Statistical Models for Trisomic Phenotypes

Neil E. Lamb,[1] Eleanor Feingold,[2] and Stephanie L. Sherman[1]

Departments of [1]Genetics and Molecular Medicine and [2]Biostatistics, Emory University, Atlanta

Summary

Certain genetic disorders are rare in the general population but more common in individuals with specific trisomies, which suggests that the genes involved in the etiology of these disorders may be located on the trisomic chromosome. As with all aneuploid syndromes, however, a considerable degree of variation exists within each phenotype so that any given trait is present only among a subset of the trisomic population. We have previously presented a simple gene-dosage model to explain this phenotypic variation and developed a strategy to map genes for such traits. The mapping strategy does not depend on the simple model but works in theory under any model that predicts that affected individuals have an increased likelihood of disomic homozygosity at the trait locus. This paper explores the robustness of our mapping method by investigating what kinds of models give an expected increase in disomic homozygosity. We describe a number of basic statistical models for trisomic phenotypes. Some of these are logical extensions of standard models for disomic phenotypes, and some are more specific to trisomy. Where possible, we discuss genetic mechanisms applicable to each model. We investigate which models and which parameter values give an expected increase in disomic homozygosity in individuals with the trait. Finally, we determine the sample sizes required to identify the increased disomic homozygosity under each model. Most of the models we explore yield detectable increases in disomic homozygosity for some reasonable range of parameter values, usually corresponding to smaller trait frequencies. It therefore appears that our mapping method should be effective for a wide variety of moderately infrequent traits, even though the exact mode of inheritance is unlikely to be known.

Introduction

As with all syndromes due to aneuploidy, live-born trisomy 21, leading to Down syndrome (DS), is characterized by considerable variation within its phenotype. For example, 40% of all DS individuals are born with some type of heart defect (Ferencz et al. 1989), while 8% of the DS population exhibits some form of congenital gut abnormality (Bergsma 1979, p. 215). Even mental impairment, the only constant finding, varies in its expressivity and severity. Such variation has previously been ascribed to genetic, stochastic, and environmental factors (Epstein 1993). From a genetics standpoint, some unknown proportion of the variable expression may be caused by specific effects of different alleles at one or a few genetic loci on the trisomic chromosome. The red cell acid phosphatase gene is a classic example of an allelic effect on a quantitative system in a normal euploid environment. Various combinations of alleles of the red cell acid phosphatase gene result in different levels of total enzymatic activity (Hopkinson et al. 1963). In a trisomic cell, such differences would be magnified. Thus, it has been hypothesized that certain trisomic genotypes may lead to greater liability or susceptibility for a phenotypic trait, because of various causes, such as (1) different levels of gene regulation; (2) altered enzymatic activity; (3) altered molecular structural arrangements; (4) different physiological or metabolic responses by the body to a trisomic product; or (5) varying reactions to an environmental insult.

We have elsewhere described a method to map genes involved in the etiology of phenotypic traits that appear in only a subset of trisomic individuals (Feingold et al. 1995). We proposed that susceptible trisomic genotypes are likely to arise in cases where the two chromosomes inherited from the nondisjoining parent are partially identical, resulting in the inheritance of double copies of "susceptibility" alleles at some specific locus. For this reason, these traits are much more frequent in the trisomic population than in the population at large. Traits that behave dominantly (i.e., require only a single copy of a susceptibility allele) are not included in this model; such traits should occur with equal frequency in both the trisomic and disomic populations. Our gene-dosage model bears resemblance to previous models described by Engel (1980) and Carothers (1983). The identical chromosomal regions are

examples of disomic homozygosity, defined as homozygosity by descent of the two alleles inherited from the parent in whom the nondisjunction event occurred (i.e., the nondisjoining parent). Thus, a subset of the DS population (i.e., those affected with a specific phenotypic trait) can be screened for shared regions of disomic homozygosity to identify a candidate chromosomal region that may contain genes that are involved in the trait etiology. Recently, DS individuals with transient leukemia or acute megakaryoblastic leukemia (ANLL subtype M7) were collected and screened for increased disomic homozygosity (Shen et al. 1995). For both leukemic subgroups, levels of disomic homozygosity were found to be increased when compared with nonleukemic DS individuals. This increase was most notable in the proximal region of chromosome 21. Much of the increase was attributable to unusual numerical or structural abnormalities leading to reduction to homozygosity at all loci. Increased disomic homozygosity screenings have also been initiated for DS individuals affected with duodenal atresia; sufficient sample sizes, however, have yet to be attained (Lamb et al. 1994).

Although Feingold et al. (1995) described a simple two-allele model, the mapping method can be applied to any trait for which excess disomic homozygosity is expected. While the method does not require specification of the model, it is likely to be more or less successful according to *how much* excess disomic homozygosity can be expected.

In this paper, we explore the utility of disomic homozygosity mapping by examining which models of trait etiology yield detectable levels of excess disomic homozygosity. There is, of course, a rich literature of statistical models of phenotypic variation in disomic individuals, but, to our knowledge, there has been no systematic extension of these models to trisomic individuals. Thus, we begin by describing a number of basic statistical models for trisomic phenotypes. These include (1) a general single locus, two-allele model, (2) models that incorporate heterogeneity and environmental effects, (3) a model that takes into account the high level of selection against the trisomic fetus with the trait, and (4) a model that describes allele loss yielding disomy in a particular tissue or cellular subset of interest (mosaicism). Although the models presented are assumed to cause the presence or absence of a trait (all-or-none traits), some are also applicable to quantitative traits. We investigate which models and which parameter values for each model yield excess disomic homozygosity and whether that excess is sufficient to be detected with a reasonable sample size.

Models

Single-Locus Models

In order to investigate the power of our mapping method, we need to describe traits by associating a pene-

trance, *f*, with each genotype, G. These penetrances, along with allele frequencies, are the parameters that will determine whether a particular trait can be mapped. Various formulas for the penetrances imply different characteristics of the trait etiology, such as dominance, phenocopy prevalence, etc.

One historic approach to modeling penetrances of all-or-none traits has been to treat the penetrance itself as a quantitative character (e.g., James 1971; Suarez et al. 1978; Risch 1990) and to apply standard linear models for quantitative traits (Kempthorne 1957, p. 316). This type of model is more mathematically than genetically motivated and has been useful for genetic linkage analysis because it provides tractable expressions for penetrances that often are good approximations to important genetic models. In this approach, the model for the penetrance of a one-locus disomic trait with alleles $A_1 \ldots A_n$ is that the genotype $A_i A_j$ (where i may equal j) has penetrance

$$f_{ij} = \mu + \alpha_i + \alpha_j + d_{ij} . \qquad (1)$$

The parameter μ is the overall mean penetrance, α_i is the contribution of allele A_i to the penetrance, and d_{ij} is the interaction between alleles A_i and A_j—the "dominance" contribution. The trisomic extension of this is

$$f_{ijk} = \mu + \alpha_i + \alpha_j + \alpha_k + d_{ij} + d_{ik} + d_{jk} + r_{ijk} , \qquad (2)$$

which includes additive contributions of all three alleles, all two-way interactions, and the three-way interaction. It is clear that this model is too complex to explore in practice (at least for an arbitrary number of alleles), though the additive version of it, $f_{ijk} = m + \alpha_i + \alpha_j + \alpha_k$, may be of some interest.

