Complex Segregation Analysis. II. Multiple Classification

REGINA C. ELANDT-JOHNSON[1]

INTRODUCTION

Modes of inheritance of human diseases are usually studied by analyzing family and pedigree data. A child having a disease under consideration is called "affected," otherwise it is called "normal." It appears that many rare diseases are inherited as simple recessives or as simple dominants. For *rare* diseases, the carriers of deleterious genes are usually heterozygotes. It follows that each of three basic parental phenotypic matings (Normal × Normal, Normal × Affected, and Affected × Affected) corresponds to genotypic matings in which one or both parents are heterozygotes.

Let θ be the probability that a child born in a family with a given parental genotypic mating type will be affected. Thus, the probability that in such a family of size s there will be r affected children is

$$\Pr(r) = \binom{s}{r} \theta^r (1 - \theta)^{n-r}, \qquad r = 0, 1, \ldots, s. \quad (1)$$

The probability in (1), considered as a function of the random variable r, is called the *segregation distribution*, and θ is called the *segregation parameter*. Statistical methods developed for estimating and testing hypotheses about θ are known as segregation analysis.

Since computer facilities are available, *common* traits (not necessarily diseases) such as blood groups, for instance, can also be analyzed, using family data. For common traits, there may be more than one segregation parameter for a given segregation distribution and more than one segregation distribution within a given phenotypic mating type. Statistical methods in which two or more distinct and functionally independent segregation parameters are involved in the analysis of family data will be called *complex segregation analysis.*

Complex segregation analysis, in which *two* phenotypic classes (affected and normal) were distinguished, has been presented in my previous paper [1]. In the present paper, the method will be generalized to situations in which *more than two* phenotypic classes are considered. Family data on multiallelic inheritance and multiloci inheritance with or without linkage are examples. These models can be complicated by incomplete penetrance, different fitnesses of genotypes, inbreeding, etc. General models for complex segregation analysis will be derived for autosomal and X-linked inheritance, using various sampling techniques.

Received May 11, 1970.

This work has been supported by U.S. Public Health Service research grant AI 07975 from the National Institutes of Health.

[1] Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27514.

AUTOSOMAL INHERITANCE. A RANDOM SAMPLE OF FAMILIES

In this section, consider models associated with *autosomal* inheritance so that the numbers of possible genotypes in females and males are the same. First, only experiments in which a *random sample of families* is selected will be considered. Two situations will be distinguished: (1) information on the phenotypes of parents is not available; and (2) information is available.

No Information Available About Parental Phenotypes

First consider families of fixed size s so that the phrase "families of size s" will be often omitted.

1. Consider a certain character which is supposed to be controlled by a certain genetic system and which gives rise to K phenotypic forms. Let k ($k < K$) be the number of phenotypic forms to be observed, and so classify the individuals into k mutually exclusive categories, $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$. Call this the "$\mathcal{C}$ classification." For instance, in the ABO–blood group system, six phenotypic classes (O, A_1, A_2, B, A_1B, A_2B), or four classes (O, A, B, AB), or only two classes (A and non-A), etc., may be distinguished.

2. Let f be the number of possible genotypes associated with this mode of inheritance. It is sometimes convenient to distinguish between the genotypes of the mother and of the father (e.g., certain maternal [or paternal] genotypes may influence segregation ratios). Let ψ_{uv} be the expected proportion of genotypic matings $u \times v$. For the entire population,

$$\sum_{u=1}^{f} \sum_{v=1}^{f} \psi_{uv} = 1 . \tag{2}$$

Assuming that the female and male genotype frequencies are equal and the ith genotype has the frequency ψ_i, with

$$\sum_{i=1}^{f} \psi_i = 1 ,$$

then for *random* mating

$$\psi_{uv} = \psi_u \psi_v , \qquad u, v = 1, 2, \ldots, f . \tag{3}$$

3. Let $\theta_{l(uv)}$ ($l = 1, 2, \ldots, k$) be the probability that a child born in a family with parental genotypes $u \times v$ is classified into the $\mathcal{C}_l$ category. Then

$$\sum_{l=1}^{k} \theta_{l(uv)} = 1 . \tag{4}$$

Let r_l ($l = 1, 2, \ldots, k$) be the observed number of children within the category $\mathcal{C}_l$, in a family of size s. Then

$$\sum_{l=1}^{k} r_l = s . \tag{5}$$

Thus, for a family of size s, *given* that the parental genotypes are $u \times v$, there is the following distribution of the children:

Category	$\mathcal{C}_1$	$\mathcal{C}_2$	. . .	$\mathcal{C}_k$	Total
Probability	$\theta_{1(uv)}$	$\theta_{2(uv)}$	. . .	$\theta_{k(uv)}$	1
No. of children	r_1	r_2	. . .	r_k	s.

The quantities $r_1, r_2, \ldots, r_k$ considered as random variables have the (conditional) multinomial distribution

$$P_{s(uv)}(r_1\ r_2, \ldots, r_k) = \frac{s!}{r_1!r_2! \ldots r_k!} \theta^{r_1}_{1(uv)}\theta^{r_2}_{2(uv)} \ldots \theta^{r_k}_{k(uv)} . \tag{6}$$

[Note that, in fact, $P_{s(uv)}$ $(r_1, r_2, \ldots, r_k)$ is the multinomial probability

$$\Pr(r_1, r_2, \ldots, r_k; s \,|\, \text{mating } u \times v)].$$

By analogy to (1), call (6) a (multinomial) *segregation distribution*, and the parameters $\theta_{l(uv)}$ $(l = 1, 2, \ldots, k)$ will be called *segregation parameters.*

4. Let $P_s(r_1, r_2, \ldots, r_k)$ be the probability that a family of size s selected at random from the entire population will have r_1 children in the category $\mathcal{C}_1$, r_2 children in the category $\mathcal{C}_2$, . . . , and r_k children in the category $\mathcal{C}_k$, with

$$\sum_{=1}^{k} r_l = s .$$

Thus,

$$P_s(r_1, r_2, \ldots, r_k) = \sum_{u=1}^{f} \sum_{v=1}^{f} \psi_{uv} P_{s(uv)}(r_1, r_2, \ldots, r_k) . \tag{7}$$

[Note that some terms in the right hand side of eq. (7) may be zeros.]

