A Growth and Division Model for Retinoblastoma

R. Shonkwiler, [1] M. Mistretta,

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332

V. J. Varma

Veterns Administration Hospital, Atlanta, GA 30324

Abstract

A stochastic growth and division model for studying a two hit cancer is developed and applied to retinoblastoma. Retinoblastoma occurs if both genes coding for a tumor suppressor protein on homologus chromosomes become defective. Germinal cases occur when a patient or carrier, born with one defective gene, suffers a second insult to any progeny retinal cell. Somatic cases are far less likely as two hits to the same cell during development are required.

Details of the disease, germinal or somatic, unilateral or bilateral, in combination with case data allow for the estimation of the two parameters of the model: mutation rate, estimated at $p = 7 \times 10^{-7}$ per chromosome per cell division, and carrier frequency, estimated at $f = 40$ per million. The model indicates that carriers of the disease arise from similar mutations to germ cells; in particular, heridary transmission can occur for only a generation or two before dying out.

The results show that a stochastic simulation of a multi-hit cancer is feasible and may predict tumor growth dynamics. A simulation run will have to consist of a few million cells in order to observe even a small number of mutations. And several dozens such runs will have to be simulated.

Keywords: Retinoblastoma, mathematical modelling, cancer

1 Introduction

It is currently believed that cancer is a multi-stage disease which develops within a cell as the result of a succession of mutations [1]. Each mutation breaks another thread of control over the cell's latent reproductive machinery. Each mutation gives rise to a clone of similarly

[1]Corresponding author, email: shonkwiler@math.gatech.edu

mutated cells that are now ripe for further mutations. In this way the disease progresses in steps toward malignancy.

A mathematical model may be formulated for such a multi-stage progression but such a model is too complicated to be analyzed analytically with mathematical tools. However, if there were only one stage to onset of the disease, then an analytical solution is feasible. Retinoblastoma is one of the simplest cancers in that it is believed to be the result of a single mutation only.

In this article we investigate a growth and division mathematical model for retinoblastoma incorporating the possibility of a single mutation. The probability of such a mutation during cell division is a parameter of the model which we denote by p. An altered cell then gives rise to a clone of itself which we take to be a tumor. We will see that retinoblastoma incidence data is consistent with this hypothesis. What is remarkable is that so simple a model can predict the disease so well. The success of this model indicates that modeling more complicated cancers should be undertaken, perhaps via simulation.

Retinoblastoma

Retinoblastoma is among the better understood cancers and serves as a paradigm for the generic basis of the disease. Several studies indicate that this cancer is initiated if both genes, lying on homologous chromosomes, coding for a single protein become dysfunctional during the growth period of the retinal tissue. This period of time is early childhood and retinoblastoma is a disease of early childhood.

The single gene in question codes for the tumor suppressor protein Rb. Thus if this protein is absent from the cell during its growth phase, unchecked growth ensues for that cell, and for the clone of cells it begets, resulting in a tumor. Referring to the process of rendering a gene dysfunctional as a "hit," there are two distinct ways in which a cell can suffer the required two hits. The first is that some retinal cell of a normal individual suffer two hits purely by chance.

This unlikely event is thought to account for 60% of the incident rate of retinoblastoma.These are called *somatic* cases.

The other possibility is that an individual inherit a defective gene for the disease; such a person is called a *carrier*. In this event, every retinal cell has already one hit. If, during the growth phase of the retinal tissue, a mutation occurs to any cell, then that cell will have 2 hits and tumor growth ensues. These are called *germinal* cases and account for the balance of the cases, about 40%.

Previous studies

Retinoblastoma has been mathematically modeled before. In a series of papers Knudson and co-authors have undertaken this investigation, [2-4]. These purely mathematical studies lead to the discovery of tumor suppressor proteins [5]. The original model by Knudson assumes that the retinal tumor cells appear randomly selected from a static retinal tissue according to a Poisson distribution with parameter $m = 3$. (Recall the parameter of a Poisson distribution is also the mean of the distribution, hence the average number of tumors per case is 3.) Any selected cell undergoes a mutation to one of its retinal tumor suppressor genes. What is striking about the model is its simplicity, there are only two parameters, m and the total number N of retinal cells, and yet the model fits the data quite well – namely it explains the occasional gene carrier who gets no tumor, those who develop only unilateral tumors, and those who develop bilateral tumors. The parameter N is reported to be about 4 million or about 2 million retinal cells per eye [7].

Table 1

Patient data

Retinoblastoma cases as shown in Table 1, repeated from [2], are observed as to sex, onset age, laterality, that is which eye holds tumors, number of tumors and family history of the disease.

