Am J Hum Genet 34:531-551, 1982

Utility and Efficiency of Linked Marker Genes for Genetic Counseling.

II. Identification of Linkage Phase by Offspring Phenotypes

ARAVINDA CHAKRAVARTI[1,2] AND MASATOSHI NEI[1]

SUMMARY

For a linked marker locus to be useful for genetic counseling, the counselee must be heterozygous for both disease and marker loci and his or her linkage phase must be known. It is shown that when the phenotypes of the counselee's previous children for the disease and marker loci are known, the linkage phase can often be inferred with a high probability, and thus it is possible to conduct genetic counseling. To evaluate the utility of linked marker genes for genetic counseling, the accuracy of prediction of the risk for a prospective child with a given marker gene to develop the genetic disease and the proportion of families in which a particular marker locus can be used for genetic counseling are studied for X-linked recessive, autosomal dominant, and autosomal recessive diseases. In the case of X-linked genetic diseases, information from children is very useful for determining the linkage phase of the counselee and predicting the genetic disease. In the case of autosomal dominant diseases, not all children are informative, but if the number of children is large, the phenotypes of children are often more informative than the information from grandparents. In the case of autosomal recessive diseases, information from grandparents is usually useless, since they show a normal phenotype for the disease locus. If we use information on the phenotypes of children, however, the linkage phase of the counselee and the risk of a prospective child can be inferred with a high probability. The proportion of informative families depends on the dominance relationship and frequencies of marker alleles, and the number of children. In general, codominant markers are more useful than are dominant

Received March 24, 1981; revised October 26, 1981.

[1] Center for Demographic and Population Genetics, University of Texas at Houston, Houston, TX 77025.

[2] Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261.

markers, and a locus with high heterozygosity is more useful than is a locus with low heterozygosity.

INTRODUCTION

The idea of using linked marker genes for genetic counseling was first mentioned by Hoogvliet [1], who used the color-blindness gene to detect a carrier of the hemophilia gene. Since then, this technique has been used for a number of genetic diseases. Some mathematical problems related to this technique have also been studied [2–5]. Recently, the usefulness of this technique has increased because of the discovery of many marker genes. It is now possible to diagnose β-thalassemia and sickle-cell anemia prenatally by using this technique [6–8]. In the future, this technique will be more important since many new polymorphic genes are expected to be discovered by using restriction endonucleases [9].

To evaluate the utility of marker loci for genetic counseling, two criteria are required [4, 5]. One is the accuracy of prediction or the probability with which an individual with a given marker gene contracts the disease in question. This probability depends on the recombination value between the marker and disease genes and how the linkage phase is determined. The other is the proportion of individuals or families (informative families) in which a particular marker locus can be used for genetic counseling. This proportion depends on the frequencies and dominance of marker genes. Obviously, the marker locus must be polymorphic; otherwise, it has no utility for genetic counseling.

In some cases, the linkage phase of a counselee (parent) can be determined unambiguously through information on the phenotype of grandparents. In this case, the accuracy of prediction of genetic disease is solely dependent on the recombination value. Therefore, the evaluation of the accuracy of prediction and the proportion of informative families is relatively simple, and a systematic study of this problem has been conducted by Nei [4, 5]. However, information on the phenotypes of grandparents is not always available, and if it is not available, we must use information from other relatives. The most important relative for this purpose is offspring. When we use information on offspring, however, the linkage phase cannot be determined with certainty, and it must be estimated. This can be done if we use the Bayesian probability approach. Here we evaluate the accuracy of prediction of genetic disease and the proportion of informative families by using this approach.

The Bayesian probability method has previously been used for risk estimation in a variety of genetic counseling situations when linkage information is absent [10]. For simple situations, the calculation of risk is facilitated by the use of the computer program PEDIG [11]. In more complex cases, such as those involving extensive pedigree information and familial diseases, the method of Cannings et al. [12] may be used with the aid of the computer program PAP [13]. Here we shall

consider simple family structures and present mathematical formulas that can easily be used for genetic counseling through linked marker genes.

X-LINKED RECESSIVE DISEASES

Codominant Markers

Let us first consider the genetic counseling of X-linked recessive diseases with codominant marker genes. In this case, the counselee is a female and carries a recessive disease gene, d, in heterozygous condition. We designate by D its allelic normal gene. We assume that the carrier status of the female is known because of the phenotype of her father or her previous sons or from a biochemical carrier-detection test. Our problem is to predict the risk for a son with a marker gene to develop the genetic disease. Let M_1 and M_2 be the two alleles at the marker locus, and r be the recombination value between the two loci. The counselee must be a double heterozygote for these loci. Otherwise, the marker locus has no utility for increasing the predictability of genetic diseases. It is also important to know the linkage phase of the counselee, although this information is not always available. As mentioned earlier, we infer this linkage phase from the phenotype of the counselee's offspring by using the Bayesian approach. In the case of X-linked recessive diseases, we consider only male children, since female children generally do not contract the disease.

When a counselee is a double heterozygote, her linkage phase is either coupling (DM_1/dM_2) or repulsion (DM_2/dM_1). Suppose that the counselee is in coupling phase. Then, the expected frequencies of the four different genotypes, DM_1, DM_2, dM_1, and dM_2, among her sons are $(1 - r)/2$, $r/2$, $r/2$, and $(1 - r)/2$, respectively. Therefore, if the genotype of a son is M_1 at the marker locus, the probability that he has the disease gene d is r. Thus, if the recombination value is very small, we can predict with a high probability that he is normal. On the other hand, if the counselee is in repulsion phase, the expected frequencies of DM_1, DM_2, dM_1, and dM_2 are $r/2$, $(1 - r)/2$, $(1 - r)/2$, and $r/2$, respectively. Thus, if a son has allele M_1 at the marker locus, he is expected to develop the genetic disease with probability $1 - r$.

The principle of using offspring information for predicting the linkage phase is as follows: Suppose that the genotype of a counselee is DdM_1M_2 and she has two affected sons (d) with genotype M_2 at the marker locus. If the linkage phase of the counselee (mother) is coupling (DM_1/dM_2), the probability of having two sons with these genotypes is $P(A|C) = (1 - r)^2/4$. On the other hand, if the linkage phase is repulsion, the probability is $P(A|R) = r^2/4$. We assume that the disease and marker loci are in linkage equilibrium, so that the coupling and repulsion phases are equally frequent in the population. Then, the Bayesian posterior probability that the linkage phase is coupling is

$$P(C) = \frac{P(A|C)}{P(A|C) + P(A|R)} = \frac{(1 - r)^2}{(1 - r)^2 + r^2} . \quad (1)$$

Similarly, the probability of repulsion is $P(R) = r^2/[(1 - r)^2 + r^2]$. Therefore, if $r = .01, P(C) = .9999$ and $P(R) = .0001$. Thus, we can predict the linkage phase with a high probability.

