The Age-specific Incidence of Cancer: phases, transitions and biological implications

Supplementary Information

A Hazard and Survival Function of the MSCE model

Figure 1 shows a schematic representation of the $k-$stage MSCE model. In this model, cells arrive in the first pre-initiation stage according to a Poisson process with intensity $\mu_0 X$, where X is the number of normal susceptible cells. A cell in the ith pre-initiation stage can divide into one ith pre-initiated cell and one $(i+1)$th pre-initiated cell with rate μ_i, for $i = 1, \cdots, k-3$. A $(k-2)$th pre-initiated cell can divide into one $(k-2)$th pre-initiated cell and one initiated cell with rate μ_{k-2}. Once a cell is initiated, it expands clonally via a birth-death-mutation process with rates α, β, μ_{k-1}, respectively. The hazard and survival functions for the multistage clonal expansion (MSCE) model have been derived previously [1, 2]. In particular, the survival function for the $k-$stage model, $S_k(t)$ ($k \geq 3$), can be calculated iteratively by

$$S_{k,j}(t) = \exp\left(-\mu_{k-j} \int_0^t (1 - S_{k,j-1}(t-u))\, du \right), \tag{1}$$

for $j = 3, ..., k$, where $S_k(t) = (S_{k,k}(t))^X$ and

$$S_{k,2}(t) = \left(\frac{q-p}{qe^{-pt} - pe^{-qt}} \right)^{\mu_{k-2}/\alpha}. \tag{2}$$

Here,

$$p = \frac{1}{2}(-(\alpha - \beta - \mu_{k-1}) - \sqrt{(\alpha - \beta - \mu_{k-1})^2 + 4\alpha\mu_{k-1}}), \tag{3}$$

$$q = \frac{1}{2}(-(\alpha - \beta - \mu_{k-1}) + \sqrt{(\alpha - \beta - \mu_{k-1})^2 + 4\alpha\mu_{k-1}}). \tag{4}$$

Note that $pq = -\alpha\mu_{k-1}$ and $(p+q) = -(\alpha - \beta - \mu_{k-1})$.

The hazard function of the k-stage model can be obtained from

$$h_k(t) = -\frac{\partial}{\partial t} \ln S_k(t). \tag{5}$$

In particular,

$$h_2(t) = \frac{\mu_0 X}{\alpha}\left(\frac{qp(e^{-qt} - e^{-pt})}{qe^{-pt} - pe^{-qt}}\right), \tag{6}$$

$$h_3(t) = \mu_0 X\left[1 - \left(\frac{q-p}{qe^{-pt} - pe^{-qt}}\right)^{\mu_1/\alpha}\right], \tag{7}$$

$$h_4(t) = \mu_0 X\left(1 - \exp\left\{\int_0^t \mu_1\left[\left(\frac{q-p}{qe^{-p(t-u)} - pe^{-q(t-u)}}\right)^{\mu_2/\alpha} - 1\right]du\right\}\right). \tag{8}$$

B Linear, Exponential and Quadratic Phases of the 3-stage Model Hazard

B.1 Linear Phase

The 3-stage model has the hazard function

$$h_3(t) = \mu_0 X(1 - S_2(t)), \tag{9}$$

where X is the number of susceptible stem cells, μ_0 is the mutation rate of the first hit at the tumor suppressor locus and $S_2(t)$ represents the 2-stage 'tumor survival' probability given by

$$S_2(t) = \exp\left(-\mu_1 \int_0^t (1 - S_1(t-u))\, du\right). \tag{10}$$

The survival function associated with the clonal expansion stage (i.e., the probability that a single clone originated at time 0 hasn't become malignant by time t) is:

$$S_1(t) = 1 + \frac{1}{\alpha}\frac{pq(e^{-pt} - e^{-qt})}{qe^{-pt} - pe^{-qt}}. \tag{11}$$

Let $\xi \equiv q/(-p) > 0$, $a \equiv (-p) > 0$ and $b \equiv \alpha$, then (11) reduces to:

$$S_1(t) = 1 + \frac{\xi a(1 - e^{(1+\xi)at})}{b(1 + \xi e^{(1+\xi)at})}, \tag{12}$$

and clearly

$$p_\infty \equiv \lim_{t\to\infty}(1 - S_1(t)) = 1 - \frac{b-a}{b} \approx 1 - \frac{\beta}{\alpha}. \tag{13}$$

p_∞ also represents the probability that a single premalignant clone initiated at time 0 eventually becomes malignant.

