Am J Hum Genet 26:207–222, 1974

Gonadal Mosaicism and Genetic Counseling for X-linked Recessive Lethals

EDMOND A. MURPHY,[1] DANIEL W. CRAMER,[2] RICHARD J. KRYSCIO,[2]
CHARLES C. BROWN,[2] AND EDWARD R. PIERCE[1]

INTRODUCTION

The appearance of an X-linked recessive lethal disorder in a male may mean that the mother is a carrier for the disorder or that the son is a new mutation. If the mother is a carrier, either because she inherited the mutant gene from one of her parents or because she underwent a mutation very early in ontogeny, then all her cells will carry the mutant gene, and, when the final reductional division occurs in her ovaries, 50% of her ova will be of mutant type and 50% will be wild type. If we suppose that the two kinds of gametes are equally likely to be fertilized the probability is 50% that any future son will be affected. If the affected son represents a new mutation, however, the recurrence risk for the next son is the mutation rate.

In genetic counseling it is customary to reduce the problem to these two alternatives [1–3]. There are intermediate possibilities, however, and the question arises as to how much inaccuracy is introduced by ignoring them.

It is not at once clear what is meant by saying that the son "represents a new mutation." It is believed that a new mutation may arise principally whenever copying of DNA occurs. If the error occurs in the production of the oocyte we have what corresponds to the classical notion of a mutation (i.e., that the progeny are different from the cells of the parent). But if the copying error occurs early in one of the stem-lines from which the ova are derived, then a greater or lesser proportion of ova derived from it will be affected. If such a partial gonadal mosaicism exists, that is, something less than 50% of the ova are mutant, then from the standpoint of counseling risks must be assessed by estimating the proportion of mutants. This result may be achieved by maximum likelihood; but because of the paucity of data within a sibship it is desirable to incorporate prior probabilities, especially where there is no information about the ancestors or collaterals of the mother.

In this paper a mathematical model will be devised to estimate the degree of mosaicism. It will be shown how, by using Bayesian techniques, this model can be applied to refine the estimate of the risk to the next son.

Received August 27, 1973; revised October 19, 1973.

[1] Division of Medical Genetics, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205.

[2] National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20014.

BIOLOGICAL CONSIDERATIONS

Various authors have demonstrated that oogenesis in man begins during fetal life and have attempted to enumerate germ cells in human fetal ovaries [4–6]. It is believed that the total number of diploid oocytes reaches a maximum of 5–7 million by the fifth month of fetal life but that thereafter the majority of the cells degenerate. By the age of 7, some 300,000 oocytes remain of which only the minority are ever brought to ripeness. We will assume that this degeneration of oocytes during fetal life is a random process; that the ova brought to maturity are randomly selected; and that neither mutants nor nonmutants have a selective advantage. Thus the proportion of mutant oocytes remaining at any stage is on average identical with the proportion at the 5–7-million stage.

The surviving oocytes undergo meiotic division, producing four daughter cells of which only one becomes a viable gamete. From the standpoint of the mathematical model, this meiotic division will be treated as a splitting of a diploid cell without replication of DNA. Thus, from N diploid cells, $2N$ potential gametes may arise.

The population estimate of the mutation rate *per generation of persons* (μ) is obtained by observing directly, or inferring indirectly from equilibrium arguments, the number of mutant offspring from a large normal maternal population. Although μ varies, the figure 10^{-5} is commonly supposed. But this rate is the resultant of a large number of intermediate events, any one of which may be abnormal. The mutation rate *per mitosis* (λ) would thus be a quantity such that the mean number of mutant cells agrees with μ. Thus a mean of 120 heterozygous mutant cells after 6 million cell divisions would produce 120 mutant gametes out of 12 million and thus make μ equal to 10^{-5}.

THE MATHEMATICAL MODEL

The appearance and spread of a mutant oogonium during oogenesis are in many respects similar to the propagation of a mutant bacterium in a colony. Models describing this process have been investigated previously. Luria and Delbrück [7] obtained an approximation to the mean and variance of the number of mutants in a bacterial colony when it is assumed that both mutants and nonmutants grow exponentially, while mutations of nonmutants occur randomly. Later Lea and Coulson [8] altered this model by assuming that only nonmutants grow exponentially while the mutation of nonmutants and the growth of mutants occur randomly. They present a recursive procedure for obtaining an approximation to the distribution of the number of mutants and further obtain approximations to the mean and variance of the number of mutants as a function of the total population size. In a review paper on this subject, Armitage [9] generalized the results of the foregoing investigators by obtaining general expressions for the cumulants of the number of mutants in each of these particular models. Others [10–11] discuss the more general version of the Luria-Delbrück model in which both mutants and non-mutants are assumed to grow stochastically. They derive expressions for the probability-

generating function of the number of mutants but provide no answer in closed form.