5. The probabilities ψ_{uv} can be functions of some genetic parameters such as gene frequencies, inbreeding coefficient, fitnesses of genotypes, etc. Some of these parameters might be functionally dependent (e.g., gene frequencies add up to one). Similarly, $\theta_{l(uv)}$ can be functions of other parameters such as penetrance, fitnesses, and linkage. Again, some of these parameters may be functionally dependent. In most cases, it would not be difficult to find the total number, M, of parameters which are distinct and *functionally independent.* Let us denote these parameters by $\gamma' = (\gamma_1, \gamma_2, \ldots, \gamma_M)$. Thus the probability P_s in (7) is a certain function of M functionally independent parameters γ_i's and can be written more precisely:

$$P_s(r_1, r_2, \ldots, r_k) = P_s(r_1, r_2, \ldots, r_k; \gamma) . \tag{8}$$

6. In evaluating the probability P_s from (7), it has been allowed, in the general case, for f^2 different matings and f^2 segregation distributions defined in (6). In fact, there may not be so many.

First, maternal and paternal genotypes are not usually distinguished so that the matings $u \times v$ and $v \times u$ can be considered as the same. Thus, there will be only $f(f + 1)/2$ different genotypic matings.

Second, different genotypic matings may have the same segregation distributions. For instance, in the ABO–blood group system, matings $AA \times BO$ will produce only

offspring of blood groups A and AB with equal probabilities, and the same applies to matings $AA \times AB$ (see table 1). In such cases, it can be said that these genotypic matings have the same *segregation pattern* $\mathcal{S}_t$, say, determined by the parameters of the common segregation distribution.

Suppose that for a given model there are m distinct segregation patterns. Let ϕ_t be the expected proportion (probability) of matings with the $\mathcal{S}_t$ segregation pattern in a population. Then,

$$\sum_{t=1}^{m} \phi_t = 1 . \tag{9}$$

Let

$$\theta_{1t}, \theta_{2t}, \ldots, \theta_{kt} , \quad \text{with} \quad \sum_{t=1}^{m} \theta_{lt} = 1 ,$$

be the segregation probabilities. (Note that some θ_{lt}'s may be equal to zero.) Further, let

$$P_{st}(r_1, r_2, \ldots, r_k) = \frac{s!}{r_1! r_2! \ldots r_k!} \theta_{1t}^{r_1}, \theta_{2t}^{r_2}, \ldots, \theta_{kt}^{r_k} \tag{10}$$

be the segregation distribution for the $\mathcal{S}_t$ pattern. Thus, P_s defined in (7) takes the form

$$P_s(r_1, r_2, \ldots, r_k) = \sum_{t=1}^{m} \phi_t P_{st}(r_1, r_2, \ldots, r_k) . \tag{11}$$

[Note that some terms in the right-hand side of eq. (11) may be zeros.]

7. Let n_s be the number of families, each of size s, randomly selected from a population; and $r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}$, with

$$\sum_{l=1}^{k} r_{l\alpha} = s ,$$

be the observed numbers of children in categories $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$, respectively, in the αth family ($\alpha = 1, 2, \ldots, n_s$). The likelihood function for the sample of n_s families, $L_s(\gamma)$, can be written as

$$L_s(\gamma) = \prod_{\alpha=1}^{n_s} P_s(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) = \prod_{\alpha=1}^{n_s} \left[\sum_{t=1}^{m} \phi_t P_{st}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) \right] . \tag{12}$$

8. However, since there could be more than two distinct categories in classification $\mathcal{C}$, it will be more convenient to arrange the family data, not with respect to segregation patterns, but with respect to family patterns.

A *family pattern*, $\mathcal{F}_\beta$, say, will designate a subset of distinct classes from the classification $\mathcal{C}$ present in the offspring.

For instance, for the ABO blood groups, the families with only O-group children represent a pattern; with only A-group children, another pattern; with O- and A-group children, another pattern; O, A and B groups, another pattern, etc. In general, if there are k categories in classification $\mathcal{C}$, the number of distinct family patterns is $2^k - 1$.

Notice that the family pattern is a certain function of the numbers $r_1, r_2, \ldots, r_k$. For instance, for the pattern with only O-group children, $r_1 = s$ (i.e., $r_2 = r_3 = r_4 = 0$), while for the pattern with O-, A- and B-group children, $r_1 + r_2 + r_3 = s$, $r_i \neq 0$ for $i = 1, 2, 3$, and $r_4 = 0$. To emphasize this, write

$$\mathfrak{F}_\beta = \mathfrak{F}_\beta(r_1, r_2, \ldots, r_k) \tag{13}$$

for the βth family pattern.

The probability that a family with s children will have a pattern $\mathfrak{F}_\beta$ with $r_1, r_2, \ldots, r_k$ children in categories $\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_k$, respectively, is

$$P_s(\mathfrak{F}_\beta; \gamma) = \sum_{l \in \mathfrak{F}_\beta} \phi_l P_{sl}(r_1, r_2, \ldots, r_k) . \tag{14}$$

Let $F_{\alpha(\beta)}$ denote the αth family ($\alpha = 1, 2, \ldots, n_s$), and $r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}$ denote the number of *observed* children in categories $\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_k$, respectively. The subscript (β) indicates that this family has a particular pattern $\mathfrak{F}_\beta$, that is,

$$F_{\alpha(\beta)} = \mathfrak{F}_\beta(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) . \tag{15}$$

Let $P_s(F_{\alpha(\beta)}; \gamma)$ be the probability that the αth family (with the pattern $\mathfrak{F}_\beta$) has the observed frequencies $r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}$. Thus,

$$P_s(F_{\alpha(\beta)}; \gamma) = \sum_{l \in \mathfrak{F}_\beta} \phi_l P_{sl}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) . \tag{16}$$

The likelihood function, $L_s(\gamma)$ defined in (12) can now be written in the form

$$L_s(\gamma) = \prod_{\alpha=1}^{n_s} P_s(F_{\alpha(\beta)}; \gamma) = \prod_{\alpha=1}^{n_s} \left[\sum_{l \in \mathfrak{F}_\beta} \phi_l P_{sl}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) \right] . \tag{17}$$

9. In setting up the likelihood equations, the coefficients $(s!/r_1!r_2! \ldots r_k!)$ in (10) will have no role. For this reason, it is convenient to represent (10) in a slightly different form.

