									
0.48									
0.47							1		
0.46						1			
0.45				1			1		
0.44						1	1		
0.43	2	1			1	1	1		
0.42		3		1					
0.41					1	1			
0.40	5	2							—
0.39		2	2	2	1				
0.38	4	1		1	2				
0.37		1							
0.36	2	1							
0.35		1							
0.34		3							
0.33		1	1	1				—	
0.32		1					—		
0.31			1						
0.30	4	1			—				
0.29		1							
0.28	3								
0.27									
0.26		1		—					
0.25		1							
0.24	2								
0.23	2	1							
0.22									
0.21									
0.20									
0.19			—						
0.18	1								
0.17									
0.16									
0.15									
0.14									
0.13		—							
0.12									
0.11									
0.10									
0.09									
0.08	2								
0.07									
0.06									
0.05									
0.04	—								
0.03									
0.02									
0.01									
Total number of runs	60	40	15	11	9	7	7	1	1

The horizontal lines in each column represent the maximum values of $\hat{\theta}$ for each size of run, for which $Z(\hat{\theta}) > +3$ in all the trials summarized in Table 2, including every input value (θ_1).

comparable only to a trial involving 10 sibships,* and Table 2 shows that these small trials are inadequate to detect linkage convincingly even when $\theta_1 = 0.10$, since the range of estimates at this value of θ_1 is very large (0.01–0.31). Even 20 sibships are barely adequate when $\theta_1 = 0.10$, although 40 sibships are adequate. When $\theta_1 = 0.25$, trials involving 100 sibships give reasonable success in the detection of linkage and accuracy in its estimation, while trials involving 200 sibships (a size rarely, if ever, attainable in actual surveys) are entirely adequate. At $\theta_1 = 0.30$, trials involving 200 sibships are only marginally adequate. Taken as a whole, the results of these simulated experiments substantiate the uncertainty of Sanger (1965) and Siniscalco *et al.* (1966) whether *lod* scores are effective in the detection and estimation of linkage in two-generation families when linkage is not close. In fact, Morton (1955) stated, when

TABLE 4

COMPARISON BETWEEN THE TWO- AND THREE-GENERATION METHODS IN THE AVERAGE NUMBER OF SIBSHIPS OF THREE SONS NECESSARY FOR VARIOUS VALUES OF θ (*a*) TO OBTAIN A SCORE $Z(\hat{\theta}) > +3$ WHEN $\theta = \hat{\theta}$ (n_1, n_2) AND (*b*) SO THAT $Z(\theta) < -2$ WHEN $\hat{\theta} = 0.50$ (n_3, n_4)

θ	(*a*)		(*b*)	
	Two Generation (n_1)	Three Generation (n_2)	Two Generation (n_3)	Three Generation (n_4)
0.00.....	6	4	1	1
0.05.....	9	5	5	2
0.10.....	14	7	10	4
0.15.....	23	9	17	5
0.20.....	42	12	31	7
0.25.....	83	18	60	11
0.30.....	194	28	139	18
0.35.....	590	51	411	33
0.40.....	2,887	115	1,970	76
0.45.....	45,212	460	30,329	306
0.47.....	347,195	1,279	231,989	852
0.49.....	28,054,950	11,513	18,708,050	7,674

introducing the *lod* scores, that their use in the case of loose linkage was limited because of the prohibitively large sample numbers required.

Table 4 shows the number of sibships containing three sons which are needed to demonstrate and to exclude various values of linkage (using the criteria suggested previously) both by the two- and by the three-generation methods of analysis. These calculations are based on the methods of Appendix 2. In Table 5, comparisons are made between the expected maximum values, derived from Table 4, of $\hat{\theta}$ at which $Z(\hat{\theta}) > +3$ and of θ at which $Z(\theta) < -2$ when $\hat{\theta} = 0.50$, and the observed values obtained in the second series of simulated experiments. The discrepancies between the observed values and the corresponding values calculated by the two-generation method

* Most actual surveys, however, do include some three-generation as well as two-generation data. While no attempt has been made here to simulate such mixed ascertainment, this would, of course, have the effect of increasing the amount of information available.

of analysis are small and are confined to the smaller and larger simulated trials. These minor disagreements are due partly to the fact that the calculated values are the most probable values while the observed values are the extreme of a range found in actual trials (which can be particularly wide in the smaller trials), and partly to the fact that in the simulation experiments some values of $\hat{\theta}$ did not occur at all, especially in the large trials which involved very few repetitions.