In this formulation of the problem, it will be assumed that both mutants and nonmutants divide stochastically in time; but there is little interest in time as such, merely in the numbers of the two types of progeny. Thus the process need be examined only when a new cell is produced. An exact expression for the probability distribution of the number of mutants is obtained and hence an approximate expression for the moments of the number of mutants. Similarities of these results to the findings of other investigators will be pointed out where pertinent.

The application of our model to oogenesis requires the following specific assumptions:

1. All the oogonia in the ovary are descended from a wild-type cell.
2. Mutation occurs during cell division only and with probability λ.
3. There is no back mutation; hence all mutant cells breed true.
4. Mutation in the normal X chromosome of the heterozygous mutant cell will be ignored.
5. Mutant and wild-type cells are equally likely to divide, and there is no latent period between divisions.

After the first cell division, two cells exist and two outcomes are possible. The wild-type cell may mutate during division with some small probability λ to produce one wild type and one mutant; or it may not mutate during division, with probability $(1 - \lambda)$ to produce two wild-type cells. This continues with $N + 1$ cells existing and $N + 1$ outcomes possible after the Nth cell division. Outcomes for $N = 2$ cell divisions are shown in figure 1. Since we begin with a wild type and

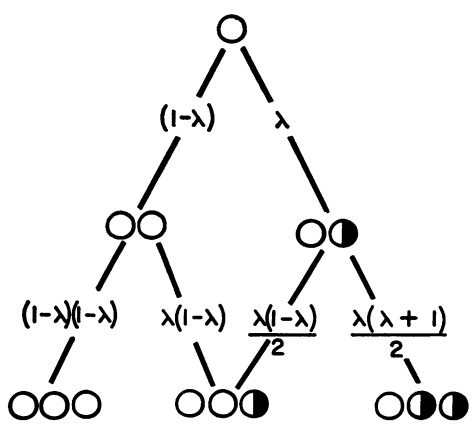

FIG. 1.—The first two steps in the proliferation of a cell line. The open circles represent wild-type cells; the black-and-white circles, heterozygous cells. The probability of a heterozygous cell being derived from a wild cell is λ. In the first division only two outcomes are possible (two wild-type cells or one wild type and one heterozygous). After two divisions there are three possible outcomes, but one of these (one heterozygous, two wild type) can be arrived at through two pathways.

since DNA is conserved, the number of mutants after the Nth cell division is at least zero but at most N.

Let $P_m(N)$, $(m = 0, 1, \ldots, N)$ denote the probability of the outcome m mutants after N cell divisions.

The outcome "zero mutants" after the Nth cell division occurs only if there exist zero mutants at the N-cell stage and no mutation occurs at the next cell division. The outcome "N mutants" after the Nth cell division occurs only if there exist $(N - 1)$ mutants at the N-cell stage and a mutation appears at the next cell division: the new mutant is produced either by the mutation of the lone wild-type cell or the division of one of the $(N - 1)$ mutant cells. At the Nth cell division, m can be attained $(0 < m < N)$ either if $(m - 1)$ mutants exist after the $(N - 1)$th cell division and a mutant appears (by mutation of a wild type during replication or by replication of a mutant) or if m mutants exist after the $(N - 1)$th cell division and the next cell is wild type. Thus:

$$P_0(N) = (1 - \lambda)P_0(N - 1),$$

$$P_m(N) = \left\{\frac{m - 1}{N} + \frac{[N - (m - 1)]\lambda}{N}\right\} P_{m-1}(N - 1) + \left[\frac{N - m}{N}(1 - \lambda)\right] P_m(N - 1), \quad (1)$$

$$m = 1, \ldots, N - 1,$$

$$P_N(N) = \left(\frac{N - 1}{N} + \frac{\lambda}{N}\right) P_{N-1}(N - 1)$$

with the initial condition $P_0(0) = 1$.

The solution to the first and last of these equations follows in a straightforward manner by induction:

$$P_0(N) = (1 - \lambda)^N,$$

$$P_N(N) = \left[\prod_{j=0}^{N-1} (j + \lambda)\right] \Big/ N!.$$

The solution for $P_m(N)$, $0 < m < N$ can be computed recursively for any given values of λ and N. However, for large N this is not practical, even on the computer. Thus, we attempt to derive expressions for these probabilities in closed form.

A Closed-Form Expression for $P_m(N)$, $0 < m < N$

For $N \geqslant 2$ and $0 < m < N$, define the nth division as that which begins with n cells and ends with $(n + 1)$ cells. Define the division in which the ith mutant cell appears as $(g_i + i - 1)$. Thus, $1 \leqslant g_1 \leqslant g_2 \leqslant \cdots \leqslant g_m \leqslant N - m + 1$ where N divisions occur producing m mutant cells.