Put

$$c_s(r_1, r_2, \ldots, r_k) = \frac{s!}{r_1!r_2! \ldots r_k!} , \tag{18}$$

and

$$T_{sl}(r_1, r_2, \ldots, r_k) = \theta_{1l}^{r_1}\theta_{2l}^{r_2} \ldots \theta_{kl}^{r_k} . \tag{19}$$

The probability $P_s(\mathfrak{F}_\beta; \gamma)$ defined in (14) can be written in the form

$$P_s(\mathfrak{F}_\beta; \gamma) = c_s(r_1, r_2, \ldots, r_k) \sum_{l \in \mathfrak{F}_\beta} \phi_l T_{sl}(r_1, r_2, \ldots, r_k) . \tag{20}$$

For the αth family, the probability $P_s(F_{\alpha(\beta)}; \gamma)$ defined in (16) takes the form

$$P_s(F_{\alpha(\beta)}; \gamma) = c_s(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) \sum_{l \in \mathfrak{F}_\beta} \phi_l T_{sl}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) . \tag{21}$$

The likelihood function (17) can now be written in the form

$$L_s(\gamma) = C_s \prod_{\alpha=1}^{n_s} \left[\sum_{t \in \mathfrak{T}_\beta} \phi_t T_{st}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) \right], \tag{22}$$

where

$$C_s = \prod_{\alpha=1}^{n_s} c_s(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) . \tag{23}$$

10. If the total sample consists of families of sizes $s = 1, 2, \ldots, S$, then the overall likelihood function, $L(\gamma)$, is

$$L(\gamma) = \prod_{s=1}^{S} L_s(\gamma) . \tag{24}$$

Parental Phenotypes Are Known

When the parental phenotypic mating type is known, repeat the same procedure for constructing the likelihood function, except that now ϕ_t's (or ψ_{uv}'s) are the *conditional* probabilities of genotypic matings, *given* the phenotypic mating type.

ESTIMATION OF GENE FREQUENCIES IN THE ABO-LIKE SYSTEMS FROM FAMILY DATA

Before proceeding to further theoretical problems, it will be useful to illustrate the theory developed above by a practical example of constructing a segregation model for estimation of gene frequencies from family data. The ABO-blood group system for four phenotypic classes (groups O, A, B, and AB) will be used and a model constructed when there is no information on the phenotypes of the parents.

Let p, q, and w, with $p + q + w = 1$, be the frequencies of alleles A, B, and O, respectively. Assuming that the population is in equilibrium, and taking a random sample of individuals, the ML estimates of gene frequencies can be calculated using standard methods. However, in random sampling we assume that the observations are independent, which, in practice, means that only one offspring in a family should be tested for the blood group. In fact, often whole families are tested, and the data are pooled together. If the total sample is large and the sizes of families are small, the assumption of independent sampling may hold approximately. On the other hand, for small samples and large families, this is not so. The observations within families are correlated, and so they do not constitute a simple random sample. In such cases, methods of complex segregation analysis are needed.

Table 1 represents a model for segregation analysis for the ABO-blood group data under the following assumptions: population in equilibrium, complete penetrance, and no selection. The table is arranged according to segregation patterns.

The gene frequencies, p, q, and w will be estimated. Of course, there will be no difficulty in modifying the table by introducing extra parameters if some of the assumptions do not hold.

When selecting families, the blood groups of the children will be tested, that is, family patterns will be observed. Table 1 must be rearranged according to family patterns. There are $2^4 - 1 = 15$ different family patterns. Table 2 gives some ex-

amples. The remaining family patterns can be reconstructed from table 1 in a similar way.

Example 1

Consider a simple example of $n_s = 3$ families, each of size $s = 4$, with family patterns as shown in table 3.

TABLE 1

SEGREGATION PATTERNS FOR THE ABO-BLOOD GROUP SYSTEM. ESTIMATION OF GENE FREQUENCIES

SEGREGATION PATTERN (S_l)	GENOTYPE MATING	EXPECTED PROPORTION OF MATINGS (ψ_{uv})	SEGREGATION PARAMETERS (θ_{lt}) O	A	B	AB	SEGREGATION DISTRIBUTIONS
1. $(r_1=s)$	$OO\times OO$	w^4	1	0	0	0	1^{r_1}
2. $(r_2=s)$	$AA\times AA$	p^4	0	1	0	0	1^{r_2}
	$AA\times AO$	$4p^3w$	0	1	0	0	
	$OO\times AA$	$2p^2w^2$	0	1	0	0	
3. $(r_3=s)$	$BB\times BB$	q^4	0	0	1	0	1^{r_3}
	$BB\times BO$	$4q^3w$	0	0	1	0	
	$OO\times BB$	$2q^2w^2$	0	0	1	0	
4. $(r_4=s)$	$AA\times BB$	$2p^2q^2$	0	0	0	1	1^{r_4}
5. $(r_1+r_2=s)$	$OO\times AO$	$4pw^3$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$(s!/r_1!r_2!)(\frac{1}{2})^{r_1+r_2}$
6. $(r_1+r_3=s)$	$OO\times BO$	$4qw^3$	$\frac{1}{2}$	0	$\frac{1}{2}$	0	$(s!/r_1!r_3!)(\frac{1}{2})^{r_1+r_3}$
7. $(r_2+r_3=s)$	$OO\times AB$	$4pqw^2$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	$(s!/r_2!r_3!)(\frac{1}{2})^{r_2+r_3}$
8. $(r_2+r_4=s)$	$AA\times BO$	$4p^2qw$	0	$\frac{1}{2}$	0	$\frac{1}{2}$	$(s!/r_2!r_4)(\frac{1}{2})^{r_2+r_4}$
	$AA\times AB$	$4p^3q$	0	$\frac{1}{2}$	0	$\frac{1}{2}$	
9. $(r_3+r_4=s)$	$AO\times BB$	$4pq^2w$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$(s!/r_3!r_4!)(\frac{1}{2})^{r_3+r_4}$
	$BB\times AB$	$4pq^3$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	
10. $(r_1+r_2=s)$	$AO\times AO$	$4p^2w^2$	$\frac{1}{4}$	$\frac{3}{4}$	0	0	$(s!/r_1!r_2!)(\frac{1}{4})^{r_1}(\frac{3}{4})^{r_2}$
11. $(r_1+r_3=s)$	$EO\times EO$	$4q^2w^2$	$\frac{1}{4}$	0	$\frac{3}{4}$	0	$(s!/r_1!r_3!)(\frac{1}{4})^{r_1}(\frac{3}{4})^{r_3}$
12. $(r_2+r_3+r_4=s)$	$AO\times AB$	$8p^2qw$	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$(s!/r_2!r_3!r_4!)(\frac{1}{2})^{r_2}(\frac{1}{4})^{r_3+r_4}$
13. $(r_2+r_3+r_4=s)$	$BO\times AB$	$8pq^2w$	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$(s!/r_2!r_3!r_4!)(\frac{1}{2})^{r_3}(\frac{1}{4})^{r_2+r_4}$
14. $(r_2+r_3+r_4=s)$	$AB\times AB$	$4p^2q^2$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$(s!/r_2!r_3!r_4!)(\frac{1}{2})^{r_4}(\frac{1}{4})^{r_2+r_3}$
15. $(r_1+r_2+r_3+r_4=s)$	$AO\times BO$	$8pqw^2$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$(s!/r_1!r_2!r_3!r_4!)(\frac{1}{4})^{r_1+r_2+r_3+r_4}$
			OBSERVED FREQUENCIES				
			r_1	r_2	r_3	r_4	