An alternative modeling approach (e.g., that of Elston and Stewart [1971] and Morton and MacLean [1974]) has been to assume that each genotype is associated with a quantitative trait value, g_{ij} (or g_{ijk} in the trisomic case), that underlies the all-or-none trait. This quantitative trait value could be biological, such as a level of enzymatic activity, or clinical, such as a score on a screening test for schizophrenia; in either case, it can be modeled as an equation of the form (1) or (2), possibly with the addition of a random term representing a polygenic or environmental component. The quantitative trait value is added to an independent random (usually normally distributed) environmental effect for each individual, yielding that individual's "liability." The liability of an individual is thus a random variable, normally distributed with mean equal to g_{ijk}. Individuals whose liability exceeds a threshold, τ, are affected with the all-or-none trait. The penetrance of genotype $A_i A_j A_k$ can be written as

$$f_{ijk} = 1 - \Phi[(\tau - g_{ijk})/\sigma] , \qquad (3)$$

where Φ is the usual distribution function of a standard normal random variable and σ is the standard deviation of the environmental noise. This type of model has been very useful for segregation analysis but is somewhat less useful for linkage mapping because expressions for penetrances are not simple functions of model parameters.

For many all-or-none traits, it is sufficient to model two alleles, with A representing all "normal" alleles and B representing all "mutant" or extreme-acting alleles. With two alleles, a disomic individual can have just three genotypes, AA, AB, and BB, with penetrances f_0, f_1, and f_2. The trisomic extension has genotypes AAA, AAB, ABB, and BBB, with penetrances f_0, f_1, f_2, and f_3, respectively. The subscript on the penetrance indicates the number of B alleles in the genotype. The penetrances can be thought of as having form (2) or form (3), if desired. In our investigations of the robustness of our mapping method, we rely primarily on two-allele models. While two-allele models cannot describe all traits, they do provide enough variety to explore the power of the mapping method under a wide range of assumptions about trait etiology.

Multilocus Models

Both approaches described above, the linear model and the threshold model of penetrance, can be extended to describe effects of additional loci and specific environmental exposures. Let $G_i, i = 1, \ldots, n$, be the *genotypes* at locus 1 and $H_j, j = 1, \ldots, m$, be the *genotypes* at locus 2. In the linear model of penetrance, we define u_{ij} to be the penetrance of the genotype combination $G_i H_j$. In the threshold model, we define v_{ij} to be the quantitative trait value associated with the genotype combination $G_i H_j$. Various formulas for the values u_{ij} or v_{ij} can then be given to describe different kinds of interactions between the loci. In both models, environmental components can be modeled in the same way as genetic loci. (When looking at a single individual, an effect due to an environmental exposure is statistically indistinguishable from an effect due to an additional genetic locus, though in family studies the distinction is generally more relevant). For our purposes, it is most useful to work in the framework of the linear model of penetrance, because of the more easily interpretable penetrance formulas.

One of the most interesting cases is heterogeneity, which we define as two or more loci or environmental effects, each of which causes the trait in a separate subset of the population. (If the cause is environmental rather than genetic, one generally refers to "phenocopies" rather than to "heterogeneity," but, as mentioned above, the two are equivalent for modeling purposes). Approximate penetrance formulas under heterogeneity are straightforward, as long as the trait is rare enough that we can essentially neglect the probability of an indi-

vidual having more than one of the causative agents (genes or environmental exposures). Risch (1990) showed that a good approximation to heterogeneity is the additive model $u_{ij} = x_i + y_j$, where x_i and y_j are the marginal penetrances for the two loci. The marginal penetrances can take the form of (1) or (2), or also of (3). This additive approximation for heterogeneity has an important implication: if only one locus is being examined in the presence of heterogeneity, the penetrance of genotype G_i still has the same form as the one-locus model we initially described. The average contribution of other loci is absorbed into μ. This means that the one-locus models discussed above are appropriate for describing the effect of a single locus even for a heterogeneous trait that may be caused by other loci or environmental effects. In the trisomy case, we may also be interested in the possibility that trisomy per se adds a certain risk, independent of genotype. Again, this can be modeled in the same way, with the extra risk due to trisomy absorbed into μ, as long as the risk due to trisomy per se is small.

Models that allow for interaction between loci (or between genetic and environmental effects) are a more complex matter and have not even been thoroughly discussed in the disomic case. Risch (1990) presents a multiplicative model that describes specific kinds of interactions. A somewhat more general model is given by Dupuis et al. (1995). Extending these general models to the trisomic case would yield models that are too complex to be of much interest. However, it is certainly feasible to construct models of specific interactions; for example, the interaction between loci on chromosomes 13 and 21 in the development of Hirshprung disease that is suggested by the work of Puffenberger et al. (1994).

Fetal Death Models

The high rate of spontaneous abortion of trisomic fetuses (reviewed by Bond and Chandley [1983]) suggests that the majority have specific defects severe enough to prevent viability. Thus, it is important to examine models that explicitly incorporate fetal death associated with the defect being studied. Fetal death due to general trisomy effects is covered under the general two-allele model above. We consider a model with two thresholds: exceeding the first, lower, liability threshold represents expression of the trait; exceeding the second, higher, liability threshold represents the more severe phenotype (i.e., fetal death due to a more severe manifestation of the trait). Trisomic individuals can then be divided into three groups: unaffected live-borns, affected live-borns, and unrecognized fetal deaths. A general statistical model for this situation associates with each genotype, G_i, a probability of survival, w_i, and a penetrance of the trait, $f_i = P\{\text{affected}|\text{survival}\}$.