The probability of the (g_1)th division producing the first mutant is

$$(1-\lambda)^{g_1-1}\lambda = \frac{(1-\lambda)^{g_1-1}(0+g_1\lambda)(g_1-1)!}{g_1!}. \tag{2}$$

After the (g_1)th division, wild-type cells only are produced until the (g_2+1)th division which begins with g_2 wild-type cells and one mutant cell and terminates with the appearance of a second mutant. The probability of this sequence of events is

$$\left[\frac{g_1(1-\lambda)}{g_1+1}\right]\left[\frac{(g_1+1)(1-\lambda)}{g_1+2}\right]\cdots\left[\frac{(g_2-1)(1-\lambda)}{g_2}\right]\left[\frac{(1+g_2\lambda)}{g_2+1}\right],$$

which may be simplified:

$$\frac{(1-\lambda)^{g_2-g_1}g_1!(g_2-1)!(1+g_2\lambda)}{(g_2+1)!(g_1-1)!}. \tag{3}$$

The joint probability of the events described by (2) and (3) is their product:

$$[(1-\lambda)^{g_2-1}(g_2-1)!(0+g_1\lambda)(1+g_2\lambda)]/(g_2+1)!.$$

Hence, by induction:

$$P(m|N, g_1, g_2, \ldots, g_m) = \frac{(1-\lambda)^{N-m}(N-m)!}{N!}\prod_{i=1}^{m}(i-1+g_i\lambda)$$

$$= \left[\frac{(1-\lambda)^{N-m}}{\binom{N}{m}}\right]\prod_{i=1}^{m}\left[\frac{(i-1+g_i\lambda)}{m!}\right].$$

Since $P_m(N)$ is the sum of expressions of this form over all possible values of the g, we may write

$$P_m(N) = \frac{(1-\lambda)^{N-m}}{\binom{N}{m}}H_m(N),$$

where

$$H_m(N) = \sum_{g_m=1}^{N-m+1}\sum_{g_{m-1}=1}^{g_m}\cdots\sum_{g_1=1}^{g_2}\frac{(0+\lambda g_1)(1+\lambda g_2)\ldots(m-1+\lambda g_m)}{m!}.$$

Expressing the summand in ascending powers of λ,

$$H_m(N) = \sum_{g_m=1}^{N-m+1}\sum_{g_{m-1}=1}^{g_m}\cdots\sum_{g_1=1}^{g_2}\left[\lambda g_1\frac{1}{m}+o(\lambda^2)\right]. \tag{4}$$

In the Appendix we establish results which allow us to rewrite (4):

$$H_m(N) = \frac{\lambda}{m}\binom{N+1}{m+1} + o(\lambda^2).$$

Where λ is very small, we may be prepared to discard terms containing λ^2 or higher powers. Substituting for $H_m(N)$,

$$P_m(N) = \frac{(1-\lambda)^{N-m}}{\binom{N}{m}}\lambda\frac{1}{m}\binom{N+1}{m+1} + o(\lambda^2),$$

which simplifies to

$$P_m(N) \cong \frac{(1-\lambda)^{N-m}(N+1)}{m(m+1)} \qquad \text{for } m \geqslant 1 \quad (5)$$

as an approximate closed form for $P_m(N)$.

Expressions for the Moments of the Distribution of Mutants

As we shall see in the next section, the individual probabilities $P_m(N)$ do not allow us to compute directly the risk to the next son. It transpires that the moments of the distribution are of greater utility.

If M is the random variable denoting the number of mutants after N cell divisions, then the expected value of M is

$$E(M) = \sum_{m=1}^{N} m\,P_m(N).$$

Substituting expression (5) for $P_m(N)$, we have

$$E(M) \cong \sum_{m=1}^{N} \frac{(1-\lambda)^{N-m}(N+1)}{m+1}. \quad (6)$$

Since λ is very small and N very large, $(1-\lambda) \cong 1$ and $N+1 \cong N$. Thus equation (6) simplifies to

$$E(M) \cong \sum_{m=1}^{N} \frac{N\lambda}{m+1} = N\lambda(\log N - 1 + C), \quad (7)$$

where $C =$ Euler's constant.