TABLE 2

FAMILY PATTERNS FOR COMPLEX SEGREGATION ANALYSIS FOR THE ABO BLOOD SYSTEMS. ESTIMATION OF GENE FREQUENCIES

FAMILY PATTERN $\mathfrak{F}_\beta(r_1, \ldots, r_4)$		GENOTYPE MATING	EXPECTED PROPORTION	SEGREGATION PATTERN	
				S_t	$\theta_{1t}^{r_1} \ldots \theta_{4t}^{r_4}$
		$OO \times OO$	w^4	1	1^{r_1}
	$r_1 = s$	$OO \times AO$	$4pw^3$	5	$(\frac{1}{2})^{r_1}$
$\mathfrak{F}_1$	Group O	$OO \times BO$	$4qw^3$	6	
		$AO \times AO$	$4p^2w^2$	10	$(\frac{1}{4})^{r_1}$
		$BO \times BO$	$4q^2w^2$	11	
		$AO \times BO$	$8pqw^2$	15	
		$AA \times AA$	p^4	2	1^{r_2}
		$AA \times AO$	$4p^3w$	2	
		$OO \times AA$	$2p^2w^2$	2	
	$r_2 = s$	$AO \times AO$	$4p^2w^2$	10	$(\frac{3}{4})^{r_2}$
$\mathfrak{F}_2$	Group A	$OO \times AO$	$4pw^3$	5	$(\frac{1}{2})^{r_2}$
		$OO \times AB$	$4pqw^2$	7	
		$AA \times BO$	$4p^2qw$	8	
		$AA \times AB$	$4p^3q$	8	
		$AO \times AB$	$8p^2qw$	12	
		$BO \times AB$	$8pq^2w$	13	$(\frac{1}{4})^{r_2}$
		$AB \times AB$	$4p^2q^2$	14	
		$AO \times BO$	$8pqw^2$	15	
etc.					
	$r_1 \neq 0, r_2 \neq 0$ $r_1 + r_2 = s$	$OO \times AO$	$4pw^3$	5	$(\frac{1}{2})^{r_1+r_2}$
$\mathfrak{F}_5$	Groups O, A	$AO \times AO$	$4p^2w^2$	10	$(\frac{1}{4})^{r_1}(\frac{3}{4})^{r_2}$
		$AO \times BO$	$8pqw^2$	15	$(\frac{1}{4})^{r_1+r_2}$
etc.					
		$OO \times AB$	$4pqw^2$	7	$(\frac{1}{2})^{r_2+r_3}$
	$r_2 \neq 0, r_3 \neq 0$ $r_2 + r_3 = s$	$AO \times AB$	$8p^2qw$	12	$(\frac{1}{2})^{r_2}(\frac{1}{4})^{r_3}$
$\mathfrak{F}_8$	Groups A, B	$BO \times AB$	$8pq^2w$	13	$(\frac{1}{4})^{r_2}(\frac{1}{2})^{r_3}$
		$AB \times AB$	$4p^2q^2$	14	$(\frac{1}{4})^{r_2+r_3}$
		$AO \times BO$	$8pqw^2$	15	
etc.					
	$r_1 \neq 0, r_2 \neq 0, r_3 \neq 0$				
$\mathfrak{F}_{11}$	$r_1 + r_2 + r_3 = s$ Groups O, A, B	$AO \times BO$	$8pqw^2$	15	$(\frac{1}{4})^{r_1+r_2+r_3}$
etc.					
		$AO \times AB$	$8p^2qw$	12	$(\frac{1}{2})^{r_2}(\frac{1}{4})^{r_3+r_4}$
	$r_2 \neq 0, r_3 \neq 0, r_4 \neq 0$ $r_2 + r_3 + r_4 = s$	$BO \times AB$	$8pq^2w$	13	$(\frac{1}{2})^{r_3}(\frac{1}{4})^{r_2+r_4}$
$\mathfrak{F}_{14}$	Groups A, B, AB	$AB \times AB$	$4p^2q^2$	14	$(\frac{1}{2})^{r_4}(\frac{1}{4})^{r_2+r_3}$
		$AO \times AO$	$8pqw^2$	15	$(\frac{1}{4})^{r_2+r_3+r_4}$

Since $p + q + w = 1$, only p and q need to be estimated. The likelihood function, $L_4(p, q)$, must be constructed for this sample. For example, the first family, F_1, has a pattern of all children with O group (according to table 2, this is pattern $\mathfrak{F}_1$), so $F_{1(1)} = \mathfrak{F}_1(r_1 = s)$. From table 2 the corresponding probabilities are obtained:

$$P_4(F_{1(1)}; p, q) = w^4 + 4w^3(p + q)(\tfrac{1}{2})^4 + 4w^2(p + q)\,(\tfrac{1}{4})^4\,,$$

$$P_4(F_{1(5)}; p, q) = 6 \cdot 4pw^2[w(\tfrac{1}{2})^4 + p(\tfrac{1}{4})^2(\tfrac{3}{4})^2 + 2q(\tfrac{1}{4})^4]\,,$$

$$P_4(F_{1(11)}; p, q) = 12 \cdot 4pq[2pqw(\tfrac{1}{2})(\tfrac{1}{4})^3 + 2qw(\tfrac{1}{2})(\tfrac{1}{4})^3 + pq(\tfrac{1}{2})^2(\tfrac{1}{4})^2 + 2w^2(\tfrac{1}{4})^4]\,,$$

substituting $w = 1 - p - q$.