Now,

$$h_3(t) = \mu_0 X(1 - S_2(t)) = \mu_0 X \left(1 - e^{-\mu_1 \int_0^t (1-S_1(t-u))du}\right). \tag{14}$$

If $(\mu_1 \int_0^t (1 - S_1(t-u))\, du) << 1$, which is the case for all $t < \frac{1}{\mu_1 p_\infty}$, then

$$h_3(t) \approx \mu_0 X \mu_1 \int_0^t (1 - S_1(t-u))\, du. \tag{15}$$

Let T^* denote the time to malignancy for a single clone, conditional on the clone not becoming extinct (i.e., destined to become malignant) and let $F^*(t)$ denote the cumulative distribution of this random variable, then

$$F^*(t) = \frac{1 - S_1(t)}{p_\infty} \tag{16}$$

and

$$T_s \equiv E[T^*] = \int_0^\infty (1 - F^*(u))du = \int_0^\infty \left(1 - \frac{1 - S_1(u)}{p_\infty}\right) du \tag{17}$$

$$= -\frac{\ln(\frac{\xi}{1+\xi})}{a} \approx -\frac{\ln(\xi)}{a} = -\frac{\ln(q/(-p))}{-p} \approx -\frac{\ln(\alpha\mu_2/(\alpha-\beta)^2)}{\alpha - \beta}, \tag{18}$$

where the approximations are valid provided $\mu_2 << 1$ (i.e., if $-p \approx (\alpha - \beta)$) and $\mu_2 < p^2/\alpha$.

A simple calculation shows that for t large enough,

$$\int_0^t (1 - S_1(t-u))\, du = p_\infty t - p_\infty T_s, \tag{19}$$

so

$$h_3(t) \approx \mu_0 X \mu_1 p_\infty (t - T_s). \tag{20}$$

We see that the 3-stage hazard behaves linearly for $t >> T_s$ until $t \sim o(1/(\mu_1 p_\infty))$. Moreover, this line is equal to zero at $t = T_s$, so the effective sojourn time can be identified and directly estimated from the intercept of the linear phase of the hazard with the time axis (see Figs.2 and 3).

B.2 Exponential and Quadratic Phases

Let $\xi \equiv q/(-p) > 0$ and $a \equiv (-p) > 0$, then the two-stage survival function (10) can be written as

$$S_2(t) = \left(\frac{(1+\xi)e^{\xi a t}}{1 + \xi e^{(1+\xi)at}}\right)^{\frac{\mu_1}{\alpha}}. \tag{21}$$

For $\xi << 1$, a Taylor series expansion of $S_2(t)$ around $\xi = 0$ gives

$$S_2(t) = 1 + \frac{\mu_1}{\alpha}\xi(1 + at - e^{at}) + o(\xi^2). \tag{22}$$

Thus,

$$h_3(t) = \mu_0 X(1 - S_2(t)) \approx \frac{\mu_0 X \mu_1}{\alpha}\xi(e^{at} - at - 1). \tag{23}$$

Clearly if $at << 1$, then $S_2(t)$ can be further approximated by $S_2(t) \approx 1 - \frac{\mu_1}{\alpha}\xi\frac{a^2t^2}{2}$, hence

$$h_3(t) \approx \frac{\mu_0 X \mu_1}{\alpha}\xi\frac{a^2t^2}{2} = \frac{1}{2}\mu_0 X \mu_1 \mu_2 t^2. \tag{24}$$

C Prevalence of Premalignant Clones

Under the 3-stage model, the probability that no premalignant clone has occurred by time t, $P_0(t)$, is given by:

$$P_0(t) = \exp\left(\frac{\mu_0 X(1 - e^{-\mu_1 t})}{\mu_1} - \mu_0 X t\right). \tag{25}$$

The prevalence of premalignant clones at age t, $PREV(t)$, can be approximated by

$$PREV(t) \approx (1 - P_0(t))p_\infty = p_\infty\left(1 - \exp\left(\frac{\mu_0 X(1 - e^{-\mu_1 t})}{\mu_1} - \mu_0 X t\right)\right). \tag{26}$$

Moreover, for early ages and assuming that $\mu_1 << 1$, the prevalence can be further approximated by $\frac{1}{2}\mu_0 X \mu_1 p_\infty t^2$. Thus, the slope of the linear phase of the hazard function can be interpreted as the curvature in the age-specific prevalence of the generated preneoplastic clone in a young population.