Expressions for the higher moments of M were derived the following way:

$$E(M^k) = \sum_{m=1}^{N} m^k P_m(N) \simeq \sum_{m=1}^{N} \frac{(1-\lambda)^{N-m}(N+1)\lambda m^{k-1}}{m+1}.$$

With the same simplifying assumptions since $m \cong m+1$, this expression becomes

$$E(M^k) \cong \sum_{m=1}^{N} N\lambda m^{k-2}.$$

Evaluating approximately by replacing the sum by an integral, we get

$$\sum_{m=1}^{N} N\lambda m^{k-2} \simeq N\lambda \int_0^N m^{k-2}dm = N\lambda \left(\frac{m^{k-1}}{k-1}\right)_0^N = \frac{N^k\lambda}{k-1}.$$

Thus,

$$E(M^k) \cong \frac{N^k\lambda}{k-1}, k \geqslant 2. \tag{8}$$

Note that the expression in equation (8) is the same as Armitage's result for the kth cumulant ($k > 1$) of the number of mutants in the Lea-Coulson model. This peculiar similarity may be due to the fact that the kth moment of M is directly proportional to N^k so that, for large values of N, the $(k-1)$th and lower moments of M are negligible in comparison. Also, in our notation, Armitage offers the quantity $N\lambda \log N$ as an approximation to the mean number of mutants. This differs from our formula (7) by the added component $N\lambda\ (C-1)$, where C is Euler's constant.

The accuracies of these approximations are explored in table 1, which compares

TABLE 1

A COMPARISON OF APPROXIMATE SOLUTION WITH EXACT SOLUTION FOR $E(M^k)$, $k = 1, \ldots, 8$, THE FIRST EIGHT MOMENTS OF THE NUMBER OF MUTANTS, WHEN $N = 150{,}000$ AND $\lambda = 1.74 \times 10^{-6}$

Moments	Approximation*	Exact
$E(M)$	3.000353	3.000345
$E(M^2)$	3.915000×10^4	3.915459×10^4
$E(M^3)$	2.936250×10^9	2.936394×10^9
$E(M^4)$	2.936250×10^{14}	2.936364×10^{14}
$E(M^5)$	3.303281×10^{19}	3.303401×10^{19}
$E(M^6)$	3.963940×10^{24}	3.964079×10^{24}
$E(M^7)$	4.954922×10^{29}	4.955101×10^{29}
$E(M^8)$	6.370614×10^{34}	6.370893×10^{34}

* The relative error in this approximation is at most one in 10,000 for all table entries.

the approximate values for the first eight raw moments of M when $N = 150{,}000$ and $\lambda = 1.74 \times 10^{-6}$ with the exact values derived recursively by computer (see Appendix). The relative error for all table entries is at most one in 10,000 even though N is only relatively small (less than 1 million).

Equation (7) may be used to estimate λ from the population estimate of the mutation rate μ. As discussed previously,

$$\frac{E(M)}{2N} = \mu \cong \frac{N(\log N - 1 + C)}{2N}.$$

Solving for λ,

$$\lambda \cong \frac{2\mu}{\log N - 1 + C}. \tag{9}$$

Table 2 shows λ values for various values of μ and N of interest.

TABLE 2

LAMBDA (λ) VALUES FOR VARIOUS POPULATION ESTIMATES OF MUTATION RATE (μ) AND FOR VARIOUS NUMBERS OF CELL DIVISIONS (N)

	Number of Cell Divisions		
μ	1 Million	6 Million	20 Million
1×10^{-4}	1.493×10^{-5}	1.317×10^{-5}	1.220×10^{-5}
5×10^{-5}	7.478×10^{-6}	6.586×10^{-6}	6.102×10^{-6}
1×10^{-5}	1.493×10^{-6}	1.317×10^{-6}	1.220×10^{-6}
5×10^{-6}	7.467×10^{-7}	6.586×10^{-7}	6.102×10^{-7}
1×10^{-6}	1.493×10^{-7}	1.317×10^{-7}	1.220×10^{-7}

APPLICATION TO GENETIC COUNSELING

When a woman without any family history has produced y sons affected by an X-linked lethal disorder and x normal sons, two hypotheses are ordinarily considered:

First, that she is a carrier for the disorder either because she inherited it from her mother or was herself the victim of a new mutation. An approximate prior probability of 4μ is associated with this hypothesis [12].

Second, that her son is a new mutant. This "hypothesis" has been treated here as a finite set of hypotheses—one of the mother's ova involved, two ova involved, three ova involved, etc.—corresponding to the various degrees of gonadal mosaicism. Each of these probabilities is multiplied by the probability that the mother is not a carrier ($1 - 4\mu$). The prior probabilities associated with each of the hypotheses are the individual probabilities $[P_m(N)]$ approximated by equation (5). The prior probabilities for each hypothesis are modified in the usual way by posterior information (i.e., in this case the mother and her descendants) to give the joint probability. The posterior probabilities for each hypothesis are determined by norming, and from them the posterior risk is calculated by multiplying by the conditional risk to the next son and summing. This conditional risk is $1/2$ under the hypothesis that the mother is a carrier and approximately $m/2N$ for the hypothesis that the mother is not a carrier and m ova are involved.

Given that the mother's ova in reproductive life are randomly selected from a finite population with a fixed number affected during her ontogeny, the number

of mutants follows, strictly speaking, the hypergeometric probability distribution rather than the binominal used here. Use of this binominal approximation will give slightly increased weight to the hypothesis that the mother is not a carrier. With large values of N this increase is negligible.