These equations can be simplified, but this is not essential if one is using a general computer program for likelihood estimates. The likelihood function is

$$L_4(p, q) = P_4(F_{1(1)})P_4(F_{1(5)})P_4(F_{1(11)})\,.$$

TABLE 3

Family $Fr_{\alpha(\beta)}$	Family Pattern $\mathfrak{F}_{\beta(r_1, \ldots, r_4)}$			
	r_1 (O)	r_2 (A)	r_3 (B)	r_4 (AB)
$F_{1(1)}$	4	. . . . .	. . . . .	. . . . .
$F_{2(5)}$	2	2	. . . . .	. . . . .
$F_{3(11)}$	. . . . .	1	1	2

ESTIMATION OF RECOMBINATION FRACTION IN LINKAGE AND GENE FREQUENCIES FROM BACKCROSSES

An illustrative example for the situation when the parental phenotypic mating is known now follows.

Let G and g be dominant and recessive alleles at the "main" locus, and D and d the dominant and recessive alleles at the "test trait" locus (usually well-known genetic marker), respectively. Let λ be the recombination fraction between these loci; p_1, q_1, with $p_1 + q_1 = 1$, the frequencies of alleles G and g and p_2, q_2, with $p_2 + q_2 = 1$, the frequencies of alleles D and d, respectively. The notation GD will be used for double dominant ($G-D-$), Gd and gD for single dominants ($G-dd$) and ($ggD-$), respectively, and gd for double recessive ($ggdd$).

Suppose that a sample of families is observed in which parental phenotypic mating type is Double Dominant $\times$ Double Recessive, that is, $GD \times gd$. For an estimation of linkage alone, only matings $GDgd \times ggdd$ will contribute effectively to the likelihood function. Of course, this genotypic mating can only be recognized when selection is through the children. Appropriate methods for detecting and testing linkage are described, for instance, by Morton [2] and Smith [3]. If, on the other hand, gene frequencies are to be estimated at the same time, families with other genotypic matings will be useful too.

Table 4 represents a model for segregation analysis when selection is through the

TABLE 4

A SEGREGATION MODEL FOR DETECTION AND ESTIMATION OF LINKAGE AND GENE FREQUENCIES. PARENTAL PHENOTYPIC MATING TYPE GD×gd

SEGREGATION PATTERN (S_t)	GENOTYPE MATINGS	EXPECTED PROPORTION OF MATINGS		SEGREGATION PARAMETERS (θ_{lt})			
		In a Population	In a Given Phenotypic Mating Type (ϕ_t)	GD	Gd	gD	gd
1.....	*GGDD*×*ggdd*	$2p_1^2q_1^2p_2^2q_2^2$	$[(1-q_1)(1-q_2)]/[(1+q_1)(1+q_2)]=(1-h_1)(1-h_2)$	1	0	0	0
2.....	*GGDd*×*ggdd*	$4p_1^2q_1^2p_2q_2^3$	$[2(1-q_1)q_2]/[(1+q_1)(1+q_2)]=(1-h_1)h_2$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
3.....	*GgDD*×*ggdd*	$4p_1q_1^3p_2^2q_2^2$	$[2(1-q_2)q_1]/[(1+q_1)(1+q_2)]=(1-h_2)h_1$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
4.....	*GD/gd*×*ggdd*	$4p_1q_1^3p_2q_2^3$	$[2q_1q_2]/[(1+q_1)(1+q_2)]=\frac{1}{2}h_1h_2$	$\frac{1}{2}(1-\lambda)$	$\frac{1}{2}\lambda$	$\frac{1}{2}\lambda$	$\frac{1}{2}(1-\lambda)$
5.....	*Gd/gD*×*ggdd*	$4p_1q_1^3p_2q_2^3$	$[2q_1q_2]/[(1+q_1)(1+q_2)]=\frac{1}{2}h_1h_2$	$\frac{1}{2}\lambda$	$\frac{1}{2}(1-\lambda)$	$\frac{1}{2}(1-\lambda)$	$\frac{1}{2}\lambda$
				NO. OF CHILDREN			
Total....		$2p_1p_2q_1^2q_2^2\times[(1+q_1)(1+q_2)]$	1	r_1	r_2	r_3	r_4

parents. The table is self-explanatory. To simplify the notation, the following two quantities have been introduced:

$$h_1 = \frac{2q_1}{1+q_1}, \quad \text{and} \quad h_2 = \frac{2q_2}{1+q_2}. \tag{25}$$

(Note that $\frac{1}{2}h_1$ and $\frac{1}{2}h_2$ are Snyder's ratios.) The likelihood function will be now $L_s(\gamma) = L_s(\lambda, h_1, h_2)$.

Table 5 has been reconstructed from table 4 and arranged according to family patterns. Since there were not many matings involved in each family pattern, table 4 gives the explicit equations for $P_s(\mathfrak{F}_\beta; \lambda, h_1, h_2)$.