In practice, a counselee may have both affected and normal sons as well as recombinant and nonrecombinant types. Let n_1, n_2, n_3, and n_4 be the numbers of sons with genotypes DM_1, DM_2, dM_1, and dM_2, respectively. If the linkage phase of the mother is coupling, the probability of having n sons with this set of genotypes is

$$P(A|C) = r^{n_2 + n_3}(1 - r)^{n_1 + n_4}/2^n , \tag{2a}$$

where $n = n_1 + n_2 + n_3 + n_4$. On the other hand, if the linkage phase is repulsion, the probability is

$$P(A|R) = r^{n_1 + n_4}(1 - r)^{n_2 + n_3}/2^n . \tag{2b}$$

Therefore, the posterior probability that the mother is in coupling phase is

$$P(C) = \frac{r^{n_2 + n_3}(1 - r)^{n_1 + n_4}}{r^{n_2 + n_3}(1 - r)^{n_1 + n_4} + r^{n_1 + n_4}(1 - r)^{n_2 + n_3}}$$

$$= \frac{1}{1 + \rho^{\alpha}} , \tag{3a}$$

where $\rho = r/(1 - r)$ and $\alpha = n_1 + n_4 - (n_2 + n_3)$. Furthermore, we have

$$P(R) = 1 - P(C) = 1/(1 + \rho^{-\alpha}) . \tag{3b}$$

Formula (3a) indicates that $P(C)$ depends on ρ and α. When ρ is small, even $\alpha = 1$ gives a high value of $P(C)$. When $\alpha = 0, P(C) = P(R) = 1/2$, and information from children is of no use. However, if r is small, this event is expected to occur very rarely, as will be discussed later. The value of $P(C)$ for various values of r is given in table 1. It is clear that when r is small and α is large the linkage phase is determined with a high probability.

Prediction of genetic disease. Suppose that a counselee wants to know the probability that her next son will develop the genetic disease in question. Consider the case where her son's genotype at the marker locus is M_1. In this case, if the mother is in coupling phase, the probability that her son will develop the genetic disease is r, whereas if the mother is in repulsion phase, the probability is $1 - r$, as mentioned earlier. From the information on her previous sons, the probabilities of coupling and repulsion have been computed to be $P(C)$ and $P(R)$ from formulas (3a) and (3b), respectively. Therefore, the Bayesian probability that he will develop the genetic disease is

$$R(M_1) = rP(C) + (1 - r)P(R) . \tag{4a}$$

TABLE 1

POSTERIOR PROBABILITIES OF COUPLING PHASE [$P(C)$] AND THE RISKS FOR A MALE OFFSPRING WITH MARKER GENE M_1 TO HAVE X-LINKED DISEASE GENE d [$R(M_1)$]

	r = .05		r = .005		r = .001	
α	$P(C)$	$R(M_1)$	$P(C)$	$R(M_1)$	$P(C)$	$R(M_1)$
0	.5000	.5000	.5000	.5000	.5000	.5000
1	.9500	.0950	.9950	.0099	.9990	.0020
2	.9972	.0525	1.0000	.0050	1.0000	.0010
3	.9998	.0501	1.0000	.0050	1.0000	.0010
$\geq$ 4	1.0000	.0500	1.0000	.0050	1.0000	.0010

NOTE: $\alpha = n_1 + n_4 - (n_2 + n_3)$, where n_1, n_2, n_3, and n_4 are the nos. of DM_1, DM_2, dM_1, and dM_2 males, respectively, and r is the recombination value.

On the other hand, the risk for a son with genotype M_2 to develop the disease is

$$R(M_2) = (1 - r)P(C) + rP(R)$$
$$= 1 - R(M_1) . \tag{4b}$$

The values of $R(M_1)$ for various values of r and α are given in table 1. $R(M_1)$ rapidly decreases with increasing α and reaches the value for the case of unambiguous determination of linkage phase when $\alpha \geq 4$ and $r \leq .05$. $1 - R(M_1)$ may be called the accuracy of prediction of genetic disease. This accuracy cannot be made close to 1, unless r is close to 0.

So far we have confined ourselves to male offspring only. However, formulas (4a) and (4b) give the probability of a female offspring being a disease-gene carrier as well, if information on the father's genotype is given. Namely, $R(M_1)$ and $R(M_2)$ give the probabilities of carrier status when the maternally inherited gene is M_1 and M_2, respectively.

Proportion of informative families. As mentioned above, linked marker genes are useful only when the counselee is a double heterozygote and the linkage phase is known. Consider a counselee who is known to be a disease-gene carrier from her previous sons. She will be a double heterozygote, if the marker locus is heterozygous. If the frequency of allele M_1 in the population is x_1, the probability of an individual to be heterozygous is $2x_1(1 - x_1)$. If there are multiple alleles at the marker locus, this probability will of course be $1 - \Sigma x_i^2$, where x_i is the frequency of the ith allele.

We have seen that the linkage phase of the counselee can be determined whenever $\alpha \neq 0$. If $\alpha = 0$, however, linked marker genes are of no use. Let us now consider the probability of $\alpha = 0$ for the case in which the counselee has n sons. First assume that the counselee is in coupling phase (DM_1/dM_2). In this case, the recombinant genotypes DM_2 and dM_1 appear with probability r among her sons, whereas the nonrecombinants DM_1 and dM_2 appear with probability $1 - r$. Therefore, the probability of $\alpha = 0$ is $\binom{n}{n/2} r^{n/2}(1 - r)^{n/2}$. When the mother is in

repulsion phase, the genotypes DM_2 and dM_1 appear with probability $1 - r$, whereas DM_1 and dM_2 appear with probability r. Therefore, the probability of $\alpha = 0$ is the same as that for the case of coupling. Thus, for computing the proportion of families with $\alpha = 0$, we do not have to consider the linkage phase of the counselee.

The probability of $\alpha = 0$ becomes $2r(1 - r)$ when $n = 2$. This value is very small when r is small. In the case of $n = 4$, this probability is even smaller, that is, $6r^2(1 - r)^2$. Thus, as long as r is small, say, smaller than .01, the probability of $\alpha = 0$ is very small, and we can assume that practically in all cases the linkage phase of the counselee can be determined from information on the phenotypes of children. The proportion of informative families is then given by

$$I = 1 - \Sigma x_i^2 , \tag{5}$$

which is the same as that for the case of determination of linkage phase through grandparents [4, 5]. It is clear from formula (5) that a locus with high heterozygosity is more useful for genetic counseling than a locus with low heterozygosity.

If one is interested in predicting the carrier status of a female child, information on the phenotype of the father is also necessary. However, the father can have any marker genotype, so that the proportion of informative families is identical with formula (5).

Dominant Markers

We designate the dominant and recessive alleles at the marker locus by M and m, respectively. In the presence of dominance, genotype Mm shows the phenotype M, and the heterozygous status of the counselee at the marker locus is detected only by the presence of recessive individuals among her children. However, once a counselee is known to be a double heterozygote, the linkage phase can be determined through information on the phenotypes of her children. Let n_1, n_2, n_3, and n_4 be the number of male children with genotypes DM, Dm, dM, and dm, respectively. The estimation of the probability of linkage phase and the prediction of genetic diseases are then exactly the same as those for codominant markers, and formulas (3) and (4) and table 1 are directly applicable. Formulas (4a) and (4b) can also be used for computing the probability of a female child being a carrier if the father has genotype m.