We conceptualize the case data as to "onset data," meaning data pertaining to temporal information, and "summary data" referring to the rest, mostly statistical summaries. Onset data involves diagnosis delay and adds a great complication to the data analysis. For this reason we will use the summary data (see Table 2) for the most part to validate our model. The summary data is comprised of 13 statistic none of which depend on observation time; these include carriers never affected, the average number of tumors among germinal cases, unilateral cases among germinal cases and others. These data, also reported in [2], are shown in Table 2.

Growth and Division Model

Our model differs from those of Knudson in that we track the growth of the retinal tissue from a single cell through repeated cell divisions. Each cell division is accompanied by the risk of a mutation. This approach allows for the prediction of several statistical consequences of the disease, see Table 2, including the prediction of the frequency of carriers within the general population, denoted by f, and the persistance of the disease among carriers.

In this work we impose a fixed time growth period followed by a cell division. Thus division takes place at the discrete times $t = 1, 2, \ldots, T$. As a consequence cell generations are synchronized, at any given time, all cells are of the same generation. Nevertheless this simple model fits the data quite well.

In the next section we apply the model to the summary patient data to determine model parameters and for validation. This data, derived from Table 1, is shown in Table 2. The equations of the model are derived in the Appendix.

Table 2

2 Application to the data

First we calculate the number of cell divisions T required produce the 2 million retinal cells per eye. Let $N(t)$ be the number of cells at time t. According to the model, $N(t+1) = 2N(t)$ and $N(0) = 1$. Hence

$$N(t) = 2^t. \tag{1}$$

Thus for $t = 22$, $N = 4194304$, which is approximately the required number. In the sequel we will assume there are $T = 22$ generations. Note that this number is consistent with the telomere theory of cell senescence which states that telomeres shorten upon each cell division and after, variously 22 to 40 cell divisions, the telomeres are too short to permit further divisions.

Next we estimate the mutation probability. Using (7) of the Appendix, in conjunction with statistic 1, carriers never affected, which should be about 5%, we get a value for p,

$$.05 = x_0(22) = (1-p)^{2^{22}-1}, \qquad \log(1-p) = \frac{\log(.05)}{2^{22}-1} = -.00000071423823202943,$$

and $(1-p) = .99999928576202303864$ so $p = 7.14 \times 10^{-7}$.

From this, equation (6) of the Appendix gives the expected number of tumors in the germinal case as 2.99. Note this is essentially the same as Knudson's figure.

Derived in the Appendix as equation (11), the infinite series, denoted by a for convenience, is

$$a = x_1(22) + \frac{1}{2}x_2(22) + \frac{1}{2^2}x_3(22) + \dots. \tag{2}$$

This gives the number of unilateral cases among germinal cases. Using equations (8) – (10) and substituting the values obtained above, gives statistic 3 as 0.337. Of course statistic 4, bilateral cases among germinal cases, is the complement of this at 0.663.

As introduced above, let f denote the fraction of carriers in the general population. This statistic was not derived by Knudson. Knowing f would allow us to estimate the other

statistics. Or we can use one of them to find f and then estimate the rest. We will estimate the value of f using the most reliable of the statistics which depend on it, this is 5, the fraction of cases which are bilateral; denote this fraction by B. As in Table 2, let u be the incidence rate among non-carriers, i.e. that some cell suffers two hits. And as in the Appendix, let $x_0(22)$ be the probability a carrier never contracts the disease. Therefore the incidence rate of retinoblastoma, counting both germinal and somatic cases, is given by

$$f(1 - x_0(22)) + (1 - f)u,$$

Since we are assuming 100% of the unilateral cases are somatic, statistic 13 of Table 2, we have for B

$$B = \frac{f(1-a)}{f(1 - x_0(22) + (1-f)u}.$$

In this, the expression in the numerator multiplying f is the fraction of germinal cases which are bilateral, as worked out above this is $1 - a$. Using 27% for the value of B and 30 per million for u, we get

$$f = \frac{Bu}{1 - a + Bu - B(1 - x_0(22))} = 0.000023.$$

On the other hand, the value of u may be derived from p as shown by equation (12) in the Appendix. Taking $p = 7.14 \times 10^{-7}$ predicts the incident rate of somatic cases among the general population to be $u = 0.000040$ or 40 per million. In turn this value of u gives $f = 0.0000274$. An incident rate of $u = 30$ per million is predicted when $p = 6.14 \times 10^{-7}$.

The remaining statistics are simple expressions in terms of these. The fraction of germinal cases among all cases is calculated by

$$\frac{f(1 - x_0(22))}{f(1 - x_0(22)) + (1 - f)u} = 0.387.$$

Of course the somatic cases are the complement fraction.