Table 3 demonstrates the intermediate steps in developing an expression for the posterior risk (R) to the next son:

$$R = \frac{4\mu(1/2)^{x+y+1} + (1-4\mu)\sum_{j=1}^{N} P_j(N)\left(\frac{j}{2N}\right)^{y+1}\left(1-\frac{j}{2N}\right)^{x}}{4\mu(1/2)^{x+y} + (1-4\mu\sum_{k=0}^{N} P_k(N)\left(\frac{k}{2N}\right)^{y}\left(1-\frac{k}{2N}\right)^{x}}. \tag{10}$$

Inserting the approximate values from equation (5) into expression (10) is tedious. The moments discussed can be used to evaluate expression (10) as follows.

In those terms of the form,

$$\sum_{k=0}^{N} P_k(N)\left(\frac{k}{2N}\right)^{y}\left(1-\frac{k}{2N}\right)^{x}$$

expand $[1-(k/2N)]^x$ to obtain

$$\sum_{k=0}^{N} P_k(N)\left(\frac{k}{2N}\right)^{y}\sum_{a=0}^{x}\binom{x}{a}(-1)^a\left(\frac{k}{2N}\right)^{a}.$$

Then, interchanging summation signs, we get

$$\sum_{a=0}^{x}\binom{x}{a}(-1)^a\sum_{k=0}^{N} P_k(N)\left(\frac{k}{2N}\right)^{y+a}$$

$$= \sum_{a=0}^{x}\binom{x}{a}(-1)^a(1/2)^{y+a}e_{y+a}, \tag{11}$$

where

$$e_{y+a} = \sum_{k=0}^{N} P_k(N)(k/N)^{y+a}.$$

Note that $e_{y+a} = 1/N^{y+a}\cdot E(M^{y+a})$. From equations (7) and (8), for $E(M^k)$,

$$e_0 = 1, \tag{12}$$

$$e_1 = (1nN - 1 + C), \tag{13}$$

$$e_i = \frac{\lambda}{(i-1)} \quad \text{for} \quad i \geqslant 2. \tag{14}$$

TABLE 3

BAYES THEOREM: X-LINKED LETHAL RECESSIVE UNDER THE ASSUMPTION OF GONADAL MOSAICISM

Probability	State A—Mother a Carrier	State B_i—Mother Not a Carrier But Having i Mutant Cells, $i = 0, \ldots, N$
Prior	4μ	$(1-4\mu)P_i(N)$
Conditional	$\binom{x+y}{x}\left(\frac{1}{2}\right)^{x+y}$	$\binom{x+y}{x}\left(\frac{i}{2N}\right)^{y}\left(1-\frac{i}{2N}\right)^{x}$
Joint	$4\mu\binom{x+y}{x}\left(\frac{1}{2}\right)^{x+y}$	$(1-4\mu)P_i(N)\binom{x+y}{x}\left(\frac{i}{2N}\right)^{y}\left(1-\frac{i}{2N}\right)^{x}$
Posterior	$\dfrac{4\mu\left(\frac{1}{2}\right)^{x+y}}{4\mu\left(\frac{1}{2}\right)^{x+y}+(1-4\mu)\sum_{k=1}^{N}P_k(N)\left(\frac{k}{2N}\right)^{y}\left(1-\frac{k}{2N}\right)^{x}}$	$\dfrac{(1-4\mu)P_i(N)\left(\frac{i}{2N}\right)^{y}\left(1-\frac{i}{2N}\right)^{x}}{4\mu\left(\frac{1}{2}\right)^{x+y}+(1-4\mu)\sum_{k=0}^{N}P_k(N)\left(\frac{k}{2N}\right)^{y}\left(1-\frac{k}{2N}\right)^{x}}$
Conditional risk to next son	$\frac{1}{2}$	$\frac{i}{2N}$
Posterior risk to next son	$\dfrac{4\mu\left(\frac{1}{2}\right)^{x+y+1}+(1-4\mu)\sum_{j=1}^{N}P_j(N)\left(\frac{j}{2N}\right)^{y+1}\left(1-\frac{j}{2N}\right)^{x}}{4\mu\left(\frac{1}{2}\right)^{x+y}+(1-4\mu)\sum_{k=0}^{N}P_k(N)\left(\frac{k}{2N}\right)^{y}\left(1-\frac{k}{2N}\right)^{x}}$	

NOTE.—x = number of sons not affected; y = number of sons affected; μ = mutation rate per generation; N =number of diploid cell divisions during oogenesis.