Allele-Loss Models

This model explores the concept of "allele loss" (i.e., the loss of one of the trisomic chromosomes among a subset of the cellular population). Such a mechanism has been discussed before (Pangalos et al. 1994) and is thought to account for 60% of chromosome 21 mosaicism. It is assumed that this chromosomal loss occurs during the early divisions of the zygote into blastomeres and happens randomly, with no preference for cell type or chromosomal origin. In this case, the statistical model must consider disomic and trisomic genotypes together, because the trisomic population consists of some individuals who are trisomic for the gene of interest and some who are disomic. For a two-allele locus, there are a total of seven genotypes to consider: AA, AB, BB, AAA, AAB, ABB, and BBB. In the most general case, each of these can be assigned a penetrance. Additionally, we define a new parameter, z, equal to the probability that an allele is lost. Thus, for example, a trisomic individual who starts out as ABB has probability $1 - z$ of staying ABB, probability $z/3$ of becoming BB, and probability $2z/3$ of becoming AB. It is mathematically useful to think of the "apparent penetrance" of an individual who starts out ABB as the appropriately weighted average of the penetrances of the genotypes that a person can become. Then this model becomes, for mathematical purposes, a special case of the general two-allele model. For example, if the locus behaves recessively, so that only BB and BBB individuals are affected, the apparent penetrances are $f_0 = (z)(0) + (1 - z)(0) = 0$ (since an AAA individual cannot become a BB or BBB); $f_1 = 0$; $f_2 = (z)(\frac{1}{3}) + (1 - z)(0) = z/3$; and $f_3 = 1$.

Methods

For a number of special cases of the models described above, we have investigated whether the trait can be mapped by the methods of Feingold et al. (1995) and for which parameter values. Table 1 lists the cases we considered for each model.

As previously mentioned, we propose that the susceptible trisomic genotypes are likely to arise in cases where the two chromosomes inherited from the nondisjoining parent are partially identical, resulting in the inheritance of double copies of "susceptibility" alleles at a specific locus. These identical chromosomes are examples of *disomic homozygosity*. Evidence of disomic homozygosity can be detected only at markers that are heterozygous in the parent in whom the nondisjunction event occurred. If the alleles contributed by that parent are homozygous at the markers in the trisomic offspring (i.e., have been reduced to homozygosity), they are homozygous by descent.

The mapping method is applicable to any trait for

Table 1

Models Examined

MODEL	PENETRANCE (P[survival]), BY GENOTYPE			
	AAA	AAB	ABB	BBB
I	f_0	f_1	f_2	f_3
IA	$3a$	$2a + b$	$a + 2b$	$3b$
IB	g	g	f	f
IC	0	0	f	f
ID	0	0	f_2	f_3
IE	g	g	g	f
IF	g	f	f	f
II	$f_0\ (w_0)$	$f_1\ (w_1)$	$f_2\ (w_2)$	$f_3\ (w_3)$
IIA	$0\ (w_0 = 1)$	$0\ (w_1 = 1)$	$f\ (w_2 = w)$	$f\ (w_3 = w)$
IIB	$0\ (w_0 = 1)$	$0\ (w_1 = 1)$	$f\ (w_2 = 1)$	$f\ (w_3 = w)$
IIC	$0\ (w_0 = 1)$	$0\ (w_1 = 1)$	$f\ (w_2 = w)$	$1\ (w_3 = 0)$
IID	$0\ (w_0 = 1)$	$0\ (w_1 = 1)$	$1\ (w_2 = w)$	$1\ (w_3 = 0)$
III	0	0	$z/3$	1

which trisomic individuals with a particular defect are expected to show greater-than-normal levels of disomic homozygosity in the chromosomal region containing the gene involved in the etiology of the defect; in other words,

$$P\{\text{homozygosity at the trait locus} \mid \text{trait}\}$$

$$> P\{\text{homozygosity at the trait locus}\}\ .$$

For brevity, we will denote homozygosity and heterozygosity at the trait locus with a minus sign $(-)$ and a plus sign $(+)$, respectively. Using Bayes theorem, this inequality can be rewritten as:

$$\frac{P\{\text{trait} \mid -\}P\{-\}}{P\{\text{trait} \mid -\}P\{-\} + P\{\text{trait} \mid +\}P\{+\}} > P\{-\}\ .$$

A bit of algebra shows that this is equivalent to

$$P\{\text{trait} \mid -\} > P\{\text{trait} \mid +\}\ .$$

Thus, our mapping method can be applied when the probability of having the defect given disomic homozygosity at the trait locus is greater than the probability of having the defect given disomic heterozygosity at the trait locus.

For example, consider model IC, a single locus model with penetrances of 0 for trisomic genotypes AAA and AAB and penetrances of f for trisomic genotypes ABB and BBB. Let the population frequencies of A and B be p and q, with $p + q = 1$. If there is reduction to homozygosity at the trait locus, the nondisjoining parent contributes either AA or BB with probabilities p and q. The

other parent contributes either A or B with probabilities p and q. Then, the offspring has the genotype AAA, AAB, ABB, or BBB with probabilities p^2, pq, pq, and q^2, respectively. The probabilities that the trait is expressed are $(p^2)(0)$, $(pq)(0)$, $(pq)(f)$, and $(q^2)(f)$. So,

$$P\{\text{trait}\,|-\} = (pq)(f) + (q^2)(f)$$
$$= qf(p + q) \qquad (4)$$
$$= qf.$$

If there is not reduction to homozygosity (i.e., at the trait, locus heterozygosity is maintained) the nondis-joining parent contributes AA, AB, or BB with probabilities p^2, $2pq$, and q^2, respectively. This gives the offspring genotypes AAA, AAB, ABB, or BBB with probabilities p^3, $3p^2q$, $3pq^2$, and q^3, respectively, with probabilities of expressing the trait of $(p^3)(0)$, $(3p^2q)(0)$, $(3pq^2)(f)$, and $(q^3)(f)$. So,

$$P\{\text{trait}\,|+\} = (3pq^2)(f) + (q^3)(f) . \qquad (5)$$

Setting up the inequality yields the following:

$$P\{\text{trait}\,|-\} > P\{\text{trait}\,|+\}$$
$$\leftrightarrow qf > (3pq^2)(f) + (q^3)(f)$$
$$\leftrightarrow q < 0.5$$

Therefore, under model IC when the frequency of the B allele is $<.5$ (i.e., $q < .5$), any value of penetrance will yield excess disomic homozygosity.

In some applications, it may be more informative to describe such results in terms of trait frequency (i.e., the frequency of trisomic individuals with the trait). The trait frequency is referred to as K, where,

$$K = P\{\text{trait}\,|-\}P\{-\} + P\{\text{trait}\,|+\}P\{+\} .$$

Elsewhere, we determined the probability of disomic homozygosity along the long arm of chromosome 21 from the empirical data, using our DS study population of >600 individuals (Feingold et al. 1995). This value varies between .2 and .3. For the following conversions, we use the approximate value $P(-) = .25$. Using model IC again as an example,

$$K = (qf)(0.25) + (3pq^2f + q^3f)(0.75) .$$

Thus, for the boundary value $q = .5$, the corresponding K value is

$$K = [(0.5)f](0.25)$$
$$+ [3(0.5)(0.5)^2(f) + (0.5)^3(f)](0.75)$$
$$= f/2 .$$

Thus, model IC predicts excess disomic homozygosity for $0 < K < f/2$.