The fraction of all cases which are unilateral is given by the sum of the germinal and somatic ones divided by all cases,

$$\frac{fa + (1-f)u}{f(1-x_0(22)) + (1-f)u} = 0.750.$$

The fraction of unilateral germinal cases among all cases is given by

$$\frac{fa}{f(1-x_0(22)) + (1-f)u} = 0.137.$$

And finally the fraction of unilateral cases which are germinal is given by

$$\frac{fa}{fa + (1-f)u} = 0.183.$$

3 Persistence of Germinal Cases

It is natural to wonder how germinal cases arise, familial or new. In this section we show that carriers are not persistent in society in that the genetic defect lasts at most two generations normally. Therefore the condition is also the result of a chance mutation.

Let q_k be the probability that a carrier zygote will survive and beget k offspring who are also carriers and let $F(s) = q_0 + q_1 s + q_2 s^2 + \ldots$ be the probability generating function for the q_k. To compute the q_k we also need the probabilities c_i that a surviving carrier will beget i offspring. Finally let p_0 be the probability that a carrier will survive to adulthood, taken as 0.05 from Table 2.

Assuming that a carrier mates with a non-carrier, note that $\binom{i}{k}/2^i$ is the probability that a mating resulting in i offspring will consist of k carriers and $i-k$ non-carriers. Then

$$p_0 c_i \binom{i}{k} \frac{1}{2^i}$$

is the probability that a newly born carrier will survive to adulthood, have i offspring and k of them will be carriers. Sum this over $i = k, k+1, \ldots$ to get q_k,

$$q_k = p_0 \left(c_k \frac{1}{2^k} + c_{k+1} \binom{k+1}{k} \frac{1}{2^{k+1}} + c_{k+2} \binom{k+2}{k} \frac{1}{2^{k+2}} + \ldots \right), \qquad k > 0.$$

The case $k = 0$, begetting no carrier offspring, is slightly different,

$$q_0 = 1 - p_0 + p_0(c_0 + c_1\frac{1}{2} + c_2\frac{1}{2^2} + \ldots).$$

Now it is well known that a trait will persist with probability $1 - L$, and die-out with probability L, where L is the smallest fixed point of F, i.e. smallest solution of $F(L) = L$, see [6]. Since already $F(0) = q_0 > 0.95$; L must be very close to 1.

Actually, $L < 1$ if and only if the derivative $F'(1) \geq 1$. But $F'(1) = q_1 + 2q_2 + 3q_3 + \ldots$ and since each term is multiplied by $p_0 = .05$, even with extraordinary high values of the c_i for large i, $F'(1)$ will be less than 1. For example, suppose $c_5 = 1$, i.e. an average carrier has 5 children. Then $q_k = .05\binom{6}{k}/2^6$, $k = 1, \ldots, 6$ and $F'(1) = 192/1280$.

4 Conclusions

The discrete growth and division model examined in this paper adequately explains all the observed frequency and incidence data for cases of retinoblastoma. Hence a two hit model for the initiation of this cancer, with a mutation rate on the order of 7×10^{-7} per chromosome involving the critical gene, is consistent with the observed data.

The results also show that a growth and division stochastic simulation of this disease is feasible and will give good results. However, in view of the order of magnitude of the mutation rate, a simulation run will have to consist of a few million cells in order to observe even a small number of mutations. Moreover, to obtain statistical information, several dozens such runs will have to be simulated. Therefore such a simulation will entail a large computational effort. Furthermore, the same will be true to an even greater extent for simulations of more complicated cancers.

Finally, the probability of persistence of carriers through several generations is very low and hence the disease reoccurs mainly as a result of somatic cases.

5 Appendix: Derivation of the Equations

The model assumes retinal tissue starts from a single cell which, together with its progeny, repeatedly divides until the requisite number of cells are formed. For each cell, independently of the others, division takes place upon waiting 1 time unit and gives rise to two daughter cells. Thus cell generations are synchronous. We ignore cell death. The two daughter cells are genetically the same as their shared parent excepting that, with probability p, one of them has mutated. A mutant cell then begets a clone of like cells.