Substituting equation (11) into (10) and simplifying,

$$R = \frac{1}{2}\left[\frac{1 + \left(\frac{1-4\mu}{4\mu}\right)\sum_{j=0}^{x}\binom{x}{j}(-1)^j 2^{x-j} e_{y+j+1}}{1 + \left(\frac{1-4\mu}{4\mu}\right)\sum_{k=0}^{x}\binom{x}{k}(-1)^k 2^{x-k} e_{y+k}}\right]. \tag{15}$$

It can be shown that this is a strictly increasing function of y such that the limit as $y \to \infty$ is $1/2$, which demonstrates that the maximum risk is $1/2$, that is, the risk under the assumption that the mother is a carrier.

If into formula (15) the appropriate e values given by formulas (12)–(14) are inserted, a table of posterior risks to the next son can be derived for various pedigrees (i.e., various integer values of x and y). It is necessary to choose an N, μ, and the corresponding λ. Table 4 shows these values for $N = 6$ million, $\mu = 10^5$, and $\lambda = 1.217 \times 10^{-6}$. Table 5 shows posterior risks calculated using classic methods.

TABLE 4

POSTERIOR RISKS CALCULATED BY BAYESIAN METHOD USING MOSAICISM MODEL

No. of Affected Sons	No. of Normal Sons					
	0	1	2	3	4	5
0	.00003	.00002	.00001	.00001	.00001	.00001
1	.34432	.26675	.18665	.11971	.07288	.04392
2	.49203	.48693	.47834	.46410	.44121	.40650
3	.49730	.49597	.49387	.49053	.48517	.47662
4	.49864	.49811	.49730	.49609	.49424	.49139
5	.49918	.49891	.49853	.49797	.49716	.49596

TABLE 5

POSTERIOR RISKS CALCULATED BY BAYESIAN METHODS IN CLASSICAL MODEL

No. of Affected Sons	No. of Normal Sons					
	0	1	2	3	4	5
0	.00003	.00002	.00002	.00001	.00001	.00001
1	.33334	.25001	.16668	.10001	.05557	.02942
2	.50000	.49999	.49998	.49996	.49992	.49984
3	.50000	.50000	.50000	.50000	.50000	.50000
4	.50000	.50000	.50000	.50000	.50000	.50000
5	.50000	.50000	.50000	.50000	.50000	.50000

DISCUSSION

Hartl [13] has explored this same problem using a different set of assumptions about the propagation of germinal cells. Rather than supposing, as we have done, that the mitoses are completely time homogeneous so that the whole problem reduces itself to analyzing the Lea-Coulson process, he supposes that multiplication is a regular synchronous division so that after i generations any particular cell will have 2^i descendants. The resulting algebra is considerably easier.

Neither model is completely plausible: to suppose that there is no latent interval between divisions, as we have done, is at variance with the facts [14]—ideally some kind of a "displaced" Poisson process might be appropriate with a zero probability density of division during an interval following the last division. On the other hand, to suppose that there is no, or at most a trivial, variance in generation time among cells of any generation (which Hartl's model implies) does not seem plausible either. Doubtless, the truth lies somewhere in between. It is therefore of some significance that the calculations from these two models, which represent, as it were, the extremes of synchrony and asynchrony, lead to the same general conclusions.

A comparison of tables 3 and 4 shows the effects on the risks of recurrence. The largest differences appear for one and two affected sons, especially where there are many normal sons. With one affected son, the risks under the mosaicism model are slightly greater than those under the classical model. As the number of normal sons increases the weight of the risk from the carrier state decreases while the weight of the risk from gonadal mosaicism is apparently much less influenced. Thus a sibship containing five normal and one abnormal son is compatible with a low degree of gonadal mosaicism which predicts a risk one and one-half times as great as that calculated under the classic assumptions. On the other hand, the relative risk calculated where two sons are affected is less if allowance is made for mosaicism. In the classical model, the stigma of two affected sons "wears off" more slowly, and the likelihood of the mother being a carrier is little offset by normal offspring. But in the mosaicism model a family with two affected sons is compatible with mosaiscism, and the likelihood that the mother is heterozygous does not gain much support, especially if she has several normal offspring. In general, the greater the plausibility of mosaicism, the less likely the hypothesis of the carrier state becomes. And, in general, the maximum risk of 1/2 is approached more slowly in the mosaicism model.

This model may offer tools applicable to other areas of genetics. Equation (9) offers a method-of-moments estimator of the probability of mutation per mitosis (λ) when the mutation rate per generation (μ) is known. Obviously this same formula may be used to give better predictions of μ should a more rational means of calculating λ become available.

While we have applied this model to describe oogenesis, with modification it is perhaps also applicable to spermatogenesis, which differs from oogenesis in that the number of divisions of spermatogonia is very large. Hence, other things

being equal, there is more opportunity for mutation, particularly as the male grows older. The model we have described permits a graphic representation of this paternal age effect [3].