TABLE 5

FAMILY PATTERNS WITHIN MATING TYPE GD×gd. ESTIMATION OF LINKAGE AND GENE FREQUENCIES

Family Pattern $\mathfrak{F}_\beta(r_1, \ldots, r_4)$	Probability $P_s(\mathfrak{F}_\beta; \lambda, h_1, h_2)$
1. GD; $r_1=s$	$\frac{1}{2}(1-h_1)\{1-h_2[1-(\frac{1}{2})^{s-1}]\}+\frac{1}{2}(1-h_2)\times\{1-h_1[1-(\frac{1}{2})^{s-1}]\}+(\frac{1}{2})^{s+1}h_1h_2[\lambda^s+(1-\lambda)^s]$
2. Gd; $r_2=s$	$(\frac{1}{2})^s\{(1-h_1)h_2+\frac{1}{2}h_1h_2[\lambda^s+(1-\lambda)^s]\}$
3. gD; $r_3=s$	$(\frac{1}{2})^s\{(1-h_2)h_1+\frac{1}{2}h_1h_2[\lambda^s+(1-\lambda)^s]\}$
4. gd; $r_4=s$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^s+(1-\lambda)^s]$
5. GD,Gd; $\begin{Bmatrix} r_1\neq 0, r_2\neq 0 \\ r_1+r_2=s \end{Bmatrix}$	$(\frac{1}{2})^s\{(1-h_1)h_2+\frac{1}{2}h_1h_2[\lambda^{r_1}(1-\lambda)^{r_2}+\lambda^{r_2}(1-\lambda)^{r_1}]\}$
6. GD,gD; $\begin{Bmatrix} r_1\neq 0, r_3\neq 0 \\ r_1+r_3=s \end{Bmatrix}$	$(\frac{1}{2})^s\{(1-h_2)h_1+\frac{1}{2}h_1h_2[\lambda^{r_1}(1-\lambda)^{r_3}+\lambda^{r_3}(1-\lambda)^{r_1}]\}$
7. Gd,gd; $\begin{Bmatrix} r_1\neq 0, r_4\neq 0 \\ r_1+r_4=s \end{Bmatrix}$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^s+(1-\lambda)^s]$
8. Gd,gD; $\begin{Bmatrix} r_2\neq 0, r_3\neq 0 \\ r_2+r_3=s \end{Bmatrix}$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^s+(1-\lambda)^s]$
9. Gd,gd; $\begin{Bmatrix} r_2\neq 0, r_4\neq 0 \\ r_2+r_4=s \end{Bmatrix}$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^{r_2}(1-\lambda)^{r_4}+\lambda^{r_4}(1-\lambda)^{r_2}]$
10. gD,gd; $\begin{Bmatrix} r_3\neq 0, r_4\neq 0 \\ r_3+r_4=s \end{Bmatrix}$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^{r_3}(1-\lambda)^{r_4}+\lambda^{r_4}(1-\lambda)^{r_3}]$
11. GD,Gd,gD; $\begin{Bmatrix} r_1\neq 0, r_2\neq 0, r_3\neq 0 \\ r_1+r_2+r_3=s \end{Bmatrix}$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^{r_1}(1-\lambda)^{r_2+r_3}+\lambda^{r_2+r_3}(1-\lambda)^{r_1}]$
12. GD,Gd,gd; $\begin{Bmatrix} r_1\neq 0, r_2\neq 0, r_4\neq 0 \\ r_1+r_2+r_4=s \end{Bmatrix}$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^{r_2}(1-\lambda)^{r_1+r_4}+\lambda^{r_1+r_4}(1-\lambda)^{r_2}]$
13. GD,gD,gd; $\begin{Bmatrix} r_1\neq 0, r_3\neq 0, r_4\neq 0 \\ r_1+r_3+r_4=s \end{Bmatrix}$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^{r_3}(1-\lambda)^{r_1+r_4}+\lambda^{r_1+r_4}(1-\lambda)^{r_3}]$
14. Gd,gD,gd; $\begin{Bmatrix} r_2\neq 0, r_3\neq 0, r_4\neq 0 \\ r_2+r_3+r_4=s \end{Bmatrix}$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^{r_2}(1-\lambda)^{r_2+r_3}+\lambda^{r_2+r_3}(1-\lambda)^{r_4}]$
15. GD,Gd,gD,gd; $\begin{Bmatrix} r_1\neq 0, r_2\neq 0, r_3\neq 0, r_4\neq 0 \\ r_1+r_2+r_3+r_4=s \end{Bmatrix}$	$(\frac{1}{2})^{s+1}h_1h_2[\lambda^{r_1+r_4}(1-\lambda)^{r_2+r_3}+\lambda^{r_2+r_3}(1-\lambda)^{r_1+r_4}]$

AUTOSOMAL INHERITANCE. TRUNCATED SELECTION

Suppose only those families which have at least one child with a specified trait (or traits) are to be included in the analysis. For instance, in estimating linkage between two loci with dominance, a sample of families with at least one child which is double recessive, or a sample of families in which one child is single recessive at one locus and simultaneously another child is single recessive at the second locus, will be more appropriate than a sample of families of all kinds of matings. In such cases, a random selection of families is still made, but those which do not satisfy our condition are excluded from analysis. This kind of sampling will be called a *truncated selection of families*. It is not, strictly speaking, selection through the children, and so the probability of ascertainment is not introduced. Let

$$R = \sum_i r_{l_i} \tag{26}$$

denote the condition on which the truncated selection is based, and $Q_{st}(R)$ the probability that the condition R is satisfied for a family of size s belonging to the $\mathfrak{S}_t$ segregation pattern.

Let $P_{st}(\bar{R})$ be the probability that the condition R is *not satisfied* for families of size s within the $\mathfrak{S}_t$ segregation pattern. Thus

$$Q_{st}(R) = 1 - P_{st}(\bar{R}) . \tag{27}$$

Let $Q_s(R)$ denote the probability that the condition R is satisfied in families of size s in the entire population (or within a given parental phenotypic mating type). Thus,

$$\begin{aligned} Q_s(R) &= 1 - \sum_{t=1}^{m} \phi_t P_{st}(\bar{R}) = \sum_{t=1}^{m} \phi_t - \sum_{t=1}^{m} \phi_t P_{st}(\bar{R}) \\ &= \sum_{t=1}^{m} \phi_t [1 - P_{st}(\bar{R})] = \sum_{t=1}^{m} \phi_t Q_{st}(R) . \end{aligned} \tag{28}$$