The proportion of informative mothers (families) is the proportion of counselees who are heterozygous for the marker locus and whose linkage phase can be determined. Under the assumption of linkage equilibrium, a counselee with a D gene is heterozygous for the M locus with probability $2x(1 - x)$, where x is the frequency of allele m. However, the heterozygosity of the counselee at the M locus cannot be detected unless she has at least one offspring with phenotype m. When there are n children born to the counselee, the probability that at least one of them has phenotype m is $1 - (1/2)^n$. As mentioned above, if the number of recombinant individuals equals that of nonrecombinants ($\alpha = 0$), children do not give any information about the linkage phase of the counselee, but this event can be neglected

unless r is large. Therefore, for families with n children, the proportion of informative families is given by

$$I = 2\left[1 - \left(\frac{1}{2}\right)^n\right]x(1 - x) \ . \tag{6}$$

It is noted that when $n = 1, I = x(1 - x)$. This is smaller than that for the case of determination of linkage phase through grandparents, since the latter is $x(1 - x^2)$ [4, 5]. However, if n is large, the information from children can be equally or even more useful than that from grandparents.

When the carrier status of a female child is to be determined, the father's genotype at the M locus must be m. Therefore, the proportion of informative families is

$$I = 2\left[1 - \left(\frac{1}{2}\right)^n\right]x^2(1 - x) \ . \tag{7}$$

AUTOSOMAL DOMINANT DISEASES

The linkage method for antenatal diagnosis and genetic counseling is very important for autosomal dominant diseases, since in these diseases biochemical tests are not generally available. We assume that the frequency of disease gene D is so low that the mutant homozygote does not occur, and all affected individuals have genotype Dd, where d is the normal allele. The spouse of a counselee is assumed to be normal with genotype dd. As before, r is the recombination value between the two loci, and we assume that there is linkage equilibrium between the two loci.

Codominant Markers

Let us assume that the marker locus has two codominant alleles: M_1 and M_2. Then, a counselee has the genotype of either DM_1/dM_2 or DM_2/dM_1. The spouse of the counselee can have any of the three genotypes: dM_1/dM_1, dM_1/dM_2, and dM_2/dM_2. For codominant markers, all children will give some information for determining linkage phase probabilistically except in the mating $DdM_1M_2 \times ddM_1M_2$, where only the M_1M_1 and M_2M_2 offspring are useful [5]. This can be seen from table 2, where the expected frequencies of different genotypes in the offspring from the mating $DM_1/dM_2 \times dM_1/dM_2$ are given. It is clear that if the genotype of

TABLE 2

RELATIVE FREQUENCIES OF DIFFERENT GENOTYPES FROM THE MATING $DM_1/dM_2 \times dM_1/dM_2$

Gamete	dM_1 1/2	dM_2 1/2
DM_1 $(1 - r)/2$	DM_1/dM_1	DM_1/dM_2
DM_2 $r/2$	DM_2/dM_1	DM_2/dM_2
dM_1 $r/2$	dM_1/dM_1	dM_1/dM_2
dM_2 $(1 - r)/2$	dM_2/dM_1	dM_2/dM_2

an offspring is M_1M_2, it gives no information about the disease gene since the probability of having gene D is equal to that of having gene d. We note that genotype M_1M_2 appears with probability 1/2 in the offspring of $DdM_1M_2 \times ddM_1M_2$.

Consider the mating $DdM_1M_2 \times ddM_1M_1$, and let n_1, n_2, n_3, and n_4 be the numbers of children with genotypes DdM_1M_1, DdM_1M_2, ddM_1M_1, and ddM_1M_2, respectively. Estimation of the probability of linkage phase and the risk of having genetic disease is then exactly the same as that for codominant markers linked to X-linked diseases, and formulas (3) and (4) and table 1 directly apply: namely, the probabilities that the offspring with marker genotypes M_1M_1 and M_1M_2 develop the genetic disease, $R(M_1M_1)$ and $R(M_1M_2)$, are given by $R(M_2)$ and $R(M_1)$ in formula (4), respectively. Essentially the same computation can be made by using formula (4) for the mating $DdM_1M_2 \times ddM_2M_2$. For the family $DdM_1M_2 \times ddM_1M_2$, the same comment applies, but only the M_1M_1 and M_2M_2 offspring are useful.

Let us now consider the proportion of informative families. We denote by x_1 and x_2 the frequencies of the marker alleles M_1 and M_2, respectively. A counselee who is a carrier of the disease gene is informative if he or she has marker genotype M_1M_2 with known linkage phase and $\alpha \neq 0$ among his or her offspring. As in the case of sex-linked diseases, the probability of $\alpha = 0$ is very low if r is small, so that the possibility of $\alpha = 0$ can be neglected in practice. As mentioned above, all offspring of DdM_1M_2 are informative except when the counselee mates with ddM_1M_2 and the genotypes of the offspring are all M_1M_2 at the M locus. When there are n children born to a counselee, the probability that all of them have genotype M_1M_2 is $(1/2)^n$. Therefore, the proportion of informative families for n children is

$$I = 2x_1x_2\left[1 - \left(\frac{1}{2}\right)^{n-1} x_1x_2\right] \tag{8}$$

for $n \geq 1$. It is noted that if $n = 1$, $I = 2x_1x_2(1 - x_1x_2)$, which is equal to that for the case of determination of linkage phase through grandparents [4, 5]. When $n \geq 2$, it is higher than the latter value.

When there are multiple alleles at the marker locus, the proportion of informative families is

$$I = 2\sum_{i<j} x_ix_j\left[1 - \left(\frac{1}{2}\right)^{n-1} x_ix_j\right], \tag{9}$$

where x_i is the frequency of the ith allele.

Dominant Markers

We denote the dominant and recessive alleles at the marker locus by M and m, respectively. In the presence of dominance at the marker locus, we must first know the heterozygous status (Mm) of the counselee who is a known carrier of the disease gene D. The heterozygous status is established if the counselee has the dominant

phenotype M and one of the offspring has phenotype m. Therefore, the spouse of the counselee must have marker genotype Mm or mm. The informative families would then be $DdMm \times ddMm$ and $DdMm \times ddmm$.

Let us consider the family $DdMm \times ddmm$, and let n_1, n_2, n_3, and n_4 be the numbers of children with genotype $DdMm$, $Ddmm$, $ddMm$, and $ddmm$, respectively. Since the genotypes of all children can be identified in this type of family, formulas (3) and (4) and table 1 are directly applicable, except that in formula (4), $R(M_1)$ and $R(M_2)$ refer to the risks [$R(m)$ and $R(M)$] for the offspring with phenotypes m and M (genotypes mm and Mm), respectively.

For the family $DdMm \times ddMm$, let n_1, n_2, n_3, and n_4 be the numbers of children with phenotypes DdM-, $Ddmm$, ddM-, and $ddmm$, respectively, where M- represents either MM or Mm. If the counselee is in coupling phase (DM/dm), the expected frequencies of the four phenotypes are $(2 - r)/4$, $r/4$, $(1 + r)/4$, and $(1 - r)/4$, respectively. Therefore, the probability of obtaining n_1 DdM-, n_2 $Ddmm$, n_3 ddM-, and n_4 $ddmm$ is

$$P(A|C) = (2 - r)^{n_1} r^{n_2} (1 + r)^{n_3} (1 - r)^{n_4} / 4^n , \tag{10a}$$

where $n = n_1 + n_2 + n_3 + n_4$. Similarly, if the counselee is in repulsion phase (Dm/dM), the expected frequencies of DdM-, $Ddmm$, ddM-, and $ddmm$ are $(1 + r)/4$, $(1 - r)/4$, $(2 - r)/4$, and $r/4$, respectively. Therefore, we have

$$P(A|R) = (1 + r)^{n_1} (1 - r)^{n_2} (2 - r)^{n_3} r^{n_4} / 4^n . \tag{10b}$$

Thus, the posterior probability that the counselee is in coupling phase is

$$P(C) = \frac{1}{1 + \rho_1^{\alpha} \rho_2^{\beta}} , \tag{11}$$

where $\rho_1 = r/(1 - r)$, $\rho_2 = (1 + r)/(2 - r)$, $\alpha = n_4 - n_2$ and $\beta = n_1 - n_3$. Also, $P(R) = 1 - P(C) + 1/(1 + \rho_1^{-\alpha} \rho_2^{-\beta})$. It is clear from formula (11) that $P(C)$ is different from $1/2$ unless both α and β are zero. When $\alpha = \beta = 0$, the offspring phenotypes give no information on linkage phase and thus the marker genes are not useful for genetic counseling. However, if r is small, the probability of this event is negligibly small. One can also show that $\alpha = \beta = 0$ occurs only when n is even, so that the offspring are always informative when the sibship size is an odd number.