For an arbitrary λ (say half that predicted in the female), the mean proportion of mutants for increasing N is calculated (fig. 2). The mean number of mutants

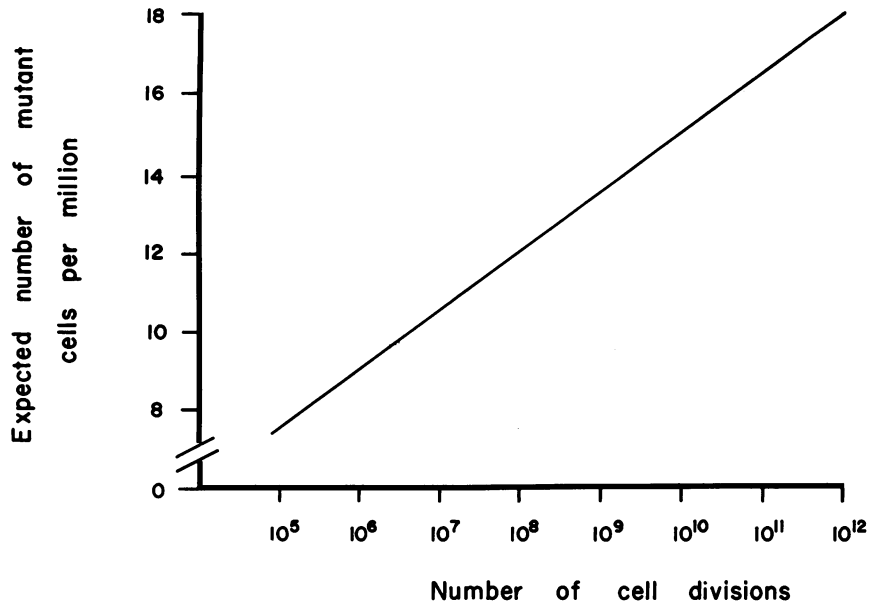

FIG. 2.—When the process shown in fig. 1 is carried through an arbitrary number of cell divisions (on the abscissa) with $\lambda = 0.6085 \times 10^{-6}$, the expected (average) number of heterozygous cells per million (on the ordinate) can be computed by formula (7). The increase is nearly linear over this range which is applicable to spermatogenesis.

steadily increases as shown. Thus a model which allows a parametric approach to the paternal age effect from information on the number of spermatogonial divisions and λ becomes available.

SUMMARY

In this paper a first-order refinement of the theory of genetic counseling for X-linked recessive disorders is presented. Biological data suggest that some 6 million oocytes are produced during the fetal life of the female. A model describing the random appearance and spread of mutant oocytes—gonadal mosaicism—as a branching process is derived. With no selection, the expected proportion of mutant oocytes remaining at any time is identical with the proportion of mutants existing at the 6-million stage. A method-of-moments estimator of the probability of a particular proportion of mutants being produced in terms of the number of cell divisions and the probability of a mutation occurring during mitosis (λ) is derived in terms of the mutation rate per generation [eq. (9)]. Approximations

to the moments of the distribution of mutants were obtained from the formula for the individual probabilities, as was a method for inserting these values of the moments into an expression describing the risk to the next son. Partial gonadal mosaicism has a small effect only on the classical risk of recurrence.

From the clinical standpoint the main implication of the paper is that it provides reassurance that, in any realistic size of family, ignoring the effect of gonadal mosaicism will have little effect on the estimate of the risk for the next child.

The application of the model to other areas, particularly spermatogenesis, is briefly discussed.

ACKNOWLEDGMENTS

We wish to thank Miss C. Bartkowiak and Mrs. Daniel W. Cramer for their help in the preparation of this manuscript.

APPENDIX

In this Appendix we will establish a method for deriving the moments of the distribution by a recursive procedure that gives exact results. Also, we present the details of the derivation of the closed-form expression for $P_m(N)$. We begin with the former.

Recursive formula for exact value of the moments. Let the random variable M_N represent the number of mutants after N cell divisions and let $\mu_{i,N} = E(M_N{}^i)$ be the ith moment of M_N. By the definition of conditional expectation, from formula (1),

$$\mu_{1,N} = E(M_N) = E[E(M_N|M_{N-1})]$$

$$= E\left\{\left(\frac{N - M_{N-1}}{N}\right)(1-\lambda)M_{N-1} + \left[\left(\frac{N - M_{N-1}}{N}\right)\lambda + \frac{M_{N-1}}{N}\right](M_{N-1}+1)\right\}$$