The probability, $P_s(r_1, r_2, \ldots, r_k; \gamma | R)$, that a family among those satisfying the condition R will have r_1 children in the category $\mathfrak{C}_1$, r_2 children in the category $\mathfrak{C}_2, \ldots, r_k$ children in the category $\mathfrak{C}_k$, with

$$\sum_{l=1}^{s} r_l = s \quad \text{and} \quad \sum_i r_{l_i} = R ,$$

is

$$P_s(r_1, r_2, \ldots, r_k; \gamma | R) = \Big[\sum_{t=1}^{m} \phi_t P_{st}(r_1, r_2, \ldots, r_k) \Big] \Big/ \Big[\sum_{t=1}^{m} \phi_t Q_{st}(R) \Big] . \tag{29}$$

By arranging the family data according to family patterns and using arguments similar to those for complete selection through parents, the probability P_s defined in (29) for the family $F_{\alpha(\beta)}$ is

$$P_s(F_{\alpha(\beta)}; \gamma | R) = \Big[\sum_{t \in \mathfrak{F}_\beta} \phi_t P_{st}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) \Big] \Big/ \Big[\sum_{t \in \mathfrak{F}_\beta} \phi_t Q\{R(\alpha)\} \Big] , \tag{30}$$

where $R(\alpha)$ denotes the condition R defined in (26) observed in the αth family, that is,

$$R(\alpha) = \sum_i r_{l_i\alpha} . \tag{31}$$

Denoting

$$c_s\{R(\alpha)\} = \frac{s!}{r_{l_1\alpha}!r_{l_2\alpha}!\ldots}, \quad \text{and} \quad W_{sl}\{R(\alpha)\} = \theta_{l_1}^{r_{l_1\alpha}}\theta_{l_2}^{r_{l_2\alpha}}\ldots \tag{32}$$

with

$$\sum_i r_{l_i\alpha} = R(\alpha) ,$$

(30) may be represented in the form analogous to (21), that is,

$$P_s(F_{\alpha(\beta)};\ \gamma|R) = c_s'(\alpha)\Big[\sum_{l\epsilon\mathfrak{F}_\beta}\phi_l T_{sl}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha})\Big] \Big/ \Big[\sum_{l\epsilon\mathfrak{F}_\beta}\phi_l W_{sl}\{R(\alpha)\}\Big] , \tag{33}$$

with

$$c_s'(\alpha) = [c_s(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha})]/[c_s\{R(\alpha)\}] , \tag{34}$$

where $c_s(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha})$ and $T_{sl}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha})$ are evaluated for the αth family from (18) and (19), respectively.

If n_s families, each of size s, are observed, the likelihood function, $L_s(\gamma)$, takes the form

$$\begin{aligned} L_s(\gamma) &= \prod_{\alpha=1}^{n_s} P_s(F_{\alpha(\beta)};\ \gamma|R) \\ &= C_s'\Big[\sum_{l\epsilon\mathfrak{F}_\beta}\phi_l T_{sl}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha})\Big] \Big/ \Big[\sum_{l\epsilon\mathfrak{F}_\beta}\phi_l W_{sl}\{R(\alpha)\}\Big] , \end{aligned} \tag{35}$$

where

$$C_s' = \prod_{\alpha=1}^{n_s} c_s'(\alpha) . \tag{36}$$

Modification for cases in which parental phenotypic mating types are known is straightforward. Modifications when families are ascertained through the children with ascertainment probability π should not cause difficulties (see [1]).

X-LINKED INHERITANCE

When the character is X-linked, the genotypes of the mother and of the father must be distinguished; in the general case, there may be f_1 distinct female genotypes and f_2 male genotypes. The matings *uv* and *vu* will now not be the same. Also, female and male offspring must be distinguished. Following arguments in the previous section, it should not be difficult to construct the likelihood function for daughters and for sons and combine them to obtain estimates of parameters under consideration.

ESTIMATION AND TESTING HYPOTHESES

If the total sample of all families, N, is sufficiently large, approximate methods of evaluating multinomial scores and the information matrix are applied in order to

obtain the maximum likelihood estimates, $\hat{\gamma}' = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_M)$, and to construct test statistics for testing hypotheses about these parameters and about homogeneity of the analyzed material. The general theory, with some numerical examples, has been presented in a previous paper [1]. However, to make this paper self-contained, the standard procedures will be briefly outlined, using for illustration the likelihood function defined in (35).

Likelihood Equations and Information Matrix

Put

$$A_s = A_s(\alpha, \gamma) = \sum_{t \in \mathfrak{F}_\beta} \phi_t T_{st}(r_{1\alpha}, r_{2\alpha}, \ldots, r_{k\alpha}) ,$$

and (37)

$$B_s = B_s(\alpha, \gamma) = \sum_{t \in \mathfrak{F}_\beta} \phi_t W_{st}[R(\alpha)] .$$

The likelihood function in (35) can be written as

$$L_s(\gamma) = L_s = C_s' \prod_{\alpha=1}^{n_s} [A_s(\alpha, \gamma)/B_s(\alpha, \gamma)] = C_s' \prod_{\alpha=1}^{n_s} (A_s/B_s) , \tag{38}$$

and its logarithm

$$\log L_s = \log C_s' + \sum_{\alpha=1}^{n_s} (\log A_s - \log B_s) . \tag{39}$$

Differentiating (39) with respect to γ_i, we obtain

$$\frac{\partial \log L_s}{\partial \gamma_i} = \sum_{\alpha=1}^{n_s} \left(\frac{1}{A_s} \frac{\partial A_s}{\partial \gamma_i} - \frac{1}{B_s} \frac{\partial B_s}{\partial \gamma_i} \right), \qquad i = 1, 2, \ldots, M . \tag{40}$$

Put

$$u_{i:s}(\alpha; \gamma) = u_{i:s}(\alpha) = \frac{1}{A_s} \frac{\partial A_s}{\partial \gamma_i} - \frac{1}{B_s} \frac{\partial B_s}{\partial \gamma_i} , \tag{41}$$

$i = 1, 2, \ldots, M, \alpha = 1, 2, \ldots, n_s$. The quantity $u_{i;s}(\alpha)$ denotes an individual score for the αth family with respect to the parameter γ_i.