Germinal Case

Should a mutation occur to some retinal cell of a carrier, the result is a doubly mutated cell and hence a tumor. We are not interested in the number of such doubly mutated cells per se but rather in the number of tumors which here is the same as the number of mutation events occurring to the growing tissue. Let $X(t)$ be the random variable denoting the number of tumors at time t, and put $x_k(t) = \Pr(X(t) = k)$, $k = 0, 1, 2, \ldots$, the probability there are k tumors at time t. Also let $G(s,t) = \sum_0 x_k(t)s^k$ be the probability generating function for the x_k. The outcome at time t is influenced by what happens on the first cell division; there are two possibilities, either a mutation occurs or not. If no mutation occurs, then to have k mutations at time t, they must occur in the remaining $t-1$ cell divisions; otherwise $k-1$ mutations must occur in the remaining $t-1$ cell divisions. Let $\Pr(A \mid B)$ denote the probability of event A given that event B has occurred. If event A is influenced by two mututally exclusive events B and C, we can write, $\Pr(A) = \Pr(A \mid B)\Pr(B) + \Pr(A \mid C)\Pr(C)$. In our derivation, by conditioning on the possibilities on the first cell division, we can write

$$\begin{aligned} x_k(t) \;=\; & \Pr(k \text{ tumors at } t \mid \text{none at } 1)\Pr(\text{no mutation at } t=1) \\ & + \Pr(k \text{ tumors at } t \mid \text{mutation occurs at } t=1)\Pr(\text{a mutation at } t=1) \quad (3) \end{aligned}$$

Recall that p is the mutation rate or, in this context, the probability that a mutation occurs during a cell division. Thus $1-p$ is the probability a mutation does not occur. If a mutation does not occur during the first cell division, then the two daughter cells go on to produce cell lines and together these must result in k mutations. This could happen in many ways, for example, one could produce 3 mutations and the other $k-3$ mutations. In general, one cell line could produce k_1 mutations and the other k_2 so long as $k_1 + k_2 = k$. Therefore the first term of the equation above is given by

$$\Pr(k \text{ tumors at } t \,|\, \text{none at } 1) = (1-p) \sum_{k_1+k_2=k} x_{k_1}(t-1)x_{k_2}(t-1). \tag{4}$$

The other is simply $px_{k-1}(t-1)$. Thus substituting into (3) we get

$$x_k(t) = (1-p) \sum_{k_1+k_2=k} x_{k_1}(t-1)x_{k_2}(t-1) + px_{k-1}(t-1).$$

Multiply both sides by s^k and sum over k to get

$$\sum_{k=0}^{\infty} x_k(t)s^k = (1-p)\sum_{k=0}^{\infty}\left(\sum_{k_1+k_2=k} x_{k_1}(t-1)x_{k_2}(t-1)\right)s^k + ps\sum_{k=1}^{\infty} x_{k-1}(t-1)s^{k-1}$$

which gives

$$G(s,t) = (1-p)G^2(s,t-1) + spG(s,t-1). \tag{5}$$

(Multiply out $G(s,t-1)$ by itself to see that one gets the sum in (4).)

From the generating function we can feasibly calculate many of the properties of interest. One example is expectation. The expected number of tumors, $E(X(t))$ is given by

$$\begin{aligned} E(X(t)) &= \left.\frac{\partial G}{\partial s}\right|_{s=1} \\ &= 2(1-p)G(1,t-1)\frac{\partial G(1,t-1)}{\partial s} + pG(1,t-1) + p\frac{\partial G(1,t-1)}{\partial s}. \end{aligned}$$

Remembering that $G(1,t) = 1$ for all t gives the following recursion equation and initial value

$$E(X(t)) = (2-p)E(X(t-1)) + p, \qquad E(X(1)) = p.$$

This is easily solved to give

$$E(X(t)) = \frac{p}{1-p}\left((2-p)^t - 1\right). \tag{6}$$

We also calculate some individual probabilities.

Since $x_0(t) = G(0,t)$, substituting $s = 0$ in (5) gives the recursion equation and initial value

$$x_0(t) = (1-p)x_0^2(t-1), \qquad x_0(1) = 1-p.$$

This is the probability of being tumor free at time t and is also easily solved,

$$x_0(t) = (1-p)^{2^t-1}. \tag{7}$$

In similar fashion one obtains the probability that there will be one tumor, $x_1(t)$. From the definition of the generating function $x_1(t) = \frac{\partial G}{\partial s}|_{s=0}$. Differentiating (5), setting $s = 0$ and using (7) gives

$$\begin{aligned} x_1(t) &= 2(1-p)x_0(t-1)x_1(t-1) + px_0(t-1) \\ &= (1-p)^{2^{t-1}-1}\left[2(1-p)x_1(t-1) + p\right], \qquad x_1(1) = p. \end{aligned} \tag{8}$$

While this equation, and to a greater extent, those for the higher order probabilities which we obtain next are hard to solve in closed form, they present no difficulty numerically.