$$= E\left\{\lambda + M_{N-1}\left[1 + \frac{(1-\lambda)}{N}\right]\right\}$$

$$= \lambda + \left[1 + \frac{(1-\lambda)}{N}\right]\mu_{1,N-1}.$$

In a similar manner,

$$\mu_{2,N} = E(M_N{}^2) = E[E(M_N{}^2|M_{N-1})]$$

$$= E\left\{\left(\frac{N - M_{N-1}}{N}\right)(1-\lambda)M_{N-1}{}^2 + \left[\left(\frac{N - M_{N-1}}{N}\right)\lambda + \frac{M_{N-1}}{N}\right](M_{N-1}+1)^2\right\}$$

$$= E\left\{\lambda + \left[1 + \frac{2(1-\lambda)}{N}\right] M_{N-1}{}^2 + M_{N-1}\left[2\lambda + \frac{(1-\lambda)}{N}\right]\right\}$$

$$= \lambda + \left[1 + \frac{2(1-\lambda)}{N}\right]\mu_{2,N-1} + \left[2\lambda + \frac{(1-\lambda)}{N}\right]\mu_{1,N-1}.$$

In general,

$$\mu_{i,N} = \lambda + \mu_{i,N-1}\left[1 + \frac{i(1-\lambda)}{N}\right] + \sum_{j=1}^{i-1}\left[\lambda\binom{i}{j} + \frac{(1-\lambda)}{N}\binom{i}{j-1}\right]\mu_{j,N-1}.$$

Derivation of a closed form for $P_m(N)$.—The derivation of expression (5) required the simplification of the expression

$$H_m(N) = \sum_{g_m=1}^{N-m+1}\sum_{g_{m-1}=1}^{g_m}\cdots\sum_{g_1=1}^{g_2}\left[\lambda g_1\frac{1}{m} + o(\lambda^2)\right].$$

We may rewrite $H_m(N)$ by factoring constants:

$$\lambda\frac{1}{m}\left[\sum_{g_m=1}^{N-m+1}\sum_{g_{m-1}=1}^{g_m}\cdots\sum_{g_1=1}^{g_2} g_1\right] + o(\lambda^2).$$

By repeated application of the standard identity, verifiable by induction,

$$\sum_{j=1}^{n}\binom{j+m-1}{m} = \binom{m+n}{m+1};$$

the term in parentheses becomes

$$\binom{N+1}{m+1}$$

as we illustrate below:

$$\sum_{g_m=1}^{N-m+1}\cdots\sum_{g_2=1}^{g_3}\sum_{g_1=1}^{g_2}\binom{g_1}{1} = \sum_{g_m=1}^{N-m+1}\cdots\sum_{g_2=1}^{g_3}\binom{g_2+1}{2} = \sum_{g_m=1}^{N-m+1}\cdots\sum_{g_3=1}^{g_4}\binom{g_3+2}{3}$$

and so on until the final summand

$$\sum_{g_m=1}^{N-m+1}\binom{g_m+m-1}{m} = \binom{N+1}{m+1}$$

Substituting

$$\binom{N+1}{m+1}$$

for the term in parentheses, this reduces to the desired result:

$$H_m(N) = \lambda \frac{1}{m}\binom{N+1}{m+1} + o(\lambda)^2.$$

REFERENCES

1. MURPHY EA: The rationale of genetic counseling. *J Pediatr* 72:121–130, 1968
2. EMERY AEW: *Heredity, Disease and Man: Genetics in Medicine.* Berkeley, Univ. California Press, 1968
3. MCKUSICK VA: *Human Genetics,* 2d ed. Englewood Cliffs, N.J., Prentice-Hall, 1969
4. SIMKINS CS: Development of the human ovary from birth to sexual maturity. *Am J Anat* 51:465–493, 1932
5. BLOCK E: A quantitative morphological investigation of the follicular system in newborn female infants. *Acta Anat* (Basel) 17:201–206, 1953
6. BAKER TG: A quantitative and cytological study of germ cells in human ovaries. *Proc R Soc Lond* [*Biol*] 158:417–433, 1963
7. LURIA SE, DELBRÜCK M: Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491–511, 1943
8. LEA DE, COULSON CA: The distribution of the numbers of mutants in bacterial populations. *J Genet* 49:264–285, 1949
9. ARMITAGE P: The statistical theory of bacterial populations subject to mutation. *J R Stat Soc* [*B*] 14:1–33, 1952
10. KARLIN S: *A First Course in Stochastic Processes.* New York, Academic Press, 1969
11. BAILEY NT: *The Elements of Stochastic Processes with Application to the Natural Sciences.* New York, Wiley, 1964
12. MURPHY EA, MUTALIK GS: The application of Bayesian methods in genetic counseling. *Hum Hered* 19:126–151, 1969
13. HARTL DL: Recurrence risks for germinal mosaics. *Am J Hum Genet* 23:124–134, 1971
14. BLANDAU RJ: Observations on living oogonia and oocytes from human embryonic and fetal ovaries. *Am J Obstet Gynecol* 104:310–319, 1969