The comparisons of Tables 4 and 5 confirm the superiority of the three-generation method of analysis. Both in the case of demonstration and that of exclusion of a particular linkage value, much smaller samples are needed for the three-generation method to obtain a result of comparable significance. The information for each son is the same in the three-generation method whatever the size of the sibship, whereas in

TABLE 5

COMPARISON OF VALUES OF θ (*a*) SUCH THAT $Z(\hat{\theta}) > +3$ WHEN $\theta = \hat{\theta}$ AND (*b*) SUCH THAT $Z(\theta) < -2$ WHEN $\hat{\theta} = 0.50$ FOR VARIOUS NUMBERS OF SIBSHIPS OF THREE SONS BETWEEN THE SIMULATED EXPERIMENTS OF TABLE 2 (θ) AND THE CALCULATIONS OF APPENDIX 2 FOR THE TWO- AND THREE-GENERATION METHODS (θ_2 AND θ_3)

NUMBER OF SIBSHIPS	(*a*)			(*b*)		
	θ	θ_2	θ_3	θ	θ_2	θ_3
10	0.04	0.06	0.17	0.20	0.10	0.24
20	0.13	0.13	0.26	0.25	0.16	0.31
40	0.19	0.19	0.33	0.26	0.22	0.36
100	0.26	0.26	0.39	0.34	0.28	0.41
200	0.30	0.30	0.42	0.35	0.31	0.43
400	0.32	0.33	0.44	0.37	0.34	0.45
1000	0.32	0.36	0.46	0.40	0.38	0.47
2000	0.33	0.39	0.47	0.41	0.40	0.48
5000	0.40	0.41	0.48	0.42	0.42	0.48

the two-generation method it increases with the size of sibship (see Appendix 1). The three-generation method, therefore, will be even more superior in the case of sibships of two sons, but its advantages will decline somewhat as the number of sons increases above three. It should be pointed out that, if the actual number of samples which can be collected and tested is not a factor which need be considered, a population will provide far more families for the two- than for the three-generation method, since the maternal grandfather will often be unavailable. However, shortage of serum will restrict the possibilities of Xg testing for the foreseeable future, and it is this rather than the availability of families which is likely to be the limiting factor in linkage studies. Therefore, the fact that smaller numbers are needed for the three-generation method is a very real advantage. Moreover, certain families who cannot be used for the two-generation method can be used for the three-generation method. Thus (*a*) family units of grandfather, mother, and one son can provide information in the first and not in the second, and (*b*) family units in which the mother has only Xg(a+) sons are wasted in the two-generation method, whereas, if the grandfather is Xg(a−), they can be used in the three-generation method.

THE BEARING OF THE SIMULATION RESULTS ON SURVEYS OF RECOMBINATION BETWEEN THE Xg AND G6PD LOCI

The first surveys undertaken to study the recombination fraction between the loci determining the Xg blood group and G6PD deficiency in Israel (Adam *et al.*, 1963; Sanger and Adam, 1964) and Greece (Fraser *et al.*, 1964) by the two-generation method may be regarded as approximately equivalent in size to the simulated trials of 20 sibships (Table 2), and the survey in Sardinia (Siniscalco *et al.*, 1966) is somewhat larger than a trial of 40 sibships. The total information from the three surveys may be regarded as approximately equivalent to a trial of 100 sibships. The estimates of θ derived by the two-generation method from the data of these surveys together with the maximum *lod* scores ($Z[\hat{\theta}]$) are given in Table 6.