The total score for a sample of n_s families, $U_{i;s}$, is

$$U_{i:s} = \sum_{\alpha=1}^{n_s} u_{i;s}(\alpha) , \qquad i = 1, 2, \ldots, M , \tag{42}$$

and the overall score, U_i, when families of sizes $s = 1, 2, \ldots, S$ are taken into account, is

$$U = \sum_{s=1}^{S} U_{i:s} = \sum_{s=1}^{S} \sum_{\alpha=1}^{n_s} u_{i;s}(\alpha) . \tag{43}$$

To obtain the ML estimates, $\hat{\gamma}' = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_M)$, solve M (linearly independent) equations of the form

$$U_i = \sum_{s=1}^{S} \sum_{\alpha=1}^{n_s} \left[\frac{1}{A_s} \frac{\partial A_s}{\partial \gamma_i} - \frac{1}{B_s} \frac{\partial B_s}{\partial \gamma_i} \right] = 0 , \qquad i = 1, 2, \ldots, M . \tag{44}$$

A unique set of solutions exists provided

$$M < \sum_{s=1}^{S} (s - 1) = S(S - 1)/2 ,$$

which is usually the case.

The elements of the expected information matrix, $I_s(\gamma) = I_s$, are

$$I_{ii;s}(\gamma) = I_{ii\cdot s} = \sum_{\alpha=1}^{n_s} \frac{A_s}{B_s} u_{i;s}^2(\alpha) , \qquad i = 1, 2, \ldots, M ,$$

and (45)

$$I_{ij;s}(\gamma) = I_{ij;s} = \sum_{\alpha=1}^{n_s} \frac{A_s}{B_s} u_{i;s}(\alpha) u_{j;s}(\alpha) , \qquad i, j = 1, 2, \ldots, M ; i \neq j .$$

These quantities should be evaluated for the expected values of the parameters $\gamma' = (\gamma_1, \gamma_2, \ldots, \gamma_M)$. If these values are unknown, substitute the estimates $\hat{\gamma}' = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_M)$ and obtain the estimated information matrix $I_s(\hat{\gamma})$.

The elements of the overall information matrix, $I(\gamma) = I$, are

$$I_{ii} = \sum_{s=1}^{S} I_{ii;s} ; \quad I_{ij} = \sum_{s=1}^{S} I_{ij;s} , \qquad i, j = 1, 2, \ldots, M ; i \neq j , \tag{46}$$

so that

$$I = \sum_{s=1}^{S} I_s . \tag{47}$$

Testing Hypotheses

Let H_0: $\gamma = \gamma_0$ be the null hypothesis which specifies all the parameters. Let

$$U_s'(\gamma_0) = [U_{1;s}(\gamma_0), U_{2;s}(\gamma_0), \ldots, U_{M\cdot s}(\gamma_0)] \tag{48}$$

be the $1 \times M$ vector of scores $U_{i;s}$ evaluated for the values $\gamma = \gamma_0$, and $I_s(\gamma_0)$ be the expected information matrix.

Let H_0' be another null hypothesis that all families obey the same segregation ratios, or that the data are *homogeneous*.

If both H_0 and H_0' are true simultaneously, the statistic

$$X^2_{\text{Total}} = \sum_{s=2}^{S} U_s'(\gamma_0) I_s^{-1}(\gamma_0) U_s(\gamma_0) \tag{49}$$

is approximately distributed as χ^2 with $\frac{1}{2}M(2S - M - 1)$ degrees of freedom (for justification of the number of df see Elandt-Johnson [1]).

If X^2_{Total} is significant, this means that either H_0 is not true or the data are not homogeneous (H_0' is not true), or both. Let

$$U(\gamma_0) = \sum_{s=1}^{S} U_s(\gamma_0) , \quad \text{and} \quad I(\gamma_0) = \sum_{s=1}^{S} I_s(\gamma_0) \tag{50}$$

be the overall vector of scores and the expected (under H_0) information matrix, respectively, for *combined* data from families of all sizes.

If H_0 is "true on the average," the statistic

$$X^2_{\text{Comb}} = U'(\gamma_0) I^{-1}(\gamma_0) U(\gamma_0) \tag{51}$$

is approximately distributed as χ^2 with M df.

To test the hypothesis H'_0, that is, about homogeneity of family data alone, use the statistic

$$X^2_{\text{Diff}} = X^2_{\text{Total}} - X^2_{\text{Comb}} . \tag{52}$$

If the data are homogeneous, X^2_{Diff} is approximately distributed as χ^2 with $\frac{1}{2}M(2S - M - 1) - M = \frac{1}{2}M(2S - M - 3)$ df.

The results of the tests using these three statistics should be interpreted jointly. For instance, if:

1. X^2_{Total} significant, X^2_{Comb} significant, X^2_{Diff} nonsignificant, this means that H_0 is probably not true and should be rejected, but the data are homogeneous with respect to another unknown hypothesis, $H_2^{(x)}$, say.
2. X^2_{Total} significant, X^2_{Comb} significant, X^2_{Diff} significant indicates that not only H_0 is not true but also the data are not homogeneous.

When all (or some) parameters are not specified by H_0, we substitute their estimators. Of course, the number of degrees of freedom has to be appropriately corrected.

Putting $B_s(\alpha, \gamma) = 1$ in (38) gives complete selection of families.

It is obvious that methods for complex segregation analysis require high-speed computer facilities. Since several statistical and genetic laboratories have general programs for solving maximum likelihood equations, details are not given about differentiation and iteration procedures. An appendix in a previous paper [1] might be useful for those who wish to design their own programs.

SUMMARY

General models for segregation analysis for common genetic traits, with offspring classified into more than two distinct categories, have been derived. Problems of estimating gene frequencies for such common traits as the ABO–blood group system and of estimating linkage and gene frequencies from family data have been presented. Likelihood equations and X^2 analysis have been briefly outlined.

REFERENCES

1. ELANDT-JOHNSON RC: Segregation analysis for complex modes of inheritance. *Amer J Hum Genet* 22:129–144, 1970
2. MORTON NE: Sequential test for the detection of linkage. *Amer J Hum Genet* 7:277–318, 1955
3. SMITH CAB: Some comments on the statistical methods used in linkage investigation. *Amer J Hum Genet* 11:289–304, 1959