Since the expected frequencies of genotypes DdM-, $Ddmm$, ddM-, and $ddmm$ are $(2 - r)/4$, $r/4$, $(1 + r)/4$, and $(1 - r)/4$ for the case of coupling counselees, the risk for a child with marker phenotype m to contract the genetic disease, $R(m)$, is r. Thus, if r is small, the risk is very small. However, if the phenotype of an offspring is M, the individual will contract the disease with a probability of $[(2 - r)/4]/[(2 - r)/4 + (1 + r)/4] = (2 - r)/3$. This is very close to $2/3$ if r is small. Namely, dominant phenotypes are not very useful for genetic counseling.

In practice, of course, the linkage phase cannot be determined with certainty and must be inferred from the phenotypes of children. Therefore, the risks are given by

$$R(m) = rP(C) + (1 - r)P(R) , \tag{12a}$$

$$R(M) = [(2 - r)P(C) + (1 + r)P(R)]/3$$

$$= [2 - R(m)]/3 . \tag{12b}$$

Table 3 presents the values of the probability of linkage phase and the risk of genetic disease for an offspring with marker phenotype m when $\alpha \geq 0$ and $\beta \gtrless 0$. It is clear from the table that when r is small, both $P(C)$ and $R(m)$ rapidly approach the values for the case of complete determination of linkage phase even for a small value of α. Thus, the offspring information is very useful for genetic counseling. In table 3, the accuracy of prediction of disease status $[1 - R(m)]$ is very high if $\alpha \geq 1$ and r is small. Note that the values of $P(R)$ and $1 - R(m)$ for the case of $\alpha < 0$ and $\beta < 0$ (or $\beta > 0$) are equal to the values of $P(C)$ and $R(m)$ for $\alpha > 0$ and $\beta > 0$ (or $\beta < 0$) in table 3. The value of $R(M)$ can be computed from formula (12b), but whatever the value of $R(m)$, $R(M)$ cannot be close to 0 or

TABLE 3

POSTERIOR PROBABILITIES OF COUPLING PHASE $[P(C)]$ AND THE RISKS FOR AN OFFSPRING WITH PHENOTYPE m TO HAVE AUTOSOMAL DISEASE GENE D $[R(m)]$

		r = .05		r = .005		r = .001	
α	β	$P(C)$	$R(m)$	$P(C)$	$R(m)$	$P(C)$	$R(m)$
0	−3	.1350	.8284	.1134	.8827	.1116	.8877
	−2	.2248	.7477	.2024	.7946	.2005	.7989
	−1	.3500	.6300	.3350	.6633	.3337	.6660
	0	.5000	.5000	.5000	.5000	.5000	.5000
	1	.6500	.3650	.6650	.3366	.6663	.2340
	2	.7752	.2523	.7976	.2053	.7995	.2011
	3	.8650	.1705	.8867	.1172	.8884	.1123
1	−3	.7479	.2769	.9622	.0424	.9921	.0089
	−2	.8464	.1882	.9806	.0242	.9960	.0049
	−1	.9110	.1301	.9901	.0147	.9980	.0029
	0	.9500	.1100	.9950	.0099	.9990	.0020
	1	.9724	.0748	.9975	.0075	.9995	.0015
	2	.9850	.0635	.9987	.0063	.9997	.0013
	3	.9919	.0573	.9994	.0056	.9999	.0011
2	−3	.9827	.0657	.9998	.0051	1.0000	.0010
	−2	.9905	.0585	.9999	.0051	1.0000	.0010
	−1	.9949	.0546	1.0000	.0050	1.0000	.0010
	0	.9972	.0523	1.0000	.0050	1.0000	.0010
	1	.9985	.0513	1.0000	.0050	1.0000	.0010
	2	.9992	.0507	1.0000	.0050	1.0000	.0010
	3	.9996	.0504	1.0000	.0050	1.0000	.0010

NOTE: $\alpha = n_4 - n_2$ and $\beta = n_1 - n_3$, where n_1, n_2, n_3, and n_4 are the nos. of *DdM-*, *Ddmm*, *ddM-*, and *ddmm* children; r is the recombination value.

1. Therefore, it is difficult to make any definite statement about the risk of genetic disease for children with the dominant phenotype.

We now consider the proportion of informative families. We have seen that there are two types of informative matings, that is, *DdMm* × *ddmm* and *DdMm* × *ddMm*. Even in these matings, if $\alpha = 0$, children are not informative. However, the probability of the event of $\alpha = 0$ is very small, so that it can be neglected, as before. Therefore, the proportion of informative families is the probability that *DdMm* mates with genotypes *ddMm* and *ddmm* and there is at least one *mm* child among the children already born. Namely, for a family of progeny size n, it is

$$I = 4x^2(1 - x)^2\left[1 - \left(\frac{3}{4}\right)^n\right] + 2x^3(1 - x)\left[1 - \left(\frac{1}{2}\right)^n\right]. \qquad (13)$$

Therefore, if x is small, dominant markers are not very useful for genetic counseling.

AUTOSOMAL RECESSIVE DISEASES

Heterozygotes for autosomal recessive genes are usually identified through their affected offspring. With the advent of biochemical techniques, they may also be identified by carrier detection tests. In the case of autosomal recessive diseases, information from grandparents is useless unless their genotypes for the disease locus are identified by a biochemical test. At any rate, the counselees for this case are usually a couple both of whom are heterozygous for the disease gene *d* and its normal allele *D*. For linked marker genes to be useful in this case, at least one of the partners must be heterozygous for the marker locus.

Codominant Markers

Suppose that there are two codominant alleles, M_1 and M_2, at a marker locus. For convenience, genotypes DM_1/dM_2 and DM_2/dM_1 will be called the coupling (C) and repulsion (R) linkage phases, respectively. It is clear that the genotypes of the counselees must be $DdM_1M_2 \times DdM_1M_1$, $DdM_1M_2 \times DdM_2M_2$, or $DdM_1M_2 \times DdM_1M_2$. For codominant markers, all children from these families will give some information for estimating the probability of linkage phase.