Comparison of Table 6 with the results of the simulation experiments when $\theta_1 = 0.50$ of Table 3 shows that individually the results of these three actual surveys are well within the range of values obtained with simulated trials of comparable size. The estimates using the cumulative totals are, however, at the very extreme of the range

TABLE 6

ESTIMATES OF THE LINKAGE VALUES ($\hat{\theta}$) AND THE MAXIMUM *LOD* SCORES ($Z[\hat{\theta}]$) DERIVED BY THE TWO-GENERATION* METHOD OF ANALYSIS FROM THE DATA OF SURVEYS IN ISRAEL, GREECE, AND SARDINIA

Survey	$\hat{\theta}$		$Z(\hat{\theta})$		Number of Informative Children on Which Estimate Is Based	
1. Israel (1st series)[a]	0.28	...	0.7968	...	51	...
2. Greece (1st series)[b]	0.26	...	0.4605	...	40	...
1+2	...	0.27	...	1.2509	...	91
3. Sardinia[c]	0.37	...	0.2142	...	166	...
1+2+3	...	0.33	...	1.1708	...	257
4. Israel (2d series)[d]	0.50	...	0	...	78	...
1+2+3+4	...	0.37	...	0.6448	...	335
5. Greece (2d series)[e]	0.50	...	0	...	16	...
1+2+3+4+5	...	0.37	...	0.5637	...	351
Sardinia	...	0.37	...	0.2142	...	166
Israel	...	0.39	...	0.1801	...	129
Greece	...	0.32	...	0.2226	...	56
All two-generation families	...	0.33	...	0.7899	...	192
All three-generation families	...	0.44	...	0.0104	...	159

[a] Adam *et al.*, 1963; Sanger and Adam, 1964.
[b] Fraser *et al.*, 1964.
[c] Siniscalco *et al.*, 1966.
[d] Adam *et al.*, 1967.
[e] Stamatoyannopoulos *et al.*, 1968.
* The method of conversion of three-generation families to two generations for analysis is discussed in the text.

of values obtained in the trials of comparable size of Table 3. The balance is somewhat redressed if the results of the second Israeli survey (Adam *et al.*, 1967) and the second Greek survey of the companion paper (Stamatoyannopoulos *et al.*, 1968) are analyzed by the two-generation method. Both these surveys gave estimates of 0.50 for the linkage value, and, therefore, the estimates of the cumulative totals have moved upward until the final total based on 351 informative children gives rise to an estimate which is at the lower end of the range for the trials containing 100 and 200 sibships of Table 3, based on an input value (θ_1) of 0.50.

Thus, these two-generation data from Israel, Greece, and Sardinia are certainly not inconsistent, on the basis of the simulated results of Tables 2 and 3, with a true linkage value of 0.50 between the loci determining the Xg blood group and G6PD deficiency. Nevertheless, on the assumption of a true linkage value of this order, these data could be regarded as being somewhat misleading, because of the extreme estimates obtained in the first two surveys in Israel and Greece.

TABLE 7

COMPARISON OF ESTIMATES OF THE LINKAGE VALUES FROM THREE-GENERATION FAMILIES WHEN ANALYZED BY THE TWO- ($\hat{\theta}_2$) AND THREE-GENERATION ($\hat{\theta}_3$) METHODS

POPULATION	$\hat{\theta}_2$	$\hat{\theta}_3$	$Z(\hat{\theta}_2)$	$Z(\hat{\theta}_3)$	NUMBER OF INFORMATIVE CHILDREN ON WHICH ESTIMATE IS BASED	
					$\hat{\theta}_2$	$\hat{\theta}_3$
Sardinian	0.32	0.47	0.4592	0.0974	94	121
Israeli	0.50	0.45	0	0.0990	43	55
Greek	0.50	0.39	0	0.3708	22	38
All	0.44	0.45	0.0104	0.3840	159	214

Table 7 presents estimates of the recombination fraction derived exclusively from families in Sardinia, Israel, and Greece for whom information was available from three generations. The estimates and the maximum *lod* scores obtained by the three-generation and two-generation methods of analysis are compared. It will be noted that use of the two-generation method involves the loss of 26% of the informative grandchildren. This is because in Tables 6 and 7 conversion of a three-generation pedigree to two generations for analysis by ignoring information from the grandfather involved exclusion of families which could not have been ascertained under a pure two-generation method of ascertainment (for example, families with a single grandchild and families in which all grandchildren were Xg[a+]). This procedure may be responsible for the slight discrepancies between the estimates of Tables 6 and 7 and the corresponding estimates of the papers of Siniscalco *et al.* (1966) and Adam *et al.* (1967).