D Linear Phase of 4-stage Model Hazard

The hazard of the 4-stage MSCE model is given by

$$h_4(t) = \mu_0 X\left[1 - \exp\left(-\mu_1 \int_0^t (1 - S_2(t - u))\, du\right)\right], \tag{27}$$

where in this case

$$S_2(t) = \left(\frac{q - p}{qe^{-pt} - pe^{-qt}}\right)^{\mu_2/\alpha}. \tag{28}$$

If $\mu_1 << 1$, then

$$h_4(t) \approx \mu_0 X \mu_1 \left(\int_0^t (1 - S_2(t-u)) \right). \tag{29}$$

As before, let T^* denote the time to malignancy for a single clone, and let $F^*(t)$ denote the cumulative distribution of this random variable. Under the 4-stage model, initiated (premalignant) clones cannot become extinct because $S_2(t) \to 0$ as $t \to \infty$. Therefore,

$$F^*(t) = (1 - S_2(t)) \tag{30}$$

and $T_s \equiv E[T^*]$ satisfies

$$T_s = \int_0^\infty S_2(u) du. \tag{31}$$

Following a similar argument as before, we can see that for $1/\mu_1 > t >> T_s$, the 4-stage hazard can be approximated by

$$h_4(t) \approx \mu_0 X \mu_1 (t - T_s). \tag{32}$$

E Parameter Estimates

Table 1: Parameter Estimates : Colorectal Cancer

Parameter		3-stage model* ($k = 3$)	4-stage model ($k = 4$)
slope†	males	0.200e-03 (0.196,0.206)×e-03	0.195e-03 (0.187,0.204)×e-03
$\mu_0 X \mu_1 p_\infty$	females	0.161e-03 (0.157,0.166)×e-03	0.198e-03 (0.185,0.213)×e-03
$\alpha - \beta$	males	0.167 (0.165,0.170)	0.161 (0.157,0.165)
	females	0.141 (0.138,0.144)	0.146 (0.141,0.151)
μ_{k-2}/α	males		1.379 (0.960,1.980)
	females		0.428 (0.342,0.535)
$\mu_{k-1}/(1 - \beta/\alpha)$	males	0.145e-04 (0.132,0.158)×e-04	0.135e-04 (0.105,0.175)×e-04
	females	0.427e-04 (0.388,0.469)×e-04	0.775e-04 (0.688,0.872)×e-04
Δ AIC‡	males	804.04	819.87
	females	339.23	401.84

* The 3-stage hazard depends also on μ_1/α, however this parameter is not identifiable when $\mu_1/\alpha \ll 1$. For the purpose of estimation, we assumed $\mu_0 = \mu_1$ and $\alpha = 9$. The model fit is insensitive to the value of α chosen, and can be rescaled as long as the slope remains unchanged.

† For the 4-stage model, $p_\infty = 1$.

‡ ΔAIC : relative to the Armitage-Doll model. AIC (Akaike Information Criterion)$\approx -2 \times \log \mathcal{L}$(loglikelihood) $+ 2\times$ no. of estimated parameters.

References

[1] Luebeck, EG, Moolgavkar, SH (2002) Multistage carcinogenesis and the incidence of colorectal cancer. Proc. Natl. Acad. Sci. U S A 99:15095–15100.

[2] Meza, R, Luebeck, EG, Moolgavkar, SH (2005) Gestational mutations and carcinogenesis. Math. Biosci. 197:188–210.

Table 2: Parameter Estimates : Pancreatic Cancer

Parameter		3-stage model* ($k=3$)	4-stage model ($k=4$)
slope†	males	0.390e-04 (0.370,0.412)×e-04	0.468e-04 (0.385,0.570)×e-04
$\mu_0 X \mu_1 p_\infty$	females	0.277e-04 (0.255,0.302)×e-04	0.273e-04 (0.250,0.298)×e-04
$\alpha - \beta$	males	0.179 (0.172,0.187)	0.192 (0.177,0.208)
	females	0.162 (0.156,0.168)	0.156 (0.148,0.164)
μ_{k-2}/α	males		0.401 (0.204,0.775)
	females		1.299 (0.680,3.480)
$\mu_{k-1}/(1-\beta/\alpha)$	males	0.138e-04 (0.107,0.178)×e-04	0.206e-04 (0.149,0.283)×e-04
	females	0.178e-04 (0.148,0.214)×e-04	0.174e-04 (0.146,0.208)×e-04
ΔAIC‡	males	77.07	84.56
	females	245.23	232.76

* The 3-stage hazard depends also on μ_1/α, however this parameter is not identifiable when $\mu_1/\alpha \ll 1$. For the purpose of estimation, we assumed $\mu_0 = \mu_1$ and $\alpha = 9$. The model fit is insensitive to the value of α chosen, and can be rescaled as long as the slope remains unchanged.

† For the 4-stage model, $p_\infty = 1$.

‡ ΔAIC : relative to the Armitage-Doll model.

Table 3: Estimated Mean Sojourn Time, T_s

colorectal	3-stage model	4-stage model
males	56.0	55.3
females	57.5	64.0
pancreas	3-stage model	4-stage model
males	52.9	57.9
females	56.3	55.8