Supplementary Material

Mathematical Model

Definitions and assumptions

1. Let $Y_1(t)$ denote the number of normal progenitor cells per breast at age t, γ_1 the net growth rate, μ_1 be mutation rate per cell per year from normal progenitor cell to the intermediate cell compartment I_2.

2. Let $Y_i(t)$ (i=2, …, k) denote the numbers of the I_i intermediate cells at age t. Let γ_i denote the net growth rate for the I_i compartment and let μ_i be the rate per cell per year for the ith rate limiting event. $Y_{k+1}(t)$ denotes the number of fully malignant cells for a k hit model at time t.

3. Malignant transformation of a single cell leads to the development of a malignant tumor.

4. Each breast develops from 10 stem cells (17).

There are two fundamental quantities of interest. The first is the hazard, or incidence, function at time t, denoted by h(t). The hazard is the instantaneous rate of appearance of malignant tumor in a previously tumor free tissue. Let T represent the time to appearance of the first tumor in a tissue. Then

$$h(t)=\lim_{\Delta t\to 0} (\mathrm{Prob}\{t \le T \le t + \Delta t \mid T \ge t\}/ \Delta t)$$

If N(t) denotes the number of individuals of age t in a given population, then $N(t)*h(t)$ is the predicted age specific breast cancer incidence function.

The second quantity that we are interested in is P(t), the probability that a malignant cell is generated by time t. The hazard and probability are related by the equation

$$P(t) = 1- \exp(- \int_0^t h(s)\, d s)$$

P(t) represents the predicted cumulative incidence of breast cancer by age t.

For a k hit model, the incidence function can be written as follows (27):

$$h(t)= \mu_k \; E[Y_k(t) \mid Y_{k+1}(t) =0] \qquad \text{(A1)}$$

where E[$Y_k(t)$ ∣ $Y_{k+1}(t)$ =0] denotes the expectation. Since cancer is a rare disease, the probability P(t) of there being a malignant cell at time t is close to zero. Thus we can make the approximation

$$E[\,Y_k(t) \mid Y_{k+1}(t)=0] \cong E[\,Y_k(t)]. \qquad \text{(A2)}$$

We will illustrate the solution of the model for the case of the 5-hit model. The solution of the other models is done similarly. In order to estimate the expectation E[$Y_5(t)$], we set up ordinary differential equations to estimate the numbers of $Y_5(t)$ before and after age 45. For t ≤ 45, $Y_i(t)$ (i=1, …,5) should satisfy the following ordinary differential equations:

$$\begin{aligned}
\frac{dY_1(t)}{dt} &= \gamma_1 Y_1(t)\left(1-\frac{Y_1(t)+Y_2(t)+Y_3(t)+Y_4(t)+Y_5(t)}{N_0}\right)-\mu_1 Y_1(t)\\
\frac{dY_2(t)}{dt} &= \mu_1 Y_1(t)+\gamma_2 Y_2(t)-\mu_2 Y_2(t)\\
\frac{dY_3(t)}{dt} &= \mu_2 Y_2(t)+\gamma_3 Y_3(t)-\mu_3 Y_3(t)\\
\frac{dY_4(t)}{dt} &= \mu_3 Y_3(t)+\gamma_4 Y_4(t)-\mu_4 Y_4(t)\\
\frac{dY_5(t)}{dt} &= \mu_4 Y_4(t)+\gamma_5 Y_5(t)-\mu_5 Y_5(t)
\end{aligned} \qquad \text{(A3)}$$

where N_0 denotes the maximal number of progenitor cells. The initial value for (A3) is ($Y_1(0)$, $Y_2(0)$,$Y_3(0)$,$Y_4(0)$, $Y_5(0)$)=(10,0,0,0,0). We have used $\gamma_1=\log_e(N_0)/16$, which ensures that the normal progenitor cells approach to the maximal number by age 20.