Table 7 shows that $Z(\hat{\theta}_3)$ is many times greater than $Z(\hat{\theta}_2)$, both scores being derived from the same set of families. In fact, when all the information from the three-

generation families (Table 7) is pooled with the results of two-generation analysis in families where the grandfather is not available (Table 6), the final estimate of θ is 0.43. This is very near the estimate derived from the former group of families alone (0.45) and is influenced very little by the lower estimate from the much larger latter group. The maximum *lod* score of this final estimate, using all available data, is 0.6027.

Table 7 also shows that discrepancies between estimates of θ obtained by two-generation ($\hat{\theta}_2$) and three-generation ($\hat{\theta}_3$) analysis from the same data may be in either direction. Thus in the case of the Israeli data, the two estimates are very similar; in the case of the Sardinian data, $\hat{\theta}_3$ is considerably larger;* and in the case of the Greek data, it is $\hat{\theta}_2$, which is larger. In the case of all the families taken together, the estimates are virtually identical, since the two discrepancies are in opposite directions.

No very close correspondence between $\hat{\theta}_2$ and $\hat{\theta}_3$ is, in fact, to be expected, since the procedure of estimation is very different in the two cases. Thus, to take an extreme example, if all the sibships of the third generation contained only recombinants or only nonrecombinants, $\hat{\theta}_2$ would be zero, whereas $\hat{\theta}_3$ could take any value. This point was readily confirmed by further simulation experiments. Thus, in 25 runs involving 40 three-son sibships each, with an input value (θ_1) of 0.50, analyses were made both by the two-generation and the three-generation methods. In one such run, $\hat{\theta}_3$ was 0.52 while $\hat{\theta}_2$ was only 0.28; conversely, in another run, $\hat{\theta}_2$ was 0.50 while $\hat{\theta}_3$ was 0.41.

CONCLUSION

Table 4 shows that an average of 115 sibships of three sons are needed to obtain an expected score of $Z(\hat{\theta})$ which is greater than $+3$ in the three-generation method when $\theta_1 = 0.40$, while an average of 76 such sibships are needed so that the expected score of $Z(\theta = 0.40)$ should be less than -2 when $\hat{\theta}$ is 0.5. The corresponding values for the two-generation method are 2,887 and 1,970 sibships. Samples of the size needed by the three-generation method to establish or exclude linkage of the order of 0.40 are, therefore, only just within the limits of practicability; the samples required at this level by the two-generation method are out of the question. Thus, the very considerable work performed to obtain three-generation data bearing on the linkage between the loci for G6PD and Xg has resulted in 214 informative grandchildren, the equivalent of 71 three-son sibships.

For linkage values between 0.40 and 0.50, however, even the three-generation method is inadequate. Therefore, the estimate of linkage between the loci for G6PD and Xg derived by the three-generation method from these surveys in Israel, Greece, and Sardinia is 0.45. To demonstrate linkage convincingly at this level, an average of 460 sibships containing three sons are needed, while to exclude linkage of 0.45, if the true level is 0.50, an average of 306 such sibships are required (Table 4). These numbers are perhaps based on somewhat stringent criteria for the demonstration or exclusion of loose linkage of this order; nevertheless, the numbers required will be

* This discrepancy was noted as an anomaly by Siniscalco *et al.* (1966) and as a "rather disturbing and unexplained effect of *lod*-score analysis" in the paper of Adam *et al.* (1967).

impracticably large even with relaxed criteria. As Adam *et al.* (1967) state, the best hope of establishing the linkage relationship of the G6PD and Xg loci is through a study of the relationships of each with an intermediate locus.