Likewise, from the fact that $\frac{\partial^n G}{\partial s^n}|_{s=0} = n!x_n$, we obtain the recursion equations below for x_2 through x_3 along with their starting values,

$$\begin{aligned} 2x_2(t) &= 2(1-p)[x_0(t-1)x_2(t-1) + x_1^2(t-1)] \\ &\quad +2px_1(t-1), \qquad x_2(1) = 0, \end{aligned} \tag{9}$$

$$\begin{aligned} 3!x_3(t) &= 2(1-p)[3!x_1(t-1)x_2(t-1) + 3!x_0(t-1)x_3(t-1)] \\ &\quad +3!px_2(t-1), \qquad x_3(1) = 0, \end{aligned} \tag{10}$$

One use of these is to calculate the number of unilateral cases among germinal cases. This is given by the infinite series

$$x_1(22) + \frac{1}{2}x_2(22) + \frac{1}{2^2}x_3(22) + \dots \tag{11}$$

evaluated at $t = 22$ since as we have seen, there are 22 cell divisions. This equation is seen as follows. First any single mutation is unilateral giving the first term. Then, with probability 1/2, two mutations will occur in the same eye giving the seond term and with probability 1/4 three will occur in the same eye and so on.

Somatic Case

Let $r_{n,0}(t) = \Pr(\text{there are } n \text{ cells with a single mutation but no tumors at time } t)$. What we want to calculate is the probability there will be no tumors and this is the sum $\sum_{n=0}^{\infty} r_{n,0}(t)$. Let $h(s,t)$, denote the probability generating function for the r.

$$h(s,t) = \sum_n r_{n,0}(t)s^n.$$

By decomposition on the possibilities at $t = 1$, we get

$$\begin{aligned} r_{n,0}(t) &= \Pr(n \text{ w/ single mut., no tumor at } t \mid \text{a mutation at 1})2p + \\ &\quad \Pr(n \text{ w/ single mut., no tumor at } t \mid \text{no mutation at 1})(1-2p) \\ &= 2px_0(t-1)r_{n-2^{t-1},0}(t-1) + (1-2p)\sum_{n_1+n_2=n} r_{n_1,0}(t-1)r_{n_2,0}(t-1). \end{aligned}$$

This is seen as follows, if a mutation occurs at $t = 1$, the resulting clone acts just like the germinal case where we learned that the probability no (additional) mutation will occur is $x_0(t-1)$. Furthermore, since this clone will produce 2^{t-1} cells each with a single mutation, the other branch must produce $n - 2^{t-1}$ single mutation cells. The other term is what we've seen before. Note that we are using $2p$ instead of p for the first mutation probability since there are two chromosomes per cell at risk. (The exact value $2p - p^2$ is well approximated by $2p$.) As usual, multiply both sides by s^n and sum over n to get

$$h(s,t) = 2px_0(t-1)s^{2^{t-1}}h(s,t-1) + (1-2p)h^2(s,t-1).$$

Note that here, the probability generating function h is not 1 at $s = 1$. Instead, $h(1,t) = \sum_{n=0}^{\infty} r_{n,0}(t)$ and this is exactly the probability there will be no doubly mutant cell by time t as decomposed over all possible ways there could be n singly mutant cells; this is what we want to calculate here. Let $u = 1 - h(1,22)$, then u is the probability of a somatic case. From above, the recursion equation for $h(1,t)$ is

$$h(1,t) = 2px_0(t-1)h(1,t-1) + (1-2p)h^2(1,t-1). \tag{12}$$

References

[1] Bert Vogelstein and Kenneth W. Kinzler, The multistep nature of cancer, Trends in Genetics 9:4 (1993) 138-141.

[2] Alfred G. Knudson, Jr., Mutation and Cancer: Statistical Study of Retinoblastoma, Proc. Nat. Acad. Sci. USA, 68:4 (1971) 820-823.

[3] Alfred G. Knudson, Jr., Herbert W. Hethcote, and Barry W. Brown, Mutation and childhood cancer: A probabilistic model for the incidence of retinoblastoma, Proc. Nat. Acad. Sci. USA 72:12 (1975) 5116-5120.

[4] Herbert W. Hethcote and Alfred G. Knudson, Jr., Model for the incidence of embryonal cancers: Application to retinoblastoma, Proc. Nat. Acad. Sci. USA 75:5 (1978) 2453-2457.

[5] Geoffrey M. Cooper, Oncogenes, 2nd ed., Jones and Bartlett Publishers, Boston, MA. 1995.

[6] Edward Yeagers, Ronald Shonkwiler, and James Herod, The Mathematics of Biology, Birkhaüser-Boston, Boston, MA 1996.

[7] Van Buren, J. M., The Retinal Ganglion Cell Layer, Charles C Thomas, Springfield, IL 1963.