After 45 years of age, $Y_i(t)$ (i=1, …,5) should satisfy the following ordinary differential equations:

$$\begin{aligned}
\frac{dY_1(t)}{dt} &= \gamma_1' Y_1(t)-\mu_1 Y_1(t)\\
\frac{dY_2(t)}{dt} &= \mu_1 Y_1(t)+\gamma_2' Y_2(t)-\mu_2 Y_2(t)\\
\frac{dY_3(t)}{dt} &= \mu_2 Y_2(t)+\gamma_3' Y_3(t)-\mu_3 Y_3(t)\\
\frac{dY_4(t)}{dt} &= \mu_3 Y_3(t)+\gamma_4' Y_4(t)-\mu_4 Y_4(t)\\
\frac{dY_5(t)}{dt} &= \mu_4 Y_4(t)+\gamma_5' Y_5(t)-\mu_5 Y_5(t)
\end{aligned} \qquad \text{(A4)}$$

where $\mu_i(i=1,\ldots,5)$ are the same as those of (A3). We permit the growth rates of the normal and intermediate compartments to change after age 45 in response to the changing hormonal environment of the woman.

For any set of parameters, we solve the system of differential equations for the function $Y_5(t)$. This function determines the hazard function thru (A1) and (A2). The hazard function determines the age-specific cancer incidence rate per breast, which we multiply by 2 to obtain the age-specific incidence rate per woman. We determine the set of parameters that provided the best fit to the SEER age-specific cancer incidence data using the numerical optimization routine fminsearch in MATLAB.

Table 2. Estimated optimal values of net growth rates γ_i and $\gamma_i^{'}$ (i=1,2,3,4,5,6) per cell per year for each cell compartment and each model obtained by the best fit to the SEER age-specific cancer incidence data with maximum size of normal proginator compartment of 10^7. Table 2 (a): age less than 45 years, Table 2 (b): age more than 45 years.

(a)

Cell Compartment	2-Hit	3-Hit	4-Hit	5-Hit	6-Hit
Normal	1.007	1.007	1.007	1.007	1.007
I_1	0.0999	0.0226	0	0	0
I_2		0.0372	0	0	0
I_3			0.0748	0	0
I_4				0	0
I_5					0

(b)

Cell Compartment	2-Hit	3-Hit	4-Hit	5-Hit	6-Hit
Normal	-0.0437	-0.0781	-0.1347	-0.1209	-0.1308
I_1	0.0226	-0.0038	-0.0542	-0.0378	-0.0444
I_2		-0.0098	0.0623	-0.0097	-0.0001
I_3			0.0013	-0.0248	-0.142
I_4				-0.0778	0.1263
I_5					-0.2892

Table 3. Estimated optimal values of mutation rates per cell per year for each cell compartment based on normal proginator compartment of 10^4 cells.

Cell Compartment	2-Hit	3-Hit	4-Hit	5-Hit	6-Hit
Normal	$1.2x10^{-5}$	$4.1x10^{-5}$	$3.6x10^{-6}$	$1.4x10^{-5}$	$2.4x10^{-5}$
I_2	$3.5x10^{-5}$	$5.4x10^{-4}$	$6.0x10^{-4}$	$1.6x10^{-3}$	$1.3x10^{-4}$
I_3		$3.9x10^{-3}$	$1.7x10^{-1}$	$8.5x10^{-2}$	$1.8x10^{-1}$
I_4			$9.0x10^{-2}$	$8.5x10^{-2}$	$3.1x10^{-1}$
I_5				$6.6x10^{-2}$	$2.2x10^{-1}$
I_6					$1.8x10^{-1}$

Tumor Progression During the Silent Interval

For an exponentially growing population of tumor cells with a proportion η of the daughter cells remaining clonogenic, after g generations there are $G_g = (2\eta)^g$ clonogenic tumor cells. If the tumor is detected when it reaches the size of N_d cells, then the silent interval includes

$$g = \log(N_d)/\log(2\eta) \qquad [1]$$

generations of tumor growth.

The cumulative number of tumor cell divisions after i generations of growth (M_i) during the silent interval is given by the geometric series $M_{i+1} = M_i + G_i$. Since $G_i = (2\gamma)^i$, M_i is the sum of a geometric series and

$$M_g = [1-(2\eta)^g] / [1-2\eta] . \qquad [2]$$

Heidenreich WF, Luebeck EG, Hazelton WD, Paretzke HG, Moolgavkar SH. Multistage models and the incidence of cancer in the cohort of atomic bomb survivors. Radiation Research 2002; 158:607-14.