SUMMARY

With the help of computer simulation techniques, the inadequacies of family data consisting only of two generations are explored with respect to the detection and estimation of loose degrees of linkage on the X chromosome of man. A comparison is made with the method of estimation involving family data from three generations, which is considerably more efficient.

It is suggested that the results of initial surveys of the relationship on the X chromosome of the loci determining the Xg blood group and G6PD deficiency, analyzed mainly by the two-generation method, led to the misleading suggestion of measurable linkage because of inadequacies of the method. It has been shown by simulation techniques that even these initial surveys, when analyzed by the two-generation method, did not give results which are inconsistent with a true recombination fraction of 0.5, although they represented somewhat extreme values of the range expected from samples of that size from populations in which the true recombination fraction is of the order of 0.5.

The present estimate of the linkage value between these two loci, using all available three-generation families from Sardinia, Israel, and Greece, is 0.45. Despite a much lower estimate from two-generation families (0.33), the combined estimate at 0.43 is very near that derived from the three-generation families alone. The present data, therefore, are consistent with either loose linkage or with free recombination. The best hope of deciding between these alternatives lies in studying the linkage relationships of each of these loci with an intermediately situated one.

APPENDIX 1

THE VARIATION IN THE CONTRIBUTION OF EACH INFORMATIVE SON TO THE *LOD* SCORE (z_1) WITH THE SIZE OF SIBSHIP

It will be shown here that the amount of information which can be added to the uncorrected score (z_1) is greater for the same recombination fraction in a sibship of size ks than in k sibships of size s (k and s are both integers greater than 1). Thus, the score for each son will be shown to be greater in the first case.

In the first case let there be ka recombinants and kb nonrecombinants, and in the second case let the figures be respectively a and b for each sibship. Then $a + b = s$, and the recombination fraction in each case is a/s.

In the first case

$$z_1(\theta) = \log 2^{ks-1}[\theta^{ka}(1 - \theta)^{kb} + \theta^{kb}(1 - \theta)^{ka}] .$$

In the second case the sum of the scores for the k sibships is

$$\log 2^{k(s-1)}[\theta^{a}(1 - \theta)^{b} + \theta^{b}(1 - \theta)^{a}]^{k} .$$

Taking antilogs and dividing through by 2^{ks-1}, it remains to be shown that

$$\theta^{ka}(1-\theta)^{kb} + \theta^{kb}(1-\theta)^{ka} > \frac{[\theta^a(1-\theta)^b + \theta^b(1-\theta)^a]^k}{2^{k-1}},$$

or that

$$2^{k-1}\left[\theta^{ks}\left(\frac{1-\theta}{\theta}\right)^{kb} + \theta^{ks}\left(\frac{1-\theta}{\theta}\right)^{ka}\right] > \left[\theta^s\left(\frac{1-\theta}{\theta}\right)^b + \theta^s\left(\frac{1-\theta}{\theta}\right)^a\right]^k,$$

or that

$$2^{k-1}(x^k + y^k) > (x+v)^k,$$

where

$$x = \left(\frac{1-\theta}{\theta}\right)^b \quad \text{and} \quad v = \left(\frac{1-\theta}{\theta}\right)^a.$$

This is a special case of an inequality due to Chebychev which states that

$$\sum_1^n a_i/n \cdot \sum_1^n b_i/n < \sum_1^n a_i b_i/n.$$

From this it follows that

$$\sum_1^n c_i/n \cdot \sum_1^n a_i b_i/n < \sum_1^n a_i b_i c_i/n,$$

and by repeated applications that

$$\frac{\Sigma a}{n} \cdot \frac{\Sigma b}{n} \cdots \frac{\Sigma k}{n} < \frac{\Sigma ab \cdots k}{n}.$$

Let

$$\frac{\Sigma a}{n} = \frac{\Sigma b}{n} = \cdots = \frac{\Sigma k}{n} = \frac{x+y}{2}.$$

Then

$$\left(\frac{x+y}{2}\right)^k < \frac{x^k + y^k}{2},$$

or

$$2^{k-1}(x^k + y^k) > (x+y)^k.$$

This inequality holds if x and y are positive and not equal and these conditions are fulfilled by

$$\left(\frac{1-\theta}{\theta}\right)^b \quad \text{and} \quad \left(\frac{1-\theta}{\theta}\right)^a,$$

unless $a = b$ or $\theta = 0.5$. In these special cases, the inequality is replaced by an equality.