Consider the informative family $DdM_1M_2 \times DdM_1M_1$. When the double heterozygote counselee is in coupling phase, the expected proportion of offspring with phenotypes $D\text{-}M_1M_1$, ddM_1M_1, $D\text{-}M_1M_2$, and ddM_1M_2 are $(2 - r)/4$, $r/4$, $(1 + r)/4$, and $(1 - r)/4$, where D- represents DD or Dd. On the other hand, if the counselee is in repulsion phase, they become $(1 + r)/4$, $(1 - r)/4$, $(2 - r)/4$, and $r/4$, respectively. Therefore, the situation is identical with that when dominant markers were linked to autosomal dominant disease genes, and if we denote by n_1, n_2, n_3, and n_4 the numbers of $D\text{-}M_1M_1$, ddM_1M_1, $D\text{-}M_1M_2$, and ddM_1M_2, respectively, formula (11) can be used for computing the probability of linkage phase. However, the risks of having genetic diseases are not the same; they are given by

$$R(M_1M_1) = [rP(C) + (1 - r)P(R)]/2 , \quad (14a)$$

$$R(M_1M_2) = [(1 - r)P(C) + rP(R)]/2$$

$$= 0.5 - R(M_1M_1) . \quad (14b)$$

The informative family $DdM_1M_2 \times DdM_2M_2$ can be studied in exactly the same manner, and formulas (11) and (14) are directly applicable. However, in this case, the risks in formulas (14a) and (14b) correspond to those of offspring genotypes M_1M_2 and M_2M_2, respectively.

Table 4 gives the values of $P(C)$ and $R(M_1M_1)$ for $\alpha \geq 0$ and β. It is clear that the effects of α and β on $P(C)$ and $R(M_1M_1)$ are very different. α represents the contribution to $P(C)$ and hence to $R(M_1M_1)$ of the offspring who are affected and whose genotypes are completely known. On the other hand, β is the contribution of phenotypically normal children, and there is uncertainty in the genotypes of the normal children; these individuals cannot be classified as recombinants or nonrecombinants. Hence, in the case of $\alpha = 0$, the accuracy with which one can identify the linkage phase of the counselee is rather low. In the case of $\alpha \geq 1$, however, the accuracy is very high if r is small. In this table, the accuracy of prediction of the disease status is given by $1 - R(M_1M_1)$ and is greater than 98% for $\alpha \geq 1$ and

TABLE 4

PROBABILITIES OF COUPLING PHASE [$P(C)$] AND THE RISKS FOR AN OFFSPRING WITH MARKER GENOTYPE M_1M_1 TO DEVELOP A RECESSIVE GENETIC DISEASE [$R(M_1M_1)$]

		r = .05		r = .005		r = .001	
α	β	$P(C)$	$R(M_1M_1)$	$P(C)$	$R(M_1M_1)$	$P(C)$	$R(M_1M_1)$
0	−3	.1350	.4142	.1134	.4414	.1116	.4438
	−2	.2248	.3739	.2024	.3973	.2005	.3995
	−1	.3500	.3175	.3350	.3317	.3337	.3330
	0	.5000	.2500	.5000	.2500	.5000	.2500
	1	.6500	.1825	.6650	.1683	.6663	.1670
	2	.7752	.1261	.7976	.1027	.7995	.1005
	3	.8650	.0858	.8867	.0586	.8884	.0562
1	−3	.7479	.1385	.9622	.0212	.9921	.0044
	−2	.8464	.0941	.9806	.0121	.9960	.0025
	−1	.9110	.0651	.9901	.0074	.9980	.0015
	0	.9500	.0475	.9950	.0050	.9990	.0010
	1	.9724	.0374	.9975	.0038	.9995	.0008
	2	.9850	.0318	.9987	.0031	.9997	.0006
	3	.9919	.0287	.9994	.0028	.9996	.0006
2	−3	.9826	.0328	.9998	.0026	1.0000	.0005
	−2	.9905	.0293	.9999	.0025	1.0000	.0005
	−1	.9949	.0273	1.0000	.0025	1.0000	.0005
	0	.9972	.0262	1.0000	.0025	1.0000	.0005
	1	.9985	.0257	1.0000	.0025	1.0000	.0005
	2	.9992	.0254	1.0000	.0025	1.0000	.0005
	3	1.0000	.0252	1.0000	.0025	1.0000	.0005

NOTE: The parental mating considered is $DdM_1M_2 \times DdM_1M_1$. M_1 and M_2 are codominant marker alleles and r is the recombination value.

$r \leq .005$. Therefore, if r is small, both the linkage phase and the risk for M_1M_1 to contract the disease can be determined quite accurately. Unfortunately, however, if the fetus or child under consideration has genotype M_1M_2, it is difficult to increase the accuracy of risk estimation except in the case of $R(M_1M_1)$ close to .5.

In the case of mating $DdM_1M_2 \times DdM_1M_2$, both parents are double heterozygotes, so that we must determine the probability of linkage phase for both parents. There are three possibilities. (1) Both parents are in coupling phase (CC); (2) one parent is in coupling and the other in repulsion (CR); and (3) both parents are in repulsion (RR). Since we have assumed linkage equilibrium, the prior probabilities of the linkage phases CC, CR, and RR are 1/4, 1/2, and 1/4, respectively. Let n_1, n_2, n_3, n_4, n_5, and n_6 denote the numbers of children with phenotypes ddM_1M_1, ddM_1M_2, ddM_2M_2, $D\text{-}M_1M_1$, $D\text{-}M_1M_2$, and $D\text{-}M_2M_2$, respectively, the total number of children being n. When both parents are in coupling phase, the expected frequencies of these six phenotypes are $r^2/4$, $r(1 - r)/2$, $(1 - r)^2/4$, $(1 - r^2)/4$, $[1 - r(1 - r)]/2$, and $[1 - (1 - r)^2]/4$, respectively. Therefore, if a child has genotype M_1M_1, his or her risk of having genetic disease is r^2, which is very small when r is small. If a child has genotype M_1M_2, the risk is $r(1 - r)$, which is still small. However, if the child's genotype is M_2M_2, the risk is $(1 - r)^2$, and thus marker genes are very useful for predicting the risk.

Let us now consider the method of estimating the linkage phases of counselees. The prior probability of obtaining a given set of offspring phenotypes is given by

$$P_{CC} = \rho_1^{n_1}\rho_2^{n_2}\rho_3^{n_3}(1 - \rho_1)^{n_4}(1 - \rho_2)^{n_5}(1 - \rho_3)^{n_6}/2^{2n-n_2-n_5} , \qquad (15a)$$

where $\rho_1 = r^2$, $\rho_2 = r(1 - r)$, and $\rho_3 = (1 - r)^2$.

Similarly, the prior probabilities of obtaining the same set of offspring phenotypes when the two parents are in different linkage phases (CR) and when they are both in repulsion phase (RR) are

$$P_{CR} = \rho_2^{n_1+n_3}(\rho_1 + \rho_3)^{n_2}(1 - \rho_2)^{n_4+n_6}(1 + 2\rho_2)^{n_5}/4^n \qquad (15b)$$

and

$$P_{RR} = \rho_1^{n_3}\rho_2^{n_2}\rho_3^{n_1}(1 - \rho_1)^{n_6}(1 - \rho_2)^{n_5}(1 - \rho_3)^{n_4}/2^{2n-n_2-n_5} , \qquad (15c)$$

respectively.