Using this same scheme, we have examined each model described in table 1 to find the parameter values that meet these conditions. In addition, sample sizes have been calculated for each model. This was done by calculating the amount of excess disomic homozygosity that is expected under each model (see Feingold et al. 1995) and then determining the sample size necessary to detect that much excess with 80% power, using a significance level for the test of .01.

Results

Model I: General Two-Allele Models

As mentioned earlier, model I assumes a one-locus, two-allele system. Allele A contributes a low-liability product, while allele B contributes a high-liability product. The penetrance variables are labeled f_0, f_1, f_2, and f_3 for the genotypes AAA, AAB, ABB, and BBB, respectively, where each subscript identifies the number of B alleles present in the genotype. Each trisomic genotype has some probability of expressing the abnormal phenotype. In and of itself, trisomy at the trait locus may carry some small risk of affection, but this risk increases with the number of B alleles present in the genotype ($f_3 \geq f_2 \geq f_1 \geq f_0$). For example, suppose the gene of interest encodes a rate-limiting enzyme involved in a metabolic pathway. As Epstein et al. (1981) point out, the increase in enzyme levels could alter both metabolic flux and the size of the metabolite pool. Suppose allele B produces a protein with higher activity levels than the protein encoded by allele A. Disomic AB individuals will produce more total activity than AA individuals and disomic BB individuals will produce more total activity than AB individuals. Trisomic individuals will also exhibit increased activity levels with AAA < AAB < ABB < BBB. Under this model, trisomy in and of itself (i.e., the presence of three active alleles) increases the liability to develop some phenotypic defect. This liability is further increased as a result of allelic variations and interactions.

Instead of encoding an enzymatic protein, the gene of interest could produce a structural protein. For example, suppose the gene encodes protein "1," one strand of a heterotrimeric structural protein like collagen. Strands "2" and "3" are encoded by other genes. If "1" is produced by a trisomic genotype, increased levels of protein "1" could lead to altered concentrations of the normal collagen and the formation of abnormal collagen homo-

trimers consisting of three "1" chains. A similar model has been discussed in terms of collagen type VI, where two of the three chains of the heterotrimeric collagen fiber are encoded by genes on chromosome 21 (Duff et al. 1990). If allele B produces a protein strand that is more likely to form abnormal homotrimers than that produced by allele A, the level of homotrimers will increase with the number of B alleles present in the trisomic genotype. This, in turn, increases the liability of developing a phenotypic defect.

Under the assumptions of this model, equations (4) and (5) can be written:

$$P\{\text{trait} \,|\, \text{homozygosity at trait locus}\}$$
$$= p^2 f_0 + pq f_1 + pq f_2 + q^2 f_3 \,,$$

$$P\{\text{trait} \,|\, \text{heterozygosity at trait locus}\}$$
$$= p^3 f_0 + 3p^2 q f_1 + 3pq^2 f_2 + q^3 f_3 \,.$$

In its most general form, this model contains too many parameters and possible penetrance combinations to yield informative results. Thus, submodels have been examined by fixing penetrance parameters to specific values, simplifying these equations. These submodels are discussed below.

Case A.—Model IA represents the classic additive model of penetrances. It considers a trait locus with two alleles, A and B, where each A allele contributes some value a to the overall susceptibility and each B allele contributes some value b. Here, the penetrances for genotypes AAA, AAB, ABB, and BBB are as follows: $f_0 = 3a, f_1 = 2a + b, f_2 = a + 2b$, and $f_3 = 3b$, respectively. In disomic individuals, the additive model is a good approximation for rare dominant traits because the BB genotype is rare (Kempthorne 1957, p. 316), but this is not true in the trisomic case because the ABB genotype will have relatively high prevalence. Under this case, equations (4) and (5) can be written:

$$P\{\text{trait} \,|\, \text{homozygosity at trait locus}\} = 3ap + 3bq \,,$$

$$P\{\text{trait} \,|\, \text{heterozygosity at trait locus}\}$$
$$= 3ap^3 + 6ap^2 q + 3bp^2 q + 3apq^2 + 6bpq^2 + 3bq^3 \,.$$

These two expressions can be shown to be equal. Therefore, a trait with purely additive penetrances cannot be mapped by looking for excess disomic homozygosity. With a bit of algebra, this result can be extended for an arbitrary number of alleles with additive penetrances.

Case B.—Under model IB, only two penetrance values are used; g, the probability of affection given genotypes AAA and AAB (i.e., $g = f_0 = f_1$), and f, the probability
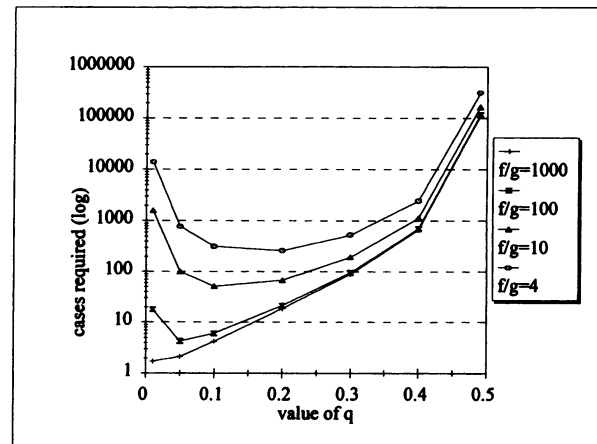

Figure 1   Sample sizes for model IB

of affection given genotypes ABB and BBB (i.e., $f = f_2 = f_3$). It is assumed that $f > g$. The above genetic scenarios can be used as examples. Each disomic and trisomic genotype produces a range of enzyme activity levels or concentration of homotrimers. As mentioned earlier, this can also be visualized as a range of cellular responses to a fixed enzyme level or homotrimer concentration for each genotype. Like the general model, case B assumes that simple trisomy at the trait locus carries some risk of affection, g. This risk does not, however, incrementally increase with each additional B allele. Instead the penetrance increases from g to f when the genotype contains a majority of the higher activity or greater "self-affinity" B alleles. The parameter g can also include risk due to some other locus or environmental effect, thus modeling heterogeneity. For this case, equations (4) and (5) are:

$$P\{\text{trait} \,|\, \text{homozygosity at trait locus}\} = qf + pg \,,$$

$$P\{\text{trait} \,|\, \text{heterozygosity at trait locus}\}$$
$$= p^3 g + 3p^2 qg + 3pq^2 f + q^3 f \,.$$

This model gives excess disomic homozygosity for all values of $q < .5$. This is equivalent to $g \leq K \leq (f + g)/2$. Figure 1 shows the sample sizes for model IB. Sample sizes remain <200 when f is at least 10 times greater than g and as long as $.05 \leq q < .3$. Sample sizes are <100 when either f is at least 100 times greater than g and $.05 \leq q \leq .2$.