APPENDIX 2

COMPARISON BETWEEN THE TWO- AND THREE-GENERATION METHODS IN THE NUMBERS OF SIBSHIPS CONTAINING THREE SONS REQUIRED TO DEMONSTRATE AND TO EXCLUDE VARIOUS VALUES OF LINKAGE

In the analysis of the second series of simulation experiments concerning the two-generation method, use was made of the maximum value of $\hat{\theta}$, such that $Z(\hat{\theta}) > +3$, and of the

maximum value of ℓ, such that $Z(\theta) < -2$ when $\hat{\theta} = 0.5$. It was pointed out that these values gave some indications, useful for comparative purposes even if not at any formal significance levels, of the size of samples of sibships containing three sons needed to indicate the presence or absence of linkage at various values.

These figures may be derived algebraically for both the two- and three-generation methods of analysis and may be used to compare the efficiency of the two methods. First, a tabulation can be made of the types of sibships containing three sons which are ascertainable with their relative frequencies in terms of the recombination fraction θ and with their scores.

Sibship	Score	Relative Frequency
gt gt gt.....	$z_1(3:0)+e_1(3:0)$	$\theta^3+(1-\theta)^3=1-3\theta(1-\theta)$
GT GT gt.....	$z_1(3:0)+e_1(2:1)$	$3[1-3\theta(1-\theta)]$
Gt Gt gT.....	$z_1(3:0)+e_1(2:1)$	$3[1-3\theta(1-\theta)]$
gT gT Gt.....	$z_1(3:0)+e_1(2:1)$	$3[1-3\theta(1-\theta)]$
gt gt GT.....	$z_1(3:0)+e_1(2:1)$	$3[1-3\theta(1-\theta)]$
gt gt gT.....	$z_1(2:1)+e_1(3:0)$	$3\theta(1-\theta)^2+3\theta^2(1-\theta)=3\theta(1-\theta)$
gT gT gt.....	$z_1(2:1)+e_1(3:0)$	$3\theta(1-\theta)$
Gt Gt gt.....	$z_1(2:1)+e_1(2:1)$	$3\theta(1-\theta)$
gt gt Gt.....	$z_1(2:1)+e_1(2:1)$	$3\theta(1-\theta)$
GT Gt gT.....	$z_1(2:1)+e_1(2:1)$	$6\theta(1-\theta)$
GT Gt gt.....	$z_1(2:1)+e_1(2:1)$	$6\theta(1-\theta)$
GT gT gt.....	$z_1(2:1)+e_1(2:1)$	$6\theta(1-\theta)$
Gt gT gt.....	$z_1(2:1)+e_1(2:1)$	$6\theta(1-\theta)$

If now $\theta(1-\theta)$ is replaced by x, $z_1(3:0)$ by y_1, $z_1(2:1)$ by y_2, $e_1(3:0)$ by y_3, and $e_1(2:1)$ by y_4, it can be seen that $13 - 3x$ sibships would be composed of $1 - 3x$ with a score $y_1 + y_3$, $12 - 36x$ with a score $y_1 + y_4$, $6x$ with a score $y_2 + y_3$, and $30x$ with a score of $y_2 + y_4$.

Thus these $13 - 3x$ sibships would have an expected score of $13(1 - 3x)y_1 + 36xy_2 + (1 + 3x)y_3 + (12 - 6x)y_4$ when $\hat{\theta} = \ell$. From this the number of sibships (n_1) needed to give an expected score of 3 can readily be calculated as

$$\frac{3(13 - 3x)}{13(1 - 3x)y_1 + 36xy_2 + (1 + 3x)y_3 + (12 - 6x)y_4},$$

and n can be tabulated by computer for various values of θ.