Let $P = P_{CC} + 2P_{CR} + P_{RR}$. From the prior probabilities of CC, CR, and RR, we then have

$$P(CC) = P_{CC}/P = \sigma_1/(2 + \sigma_1 + \sigma_2) , \qquad (16a)$$

$$P(CR) = 2P_{CR}/P = 2/(2 + \sigma_1 + \sigma_2) , \qquad (16b)$$

$$P(RR) = P_{RR}/P = \sigma_2/(2 + \sigma_1 + \sigma_2) , \qquad (16c)$$

where σ_1, 2, and σ_2 are the relative probabilities of CC, CR, and RR, respectively. σ_1 and σ_2 are given by

$$\sigma_1 = \theta_1^{n_1-n_3}\theta_2^{n_2}\theta_4^{-n_4}\theta_5^{-n_5}\theta_6^{n_6} , \tag{17a}$$

$$\sigma_2 = \theta_1^{n_3-n_1}\theta_2^{n_2}\theta_4^{-n_6}\theta_5^{-n_5}\theta_6^{n_4} , \tag{17b}$$

where $\theta_1 = \rho_1/\rho_2 = \rho_2/\rho_3$, $\theta_2 = 2\rho_2/(\rho_1 + \rho_3)$, $\theta_4 = (1 - \rho_2)/(1 - \rho_1)$, $\theta_5 = (1 + 2p_2)/[2(1 - \rho_2)]$, and $\theta_6 = (1 - \rho_3)/(1 - \rho_2)$. The formulas (16a–c) determine the probabilities of linkage phase of the counselees, and for any given values of n_1 through n_6, the posterior probabilities are always different from the prior probabilities. In other words, the use of linked marker genes will always be useful for genetic counseling.

When the linkage phases of the counselees are estimated, the risk for a child with a given marker genotype to develop the disease is given by

$$R(M_1M_1) = \rho_1P(CC) + \rho_2P(CR) + \rho_3P(RR) , \tag{18a}$$

$$R(M_1M_2) = \rho_2[P(CC) + P(RR)] + (\rho_1 + \rho_3)P(CR)/2 , \tag{18b}$$

$$R(M_2M_2) = \rho_3P(CC) + \rho_2P(CR) + \rho_1P(RR) . \tag{18c}$$

It is difficult to compute the probability of linkage phase and risk under general conditions. For the autosomal recessive diseases linked to codominant markers, affected offspring are more informative than are normal ones because their genotypes are completely known. Thus, for illustrative purposes, we consider only affected children. Table 5 gives the values of the probabilities of linkage phases and risks for developing an autosomal recessive disease for the case of $\alpha = n_2$ and $\beta = n_3 - n_1$. When the recombination value, r, is small and α is nonzero, the linkage phases of the parents can be determined with a high probability. The accuracy of prediction of disease status in table 5 is the value of $R(\cdot)$ or $1 - R(\cdot)$, whichever is larger, and is generally very high. Table 5 is given for illustrative purposes, and for a given family, the probabilities of linkage phase and risk should be determined by using information on all offspring.

Let us now consider the proportion of informative families. We denote by x_1 and x_2 the frequencies of marker alleles M_1 and M_2, respectively. The informative families are those families in which both husband and wife are a carrier of the disease gene d, and at least one of them has marker genotype M_1M_2; the spouse may have any marker genotype. Therefore, if the husband and wife are both known to be a carrier of the disease gene, the proportion of informative families is

$$\begin{aligned} I &= 4x_1^3x_2 + 4x_1^2x_2^2 + 4x_1x_2^3 \\ &= 4x_1x_2(1 - x_1x_2) . \end{aligned} \tag{19}$$

TABLE 5

PROBABILITIES OF LINKAGE PHASES AND RISKS FOR DEVELOPING AN AUTOSOMAL RECESSIVE DISEASE WHEN CODOMINANT MARKERS ARE USED

		r = .05						r = .005					
α	β	$P(CC)$	$P(CR)$	$P(RR)$	$R(M_1M_1)$	$R(M_1M_2)$	$R(M_2M_2)$	$P(CC)$	$P(CR)$	$P(RR)$	$R(M_1M_1)$	$R(M_1M_2)$	$R(M_2M_2)$
0	0	.2500	.5000	.2500	.2500	.2500	.2500	.2500	.5000	.2500	.2500	.2500	.2500
	1	.9025	.0950	.0025	.0090	.0860	.8190	.9900	.1000	.0000	.0001	.0099	.9802
	2	.9945	.0055	.0000	.0028	.0497	.8978	1.0000	.0000	.0000	.0000	.0050	.9900
	3	.9997	.0003	.0000	.0025	.0476	.9023	1.0000	.0000	.0000	.0000	.0050	.9900
1	0	.0475	.9050	.0475	.0860	.4140	.0860	.0050	.9900	.0050	.0099	.4901	.0099
	1	.4986	.5000	.0014	.0262	.2500	.4738	.5000	.5000	.0000	.0025	.0025	.4975
	2	.9499	.0501	.0000	.0048	.0678	.8596	.9950	.0050	.0000	.0001	.0074	.9851
	3	.9972	.0028	.0000	.0026	.0486	.9001	1.0000	.0000	.0000	.0000	.0050	.9900
2	0	.0055	.9891	.0055	.0519	.4481	.0519	.0001	.9999	.0001	.0050	.4950	.0050
	1	.0947	.9050	.0003	.0435	.4140	.1285	.0099	.9901	.0000	.0049	.4901	.0148
	2	.6654	.3346	.0000	.0176	.1830	.6164	.6667	.3333	.0000	.0017	.1683	.6617
	3	.9742	.0258	.0000	.0037	.0579	.8805	.9975	.0025	.0000	.0000	.0062	.9876

NOTE: The parental mating considered is $DdM_1M_2 \times DdM_1M_2$. M_1 and M_2 are codominant marker alleles and r is the recombination value. $\alpha = n_2$; $\beta = n_3 - n_1$.

This I has a maximum value of .75 when $x_1 = x_2 = .5$.

In the presence of multiple codominant alleles, the only uninformative matings are $M_iM_i \times M_iM_i$ and $M_iM_i \times M_jM_j$ $(i \neq j)$ with frequencies x_i^4 and $x_i^2x_j^2$, respectively, where x_i is the frequency of the ith allele. Therefore, the proportion of informative families is given by

$$I = 1 - \sum_i x_i^4 - \sum_{i \neq j} x_i^2 x_j^2$$

$$= 1 - \left(\sum_{i=1}^{m} x_i^2\right)^2 . \tag{20}$$

This has a maximum of $I = 1 - (1/m)^2$ when all allele frequencies are equal $(x_i = 1/m)$ and takes the value .75, .89, .94, .96, and .97 when $m = 2, 3, 4, 5$, and 6, respectively. Thus, the utility of marker genes for detecting autosomal recessive diseases is very high.

Dominant Markers

Consider a dominant marker locus with two alleles M and m, where m is recessive to M. In this case, a family will be informative when both parents are a carrier of the disease gene d, and at least one of them is heterozygous (Mm) at the marker locus. With dominance at the marker locus, the genotype Mm will not be obvious unless there is at least one mm offspring in the family. This means that the spouse of the doubly heterozygous counselee must have marker genotype Mm or mm.

Let us consider the informative family $DdMm \times Ddmm$. As mentioned above, this type of family must have at least one mm child. Then, the genotypes of both parents can be inferred, and the situation is identical with that of the mating $DdM_1M_2 \times DdM_1M_1$ with codominant markers. Therefore, formulas (11) and (14) can be used to estimate the probability of linkage phase and risk. Some representative values of these are presented in table 4.