Case C.—Case C assumes that there is no risk for affection for genotypes AAA and AAB ($f_0 = f_1 = 0$). It is further assumed that the penetrances of genotypes ABB and BBB are equal ($f_2 = f_3 = f$). In our previously published paper (Feingold et al. 1995), a simplified version of this case was presented, where $f = 1$. In terms

of the genetic examples previously discussed, a moderate increase in enzyme activity or a relatively low concentration of homotrimers presents no risk of affection to the individual. If however, the enzyme activity is greatly increased, or high levels or homotrimers are present, the possibility exists that the phenotype will be present. Under the parameters of this case, equations (4) and (5) are:

$$P\{\text{trait}\,|\,\text{homozygosity at trait locus}\} = qf,$$

$$P\{\text{trait}\,|\,\text{heterozygosity at trait locus}\} = 3pq^2f + q^3f.$$

Here, as in case B, excess disomic homozygosity is given for all values of $q < .5$. This corresponds to $0 \leq K \leq f/2$. Sample size, shown in figure 2, is dependent only on values of q and not on the values of K or f. Sample sizes are <100 when $q \leq .3$. In general, any mapping method that examines only affected individuals will erase a single penetrance parameter (or, equivalently, trait frequency) from the power calculation for a given sample size. Rather, these factors play a role in determining *how difficult* it is to collect a sample of the desired size. For example, if $q = .3$, the trait frequency is 28% if $f = 1$ but drops to 18% when $f = .5$.

Case D.—Case D assumes that the liabilities of genotypes AAA and AAB fall below the threshold. Thus the penetrance $f_0 = f_1 = 0$. In this case, individuals with the ABB genotype are less likely to be affected than their BBB counterparts (i.e., $f_3 > f_2$). When our two examples are used, low to moderate activity or homotrimer levels fall within the normal range and are tolerated by the system. Higher levels of activity or homotrimers can lead to some developmental or metabolic disorders. At this point, the higher the activity or homotrimer concentra-
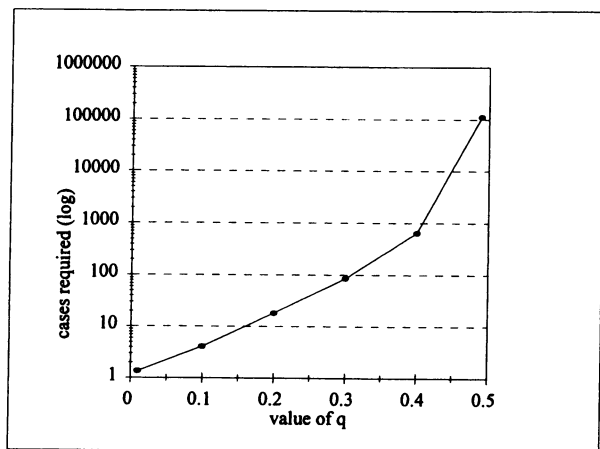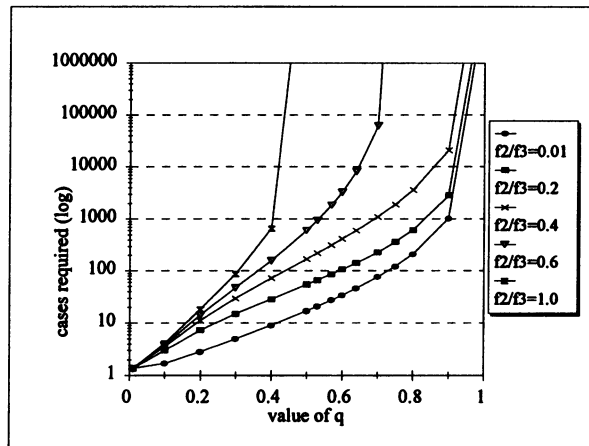

Figure 3   Sample sizes for model ID

tion, the greater the liability and the more likely the disorder will appear. Under this model, equations (4) and (5) give:

$$P\{\text{trait}\,|\,\text{homozygosity at trait locus}\} = pqf_2 + q^2f_3,$$

$$P\{\text{trait}\,|\,\text{heterozygosity at trait locus}\} = 3pq^2f_2 + q^3f_3.$$

Excess disomic homozygosity depends on the f_2/f_3 ratio. When $0 \leq f_2/f_3 \leq .5$, excess disomic homozygosity is predicted for all values of $0 \leq q \leq 1$. This corresponds to $0 \leq K \leq f_3$. When $.5 \leq f_2/f_3 \leq 1$, excess disomic homozygosity is predicted for $0 \leq q \leq [f_2/f_3]/[3(f_2/f_3) - 1]$, which corresponds to $0 \leq K \leq [4(f_2)^3/(f_3)^2]/[(3(f_2/f_3) - 1)]^2$. The sample sizes for this model are shown in figure 3. Sample size is dependent on q and f_2/f_3. Sizes are generally <100 when $f_2/f_3 \leq .2$ and $q \leq .6$ or when $q \leq .3$.

Case E.—Case E predicts that the A allele confers some "protective" effect against trait expression. Although in and of itself trisomy holds some small probability of affection ($f_0 = f_1 = f_2 = g$), as long as the trisomic genotype contains at least one A allele, the risk of affection remains low. Individuals who possess no A alleles at the trait locus (BBB) have a higher probability of affection, f. As mentioned above, g could also represent heterogeneity. This is, in many respects, the trisomic extension of a recessive trait. Under this case disomic BB individuals have a probability of affection $f - g$. If g represents heterogeneity, these individuals would have a probability of affection f. Equations (4) and (5) can be written:

$$P\{\text{trait}\,|\,\text{homozygosity at trait locus}\}$$
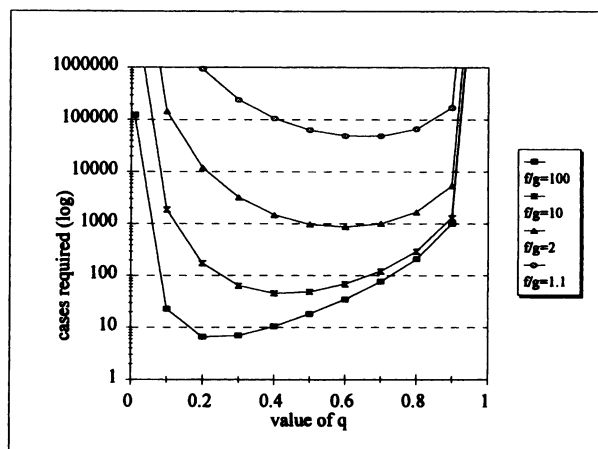
$$= p^2g + pqg + pqg + q^2f,$$

Figure 2   Sample sizes for model IC

Figure 4   Sample sizes for model IE

P{trait | heterozygosity at trait locus}

$$= p^3g + 3p^2qg + 3pq^2g + q^3f.$$

Here, all values of q ($0 \leq q \leq 1$) give excess disomic homozygosity. This corresponds to all $g \leq K \leq f$. The sample sizes for model IE are shown in figure 4. As in model IB, sample size depends on the f/g ratio as well as on q. Sample sizes are <100 when f is at least 10 times greater than g and q is roughly between .2 and .7.