A similar tabulation can be made to obtain the number of sibships n_3 which will give a score of -2 when $\hat{\theta} = 0.5$. In fact, n_3 will be equal to the above expression replacing 3 by -2 and x by $\frac{1}{4}$. Thus,

$$n_3 = \frac{-24.5}{3.25y_1 + 9y_2 + 1.75y_3 + 10.5y_4}.$$

For purposes of comparison, an expected score can be calculated for the three-generation method in which the phase of the alleles at the G and T loci in the heterozygous mother is known. If there are k sibships, this score will be

$$3k\theta \log 2\theta + 3k(1 - \theta) \log 2(1 - \theta),$$

or

$$3k\{\log 2(1 - \theta) + \theta[\log 2\theta - \log 2(1 - \theta)]\},$$

and the number of sibships n_2 needed to give an expected score of 3 will be

$$\frac{1}{\log 2(1 - \theta) + \theta \log [\theta/(1 - \theta)]}.$$

Similarly, n_4 (the number of sibships needed to give an expected score of -2 at a value of θ when $\hat{\theta} = 0.5$) is such that

$$\frac{3n_4}{2} \log 2(1 - \theta) + \frac{3n_4}{2} \log 2\theta = -2 ,$$

or

$$n_4 = \frac{-4}{3 \log 4\theta(1 - \theta)} .$$

REFERENCES

ADAM, A., SHEBA, C., SANGER, R., RACE, R. R., TIPPETT, P., HAMPER, J., GAVIN, J., and FINNEY, D. J. 1963. Data for X-mapping calculations. Israeli families tested for Xg, G6PD and for colour vision. *Ann. Hum. Genet.* (Lond.) **26**:187–194.

ADAM, A., TIPPETT, P., GAVIN, J., NOADES, J., SANGER, R., and RACE, R. R. 1967. The linkage relation of Xg to g-6-pd in Israelis: the evidence of a second series of families. *Ann. Hum. Genet.* (Lond.) **30**:211–218.

FRASER, G. R. 1968. Some complications of the use of the three-generation method in the estimation of linkage relationships on the X chromosome in man. *Ann. Hum. Genet.* (Lond.) **32**:65–79.

FRASER, G. R., DEFARANAS, B., KATTAMIS, C. A., RACE, R. R., SANGER, R., and STAMATOYANNOPOULOS, G. 1964. Glucose-6-phosphate dehydrogenase, colour vision and Xg blood groups in Greece: linkage and population data. *Ann. Hum. Genet.* (Lond.) **27**:395–403.

MAYO, O. 1967. The use of simulation by computer in population genetics. Ph. D. thesis. University of Adelaide.

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

MORTON, N. E. 1957. Further scoring types in sequential linkage tests with a critical review of autosomal and partial sex linkage in man. *Amer. J. Hum. Genet.* **9**:55–75.

RENWICK, J. H., and SCHULZE, J. 1964. An analysis of some data on the linkage between Xg and colourblindness in man. *Amer. J. Hum. Genet.* **16**:410–418.

SANGER, R. 1965. Genes on the X chromosome. *Canad. J. Genet. Cytol.* **7**:179–188.

SANGER, R., and ADAM, A. 1964. Xg and g-6-pd in Israeli families: an addendum. *Ann. Hum. Genet.* (Lond.) **27**:271–272.

SINISCALCO, M., FILIPPI, G., LATTE, B., PIOMELLI, S., RATTAZZI, M., GAVIN, J., SANGER, R., and RACE, R. R. 1966. Failure to detect linkage between Xg and other X-borne loci in Sardinians. *Ann. Hum. Genet.* (Lond.) **29**:231–252.

STAMATOYANNOPOULOS, G., SOFRONIADOU, C., AKRIVAKIS, A., and FRASER, G. R. 1968. Further data from Greece on recombination between the Xg blood group and glucose-6-phosphate dehydrogenase deficiency. *Amer. J. Hum. Genet.* **20**: 528–533.