In the mating $DdMm \times DdMm$, both parents are double heterozygotes, and thus one has to estimate the probability of linkage phase in both of them. Let n_1, n_2, n_3, and n_4 be the numbers of offspring with phenotypes *ddmm*, *ddM−*, *D-mm*, and *D-M-*, respectively. We then have

$$P_{CC} = \rho_3^{n_1}(1 - \rho_3)^{n_2+n_3}(2 + \rho_3)^{n_4}/4^n , \tag{21a}$$

$$P_{CR} = \rho_2^{n_1}(1 - \rho_2)^{n_2+n_3}(2 + \rho_2)^{n_4}/4^n , \tag{21b}$$

$$P_{RR} = \rho_1^{n_1}(1 - \rho_1)^{n_2+n_3}(2 + \rho_1)^{n_4}/4^n , \tag{21c}$$

where $n = n_1 + n_2 + n_3 + n_4$, $\rho_1 = r^2$, $\rho_1 = r(1 - r)$, and $\rho_3 = (1 - r)^2$. Let $P = P_{CC} + 2P_{CR} + P_{RR}$. As before, the prior probabilities of CC, CR, and RR are given by 1/4, 1/2, and 1/4, respectively. Thus, the posterior probabilities,

$P(CC)$, $P(CR)$, and $P(RR)$, are given by formulas (16a–c) with σ_1 and σ_2 redefined as

$$\sigma_1 = \theta_1^{-n_1}\theta_2^{n_2+n_3}\theta_3^{-n_4}$$

$$\sigma_2 = \theta_1^{n_1}\theta_4^{-(n_2+n_3)}\theta_5^{n_4} ,$$

where $\theta_1 = \rho_2/\rho_3$, $\theta_2 = (1 - \rho_3)/(1 - \rho_2)$, $\theta_3 = (2 + \rho_2)/(2 + \rho_3)$, $\theta_4 = (1 - \rho_2)/(1 - \rho_1)$, and $\theta_5 = (2 + \rho_1)/(2 + \rho_2)$.

The risks of genetic disease for children with recessive and dominant marker phenotypes are given by

$$R(m) = \rho_3 P(CC) + \rho_2 P(CR) + \rho_1 P(RR), \tag{22a}$$

$$R(M) = [1 - R(m)]/3 . \tag{22b}$$

Table 6 gives the values of $P(CC)$, $P(CR)$, $P(RR)$, and the risk of genetic disease for sibship sizes 0, 1, 2, and 3. Each sibship of size n can be partitioned into n_1, $n_2 + n_3$, and n_4 so that the above calculations can be made for all phenotypic combinations. It is clear that $R(m)$ and $R(M)$ depend on the values of n_1, $n_2 + n_3$, and n_4. If $n_1 \geq 1$ and $n_2 = n_3 = n_4 = 0$, $R(M)$ is very close to the recombination value, that is, the value of $R(M)$ when the linkage phase is determined with certainty. Actually, if $n_1 > n_2 + n_3 + n_4$, $R(M)$ is close to the recombination value. When $n_2 + n_3 > 1$ and $n_1 = n_4 = 0$, $R(m)$ shows a very small value. When $n_4 > 1$ and $n_1 + n_2 + n_3 = 0$, however, marker genes are not very useful for genetic counseling.

To compute the proportion of informative families, we denote by x the frequency of the recessive allele m. Note that the informative families are similar to those obtained for autosomal dominant disease genes linked to dominant markers, but in this case the counselee need not be heterozygous for the marker locus if his or her spouse is. Therefore, the proportion of informative families for a given value of n is

$$I = 4x^2(1 - x)^2\left[1 - \left(\frac{3}{4}\right)^n\right] + 4x^3(1 - x)\left[1 - \left(\frac{1}{2}\right)^n\right] . \tag{23}$$

This is slightly higher than that for the case of dominant genetic diseases.

NUMERICAL EXAMPLE

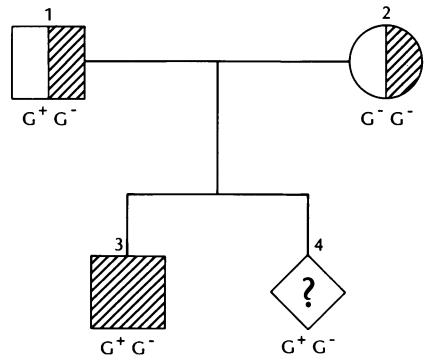
To illustrate how to use the formulas developed here, we consider the example given in figure 1. This figure shows a family with a sickle-cell anemia son [14]. His parents are obviously heterozygous for the sickle-cell anemia gene (S). The parents want to know whether their second child (fetus) will develop the disease or not. Phillips et al. [14] examined the *Hin* dIII restriction site polymorphism in the hemoglobin $(Hb)^G\gamma$ gene, which is closely linked to the sickle-cell anemia gene

TABLE 6

PROBABILITIES OF LINKAGE PHASES AND RISKS FOR DEVELOPING AUTOSOMAL RECESSIVE DISEASES WHEN DOMINANT MARKERS ARE USED

				r = .05					r = .005				
n	n_1	$n_2 + n_3$	n_4	$P(CC)$	$P(CR)$	$P(RR)$	$R(m)$	$R(M)$	$P(CC)$	$P(CR)$	$P(RR)$	$R(m)$	$R(M)$
0	0	0	0	.2500	.5000	.2500	.2500	.2500	.2500	.5000	.2500	.2500	.2500
1	0	0	1	.3225	.4550	.2225	.3132	.2289	.3322	.4456	.2222	.3311	.2230
	0	1	0	.0325	.6350	.3325	.0603	.3132	.0033	.6634	.3333	.0066	.3311
	1	0	0	.9025	.0950	.0025	.8190	.0603	.9900	.0100	.0000	.9802	.0066
2	0	0	2	.4047	.4027	.1926	.3848	.2051	.4261	.3832	.1907	.4238	.1921
	0	2	0	.0034	.6437	.3530	.0345	.3218	.0000	.6644	.3355	.0033	.3322
	2	0	0	.9945	.0055	.0000	.8978	.0341	.9999	.0001	.0000	.9900	.0033
	0	1	1	.0458	.6310	.3232	.0721	.3093	.0050	.6628	.3322	.0082	.3306
	1	0	1	.9292	.0690	.0018	.8419	.0527	.9933	.0067	.0000	.9834	.0055
	1	1	0	.4862	.5000	.0138	.4626	.1791	.4987	.5000	.0013	.4963	.1679
3	0	0	3	.4925	.3458	.1617	.4613	.1796	.5257	.3170	.1573	.5220	.1593
	0	3	0	.0003	.6350	.3647	.0314	.3229	.0000	.6634	.3367	.0033	.3323
	3	0	0	.9997	.0003	.0000	.9023	.0326	1.0000	.0000	.0000	.9900	.0033
	0	1	2	.0641	.6236	.3123	.0883	.3039	.0074	.6617	.3309	.0106	.3298
	0	2	1	.0048	.6478	.3474	.0360	.3213	.0001	.6650	.3350	.0034	.3322
	1	0	2	.9490	.0497	.0013	.8589	.0470	.9955	.0045	.0000	.9856	.0048
	1	2	0	.0882	.8862	.0256	.1218	.2927	.0099	.9876	.0025	.0147	.3284
	2	0	1	.9961	.0039	.0000	.8992	.0036	1.0000	.0000	.0000	.9900	.0033
	2	1	2	.9486	.0513	.0000	.8585	.0472	.9950	.0050	.0000	.9851	.0050
	1	1	1	.5731	.4157	.0112	.5370	.1543	.5974	.4016	.0010	.5934	.1355

NOTE: The parental mating considered is $DdMm \times DdMm$. M and m are marker alleles with m being recessive to M, and r is the recombination value.