Case F.—Case F presents the converse of case E (i.e., the trisomic extension of a dominant trait). The B allele is dominant to A. AAA individuals have some low probability of affection, g, because of trisomy per se or heterogeneity. The other genotypes have a greater probability of affection, f, because of the presence of at least one B allele. Thus, $f_0 = g$, and $f_1 = f_2 = f_3 = f$. As in the previous case, a subset of the disomic population may also express the trait. The AB and BB disomic individuals have probability $f - g$ of affection, or simply f, if g represents heterogeneity. Equations (4) and (5) can be written:

P{trait | homozygosity at trait locus}

$$= p^2g + pqf + pqf + q^2f,$$

P{trait | heterozygosity at trait locus}

$$= p^3g + 3p^2qf + 3pq^2f + q^3f.$$

Under this case, no values of q will yield excess homozygosity. Thus, a trait of this type cannot be mapped by our methods. As previously discussed, this result is expected, because inheriting an extra copy of a dominant "mutant" allele via disomic homozygosity would not increase the likelihood of developing the trait.

Model II: The Fetal Death Model

Model II also describes a single-locus, two-allele system but incorporates two distinct activity thresholds. Genotypes with liabilities below the first threshold do not express the abnormal trait, genotypes with liabilities between the two thresholds are phenotypically affected, and genotypes with liabilities above the second threshold do not survive to birth. Therefore, in addition to a penetrance parameter, each genotype is assigned a survival parameter, w ($0 < w < 1$) defined as the probability of survival given that genotype. Following our earlier notation, the survival parameters are labeled w_0, w_1, w_2, and w_3, for genotypes AAA, AAB, ABB, and BBB, respectively. Under this general model, equations (4) and (5) can be expressed as:

P{trait | homozygosity at trait locus}

$$= p^2f_0w_0 + pqf_1w_1 + pqf_2w_2 + q^2f_3w_3,$$

P{trait | heterozygosity at trait locus}

$$= p^3f_0w_0 + 3p^2qf_1w_1 + 3pq^2f_2w_2 + q^3f_3w_3.$$

We have examined four specific cases. The genetic scenarios described previously are still applicable, with the addition of the second threshold resulting in death during the fetal stage.

Case A.—Under model IIA, individuals with genotypes AAA and AAB are phenotypically normal ($f_0 = f_1 = 0$, and $w_0 = w_1 = 1$). ABB and BBB individuals, however, display both reduced penetrance ($f_2 = f_3 = f$) and reduced survival ($w_2 = w_3 = w$). Thus, some proportion of ABB and BBB individuals have liabilities that exceed both the affection and survival thresholds. When these parameters are used, equations (4) and (5) yield:

P{trait | homozygosity at trait locus}

$$= pqfw + q^2fw,$$

P{trait | heterozygosity at trait locus}

$$= 3pq^2fw + q^3fw.$$

and all values of $q < .5$ predict excess disomic homozygosity. This corresponds to all values of $K \leq (fw)/2$. Here also, sample sizes are dependent solely on q (fig. 5). Sample sizes remain <100 when $q \leq .3$.

Case B.—Under case B, AAA and AAB individuals are phenotypically normal ($f_0 = f_1 = 0$) and have survival parameters equal to 1 ($w_0 = w_1 = 1$). ABB individuals are at risk for affection ($f_2 = f$) but always survive ($w_2 = 1$). It is assumed that the mean liability of genotype

BBB is higher than that for ABB but that the same proportion of individuals are affected given survival (i.e., $f_2 = f_3 = f$). The increased mean liability of BBB is reflected by some proportion of affected individuals exceeding the second threshold and therefore not surviving ($w_3 = w$). Under these assumptions, equations (4) and (5) simplify to:

$$P\{\text{trait}\,|\,\text{homozygosity at trait locus}\} = pqf + q^2fw ,$$

$$P\{\text{trait}\,|\,\text{heterozygosity at trait locus}\} = 3pq^2f + q^3fw .$$

All values of q between 0 and $1/(3 - w)$ give excess disomic homozygosity. This is equivalent to all values of K between 0 and $(2f)/(3 - w)^2$. The sample sizes for model IIB are shown in figure 6. Sample sizes are <100 when $q < .25$. While $q < .3$, w has little effect on sample
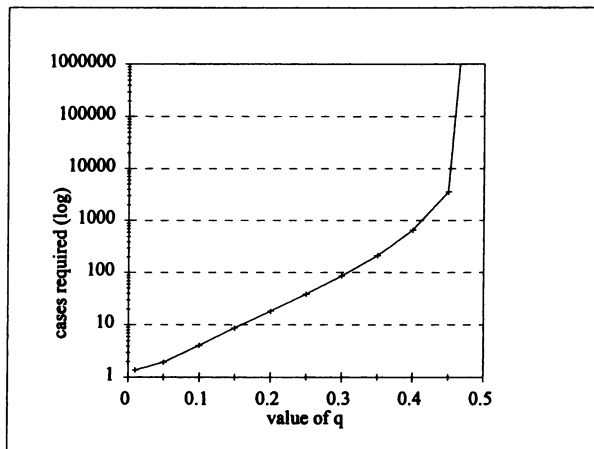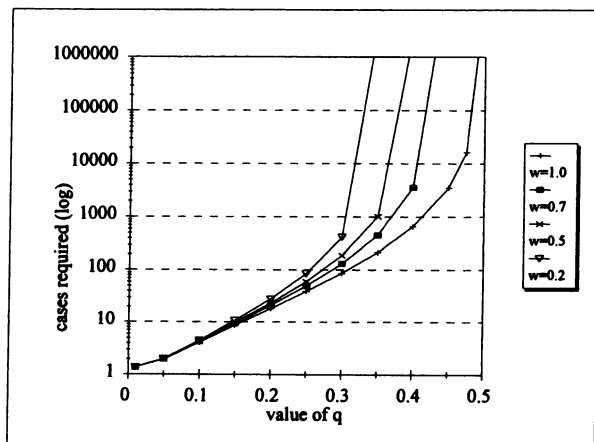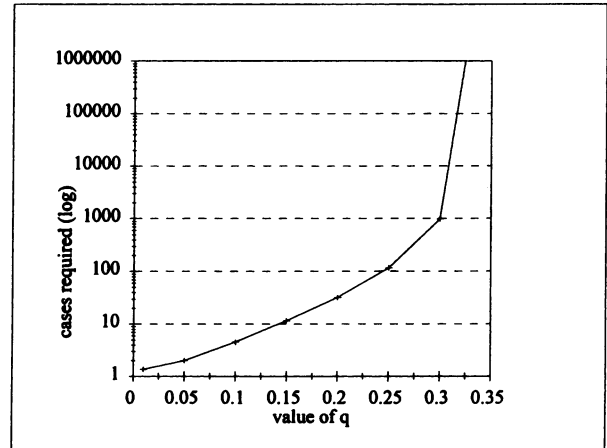

Figure 7   Sample sizes for model IIC

size. This effect does, however, increase as q increases to .5.

Case C.—As in the previous cases, AAA and AAB individuals are phenotypically unaffected ($f_0 = f_1 = 0$; $w_0 = w_1 = 1$). ABB individuals exhibit both reduced penetrance ($f_2 = f$) and reduced survival ($w_2 = w$). BBB individuals are fully penetrant ($f_3 = 1$), and do not survive ($w_3 = 0$). Thus, BBB individuals are never observed in the live-birth sample. Under this model, equations (4) and (5) are

$$P\{\text{trait}\,|\,\text{homozygosity at trait locus}\} = pqfw ,$$

$$P\{\text{trait}\,|\,\text{heterozygosity at trait locus}\} = 3pq^2fw .$$

All values of $q < .333$ predict excess disomic homozygosity. This is equivalent to $K \le (^2/_9)(fw)$. Sample sizes, shown in figure 7, depend on only q and remain <200 when $q \le .25$.

Case D.—Case D differs from case C in that the penetrance of ABB is complete ($f_2 = 1$) and, as before, only some proportion survive ($w_2 = w$). As in case C, $f_0 = f_1 = 0$, $w_0 = w_1 = 1$, and BBB individuals are not observed in the live birth population ($w_3 = 0$).Equations (4) and (5) can be rewritten as:

$$P\{\text{trait}\,|\,\text{homozygosity at trait locus}\} = pqw ,$$

$$P\{\text{trait}\,|\,\text{heterozygosity at trait locus}\} = 3pq^2w .$$

As in model IIC, all values of $q < .333$ predict excess disomic homozygosity. This is equivalent to $K \le (^2/_9)(fw)$. The results for sample size are also identical to those for model IIC, because they are independent of penetrance f (fig. 7).
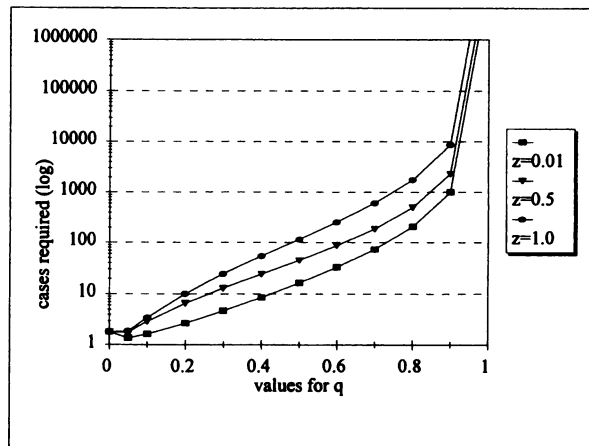

Figure 5   Sample sizes for model IIA

Figure 6   Sample sizes for model IIB

Figure 8   Sample sizes for model III

Model III: The Allele-Loss Model

As previously discussed, model III examines the concept of chromosome loss during early embryonic development or in some tissue type. Thus, this model examines the consequences of allele loss resulting in a disomic recessive genotype. We have chosen to examine an example where the trait locus behaves recessively so that only BB and BBB individuals are affected. The apparent penetrances under this model are $f_0 = 0$, $f_1 = 0$, $f_2 = z/3$, and $f_3 = 1$, where z is the probability of losing any allele. It is interesting to note that, although genetically distinct, model III is mathematically equivalent to model ID, with $f_2 = z/3$ and $f_3 = 1$. Equations (4) and (5) can be written for this model as:

$$P\{trait|homozygosity\ at\ trait\ locus\} = q^2 + \tfrac{1}{3}(zpq)\ ,$$

$$P\{trait|heterozygosity\ at\ trait\ locus\} = q^3 + zpq^2\ .$$

All values of q ($0 < q < 1$) give excess disomic homozygosity. This corresponds to $0 \leqslant K \leqslant 1$. The sample size is dependent on q and z and is shown in figure 8. As a general rule, sample size remains <100 when $q \leqslant .5$. In addition, lower values of z correspond to smaller sample sizes.

Discussion

The trisomic mapping method to detect a susceptibility gene involved in the etiology of a specific trait among trisomic individuals can be effective, in theory, for any trait, when the following condition exists: the probability of having the trait given disomic homozygosity at the susceptibility locus is greater than the probability of having the trait given disomic heterozygosity at the susceptibility locus. We have explored the robustness of the mapping method by investigating the kinds of models that give this expected increase in disomic homozygosity. We have examined several cases of a general two-allele model, incorporating various heterogeneity and environmental effects. We have also described a model that examines fetal death among trisomic individuals as well as a model that explores the development of cellular mosaicism. In each case, we have determined the relevant parameters that yield excess disomic homozygosity. These parameters were expressed in terms of both allelic and trait frequency. In addition, we have identified the sample sizes necessary to detect increased disomic homozygosity under each model. The models under which our method does not work at all are those of additive penetrance (IA) and dominance of the susceptibility allele (IF), essentially models in which the additional B allele gained by disomic homozygosity does not appreciably increase the risk for affection. The remaining models we have examined, however, have predicted excess disomic homozygosity for some subset of parameter values, generally corresponding to lower trait frequencies. Thus, it should be feasible to map any moderately infrequent trait, without knowing the specific trait etiology, as long as it is believed that a second or third copy of a "susceptibility" allele contributes an increment to the total risk that is greater than the increment contributed by the first copy. In the appendix we describe this condition more rigorously, although it is more theoretically interesting than genetically practical, since it is unlikely such precise information about trait etiology would be known during the early stages of trait mapping.

One area for future work is the examination of the robustness of the linkage parameter, y, used to estimate the distance from any marker to the trait locus. For the purposes of this paper, the trait locus and the marker being studied were assumed to be identical ($y = 0$). It will be interesting to incorporate other values of y into future model calculations as well as examine the effects of model misspecification on the outcome of y. An additional area for future study concerns examining more complex models of heterogeneity and determining their usefulness in homozygosity-by-descent mapping.