LEGEND:

□/○	NORMAL MALE/FEMALE (AA)
◪	CARRIER MALE (AS)
▨	AFFECTED MALE (SS)
◇?	PREGNANCY

FIG. 1.—Family with a sickle-cell anemia son (*3*). His parents are both heterozygous for the sickle-cell gene *S*. *Individual 4* is a prospective child (fetus).

(*Hbβ* gene), and found that the genotypes of the male and female parents are G^+G^- and G^-G^-, respectively, with respect to this polymorphism, whereas the genotype of their son is G^+G^-. Since both parents are carriers of the sickle-cell anemia gene (*AS*), the genotypes of the mother and the son are clearly AG^-/SG^- and SG^+/SG^-, respectively. This case represents our autosomal recessive disease with codominant markers, and alleles A, S, G^-, and G^+ correspond to our previous notations D, d, M_1, and M_2, respectively. Since the genotype of their previous child is $SS\ G^-G^+$ (ddM_1M_2), $n_1 = n_2 = n_3 = 0$ and $n_4 = 1$, and thus $\alpha = 1$ and $\beta = 0$. Therefore, the probability of the father having the coupling genotype AG^-/SG^+ (DM_1/dM_2) is $1 - r$ from formula (11), whereas the probability of having the repulsion genotype AG^+/SG^- (DM_2/dM_1) is r. Since the recombination value between the S gene and the G gene is apparently very small, the father's genotype is AG^-/SG^+ with a high probability.

The risk of the second child (fetus) to develop the sickle-cell anemia can be computed from formula (14). Since the marker genotype of the fetus is G^+G^-, the risk is $R(M_1M_2) = [(1 - r)^2 + r^2]/2$. Since r is practically 0, this becomes 1/2. Therefore, it is difficult to know the disease status in this case, even if we know the linkage phase of the father. If the fetus had genotype G^-G^-, then we would have predicted that the fetus is not homozygous for the S gene with a high probability. In the present family, however, Phillips et al. examined another marker locus (*Hpa* I), which indicated that the fetus with G^+G^- was a homozygote for the S gene with a high probability. Subsequently, the family decided to terminate the pregnancy.

In the present example, the computation of the probabilities of linkage phase and risk of genetic disease is quite simple, so that without our formulas they can be computed. When a family has many children with various marker genotypes, the computation is no longer simple, and our formulas will facilitate the computation to a great extent.

DISCUSSION

We have seen that information on the phenotypes of children is quite useful for genetic counseling. However, the accuracy of prediction of genetic disease depends on the recombination value and the types and number of children. Certain types of offspring genotypes increase the accuracy to a great extent, but others do not. It also depends on whether the marker genes are dominant or codominant. Generally, codominant markers are more useful than dominant markers. The proportion of informative families is also higher for codominant markers than for dominant markers, other things being equal. Therefore, it is important to use a better marker when there are several markers available for a given genetic disease.

The proportion of informative families depends on the frequency of marker alleles as well as on the number of children. Since allele frequencies vary with population, this proportion should be computed for each population separately. Nei [4, 5] has computed this proportion for a number of marker genes in several human populations when linkage phase is determined from information on grandparents. Some of his results directly apply to our case, but, in general, it is not the same as that for the case of determination of linkage phase through children.

Here we have assumed that the disease and marker loci are in linkage equilibrium. When the recombination value is small, this assumption is often violated by the effect of selection, genetic drift, or migration as well as by the uniqueness of initial mutation that occurs in a particular chromosome [15]. When there is linkage disequilibrium, however, the proportion of informative families is expected generally to increase, since it makes the association of disease gene and a certain marker gene stronger. Indeed, in the case of determination of linkage phase through grandparents, Chakravarti and Nei [16] have shown that the increase is substantial. We expect that linkage disequilibrium has a similar effect in the present case.

REFERENCES

1. Hoogvliet B: Genetische en klinische beschouwing naar aanleiding van bloederziekte en kleurenblindheit in dezelfde familie. *Genetica* 23:93–220, 1942
2. Murphy EA, Chase GA: *Principles of Genetic Counseling*. Chicago, Yearbook Medical, 1975
3. Rivas ML, Conneally PM: Application and significance of linkage in diagnosis and prevention of genetic disease, in *Genetic Counseling*, edited by Lubs HA, de la Cruz F, New York, Raven Press, 1977, pp 447–475
4. Nei M: Utility and efficiency of linked marker genes for genetic counseling, in *Proceedings 41st Session of International Statistical Institute*, New Delhi, 1977, pp 698–711
5. Nei M: Proportion of informative families for genetic counseling with linked marker genes. *Jpn J Hum Genet* 24:131–142, 1979

6. KAN YW, DOZY AM: Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci USA* 75:5631–5635, 1978
7. KAN YW, LEE KY, FURBETTA M, ANGIUS A, CAO A: Polymorphism of DNA sequence in β-globin gene region: application to prenatal diagnosis of β° thalassemia in Sardinia. *N Engl J Med* 302:185–188, 1980
8. LITTLE PFR, ANNISON G, DARLING S, WILLIAMSON R, CAMBA L, MODELL B: Model for antenatal diagnosis of β-thalassemia and other monogenic disorders by molecular analysis of linked DNA polymorphisms. *Nature* 285:144–147, 1980
9. BOTSTEIN D, WHITE RL, SKOLNICK M, DAVIS RW: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331, 1980
10. MURPHY EA, MUTALIK GS: The application of Bayesian methods in genetic counseling. *Hum Hered* 19:126–151, 1969
11. HEUCH I, LI FHF: PEDIG—A computer program for calculation of genotype probabilities using phenotype information. *Clin Genet* 3:501–504, 1972
12. CANNINGS C, SKOLNICK MH, DE NEVERS K, SRIDHARAN R: Calculation of risk factors and likelihoods for familial diseases. *Comp Biomed Res* 9:393–407, 1976
13. HASSTEDT S, CARTWRIGHT P: PAP: Pedigree analysis package, technical report #13, Dept. of Medical Biophysics and Computing, Univ. of Utah, 1979
14. PHILLIPS JA, PANNY SR, KAZAZIAN HH, BOEHM CD, SCOTT AF, SMITH KD: Prenatal diagnosis of sickle cell anemia by restriction endonuclease analysis: *Hin* dIII polymorphisms in γ-globin genes extend test applicability. *Proc Natl Acad Sci USA* 77:2853–2856, 1980
15. NEI M, LI WH: Non-random association between electromorphs and inversion chromosomes in finite populations. *Genet Res* 35:65–83, 1980
16. CHAKRAVARTI A, NEI M: The utility of linked marker genes for genetic counseling when linkage disequilibrium is present. *Am J Hum Genet* 31:121A, 1979