Acknowledgments

Appendix

We begin with the general inequality $P\{trait|-\}$ $> P\{trait|+\}$ as described under model I:

$$p^2 f_0 + pq f_1 + pq f_2 + q^2 f_3$$

$$> p^3 f_0 + 3p^2 q f_1 + 3pq^2 f_2 + q^3 f_3 .$$

The penetrances can be rewritten in terms of the incremental risk added by each successive B allele present in the genotype. Some base risk value, $f_0 = f$ exists for the AAA genotype. The presence of a single copy of the susceptibility allele (AAB) increases the risk of affection by some value a (i.e., $f_1 = f + a$). A second copy of the susceptibility allele adds an additional risk of b, so that $f_2 = f + a + b$. In the same manner, the third copy of the susceptibility allele increases the risk by some value c (i.e., $f_3 = f + a + b + c$). Substitution of these values into the inequality yields:

$$p^2 f + pq(f + a) + pq(f + a + b)$$

$$+ q^2(f + a + b + c) > p^3 f + 3p^2 q(f + a)$$

$$+ 3pq^2(f + a + b) + q^3(f + a + b + c) ,$$

which can be simplified to

$$(b - a) + q(a - 2b + c) > 0.$$

Using this, general conditions can be identified for a, b, and c that are necessary and sufficient to yield a true inequality for various ranges of q. We have examined three such ranges, $0 < q < .5$; $0 < q < 1$; and $0 < q < \varepsilon$, where ε represents an arbitrarily small value of q.

$0 < q < .5$

To obtain disomic homozygosity for all values of $0 < q < .5$, it is necessary and sufficient to have $b \geq a$ and $c \geq a$, with at least one of the inequalities strict (i.e., either $b > a$ or $c > a$). Rewriting this in terms of the penetrance parameters used throughout the paper yields the following: $f_2 - f_1 \geq f_1 - f_0$ and $f_2 - f_1 \geq f_1 - f_0$, with at least one of the inequalities strict.

$0 < q < 1$

To observe excess disomic homozygosity for all $0 < q < 1$, one of the following conditions must be met:

1) $b = a$ and $c > a$,
2) $b > a$ and $a - 2b + c \geq 0$, or
3) $b > a$ and $a - 2b + c < 0$ and $c \geq b$.

Substituting into the previous penetrance parameters yields the following:

1) $f_2 - f_1 = f_1 - f_0$ and $f_3 - f_2 > f_1 - f_0$
2) $f_2 - f_1 > f_1 - f_0$ and $3f_1 - f_0 - 3f_2 + f_3 \geq 0$

3) $f_2 - f_1 > f_1 - f_0$ and $3f_1 - f_0 - 3f_2 + f_3 < 0$ and $f_3 - f_2 \geq f_2 - f_1$.

$0 < q < \varepsilon$

To identify excess disomic homozygosity for some small value of q, one of the following conditions must be true:

1) $b = a$ and $c > a$, and
2) $b > a$.

This is equivalent to:

1) $f_2 - f_1 = f_1 - f_0$ and $f_3 - f_2 > f_1 - f_0$, and
2) $f_2 - f_1 > f_1 - f_0$.

Model II can be examined in a similar manner; however the addition of w, the survival parameter, greatly increases the complexity of the equations. For this reason, the model is not discussed here. As mentioned earlier, model III is mathematically equivalent to model ID with $a = 0$, $b = z/3$, and $c = (2z)/3$. So, $b > a$, $c > b$, and $a - 2b + c = 0$, meeting the conditions for $0 < q < 1$.

References

Bergsma D (1979) Birth defects compendium. Alan R Liss, New York

Bond DJ, Chandley AC (1983) Aneuploidy. Oxford University Press, Oxford

Carothers AD (1983) Gene dosage effects in trisomy: comment on a recent article by BL Shapiro. Am J Med Genet 16:635–637

Duff K, Williamson R, Richards S-J (1990) Expression of genes encoding two chains of the collagen type VI molecule during fetal heart development. Int J Cardiol 27:128–129

Dupuis J, Brown PO, Siegmund D (1995) Statistical methods for linkage analysis of complex traits from high resolution maps of identity by descent. Genetics 140:843–856

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Engel E (1980) A new genetic concept: uniparental disomy and its potential effect. Am J Med Genet 6:137–143

Epstein CJ (1993) The conceptual bases for the phenotypic mapping of conditions resulting from aneuploidy. In: Epstein CJ (ed) The phenotypic mapping of Down syndrome and other aneuploid conditions. Wiley-Liss, New York, pp 1–18

Epstein CJ, Epstein LB, Cox DR, Weil J (1981) Functional implications of gene dosage effects in trisomy 21. In: Burgio GR, Fraccaro M, Tiepolo L, Wolf U (eds) Trisomy 21: an international symposium. Springer, Berlin, pp 155–172

Feingold E, Lamb NE, Sherman SL (1995) Methods for genetic linkage analysis using trisomies. Am J Hum Genet 56:475–483

Ferencz C, Neill CA, Boughman JA, Rubin JD, Brenner JI, Perry LW (1989) Congenital cardiovascular malformations

associated with chromosome abnormalities: an epidemiologic study. J Pediatr 114:79–86

Hopkinson DA, Spencer N, Harris H (1963) Red cell acid phosphatase variants: a new human polymorphism. Nature 199:969–971

James JW (1971) Frequency in relatives for an all-or-none trait. Ann Hum Genet 35:47–49

Kempthorne O (1957) An introduction to genetic statistics. Harper & Row, New York

Lamb NE, Feingold E, Sherman SL (1994) Increased disomic homozygosity in the telomeric region of chromosome 21 among Down syndrome individuals with duodenal atresia. Abstract presented at the 44th Annual Meeting of the American Society of Human Genetics, Montreal, Quebec, October 18–22

Morton NE, MacLean CJ (1974) Analysis of family resemblance. III. Complex segregation of quantitative traits. Am J Hum Genet 26:489–503

Pangalos C, Avramopoulos D, Blouin J-L, Raoul O, deBlois M-C, Prieur M, Schinzel AA, et al (1994) Understanding the mechanism(s) of mosaic trisomy 21 by using DNA polymorphism analysis. Am J Hum Genet 54: 473–481

Puffenberger EG, Kauffamn ER, Bolk S, Matise TC, Washington SS, Angrist M, Wesiienbach J, et al (1994) Identity-by-descent and association mapping of a recessive gene for Hirschprung disease on human chromosome 13q22. Hum Mol Genet 3:1217–1225

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222–228

Shen JJ, Williams BJ, Zipursky A, Doyle J, Sherman SL, Jacobs PA, Shugar AL, et al (1995) Cytogenetic and molecular studies of Down syndrome individuals with leukemia. Am J Hum Genet 56:915–925

Suarez BK, Rice J, Reich T (1978) The generalized sib pair IBD distribution: its use in the detection of linkage. Ann Hum Genet 42:87–94