

**Estimating parameters of cell turnover:
Smith-Martin type models and method of rescaling**

April 12, 2004

Mathematical Biosciences Institute
Ohio State University

Sergei S. Pilyugin, University of Florida
Vitaly V. Ganusov, Rustom Antia, Rafi Ahmed, Emory University
Rob de Boer, Utrecht University
Kaja Murali-Krishna, University of Washington

Outline of the talk

1. Formulation of the problem
2. Description of the experiment and CFSE data
3. Smith-Martin model
4. Fallacies of a simple model
5. Analysis of the SM model
6. Rescaling method
7. Applications to experimental datasets
8. Comparison of different methods
9. Discussion and conclusions

The problem

Understanding the dynamics of immune responses requires the development of quantitative analytical tools to estimate the rates of kinetic parameters describing cell division and death.

Specifically, we want to know:

How rapidly immune cells divide and die with or without antigenic stimulation?

How these rates may differ for naive/effector/memory T cells and B cells?

What is the turnover rate of memory cells in homeostasis? What regulates the homeostasis?

Available data:

Earlier approaches used BrdU and *D*-glucose labels. Both labels simply allow to distinguish between divided and undivided cells. Development of CarboxyFluorescein Diacetate Succinimidyl Ester (CFSE) label allowed to accurately track the number of divisions that a given cell has undergone following transfer *in vivo*.

Description of CFSE data

A CFSE dataset includes both the total population of cells $X(t)$ as well as the CFSE division profile at several time points following labeling with CFSE. The CFSE division profile provides the fraction of cells $f_n(t)$ which have undergone n divisions by time t . The total number of cells having undergone n divisions at time t then equals $X_n(t) = X(t)f_n(t)$. We can also calculate the mean number of divisions and the variance of division numbers for labeled cells at time t , $\mu(t)$ and $\sigma^2(t)$:

$$\mu(t) = \sum_{n=0}^{n_{max}} n f_n(t), \quad \sigma^2(t) = \sum_{n=0}^{n_{max}} n^2 f_n(t) - \mu^2(t) \quad (1)$$

Since, after n divisions each cell has given rise to 2^n progeny, by dividing the number of cells in the n^{th} generation by 2^n we can estimate the number of precursors which would have generated the current cell population if no death occurred in the population. The precursor number $s(t)$ equals:

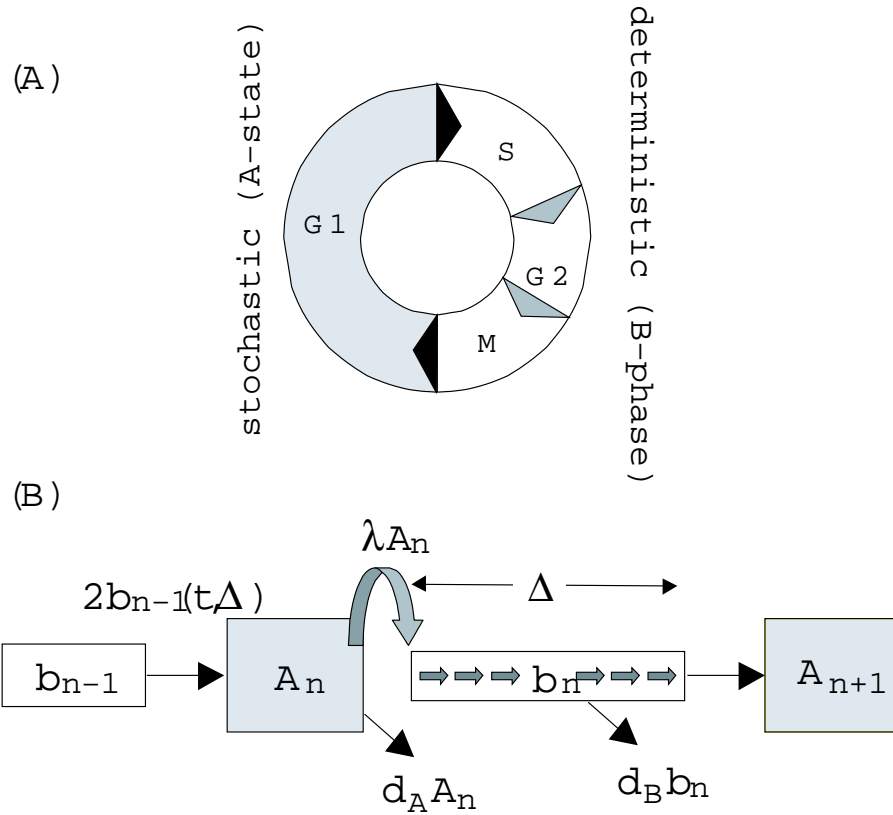
$$s(t) = \sum_{n=0}^{n_{max}} X_n(t) 2^{-n}. \quad (2)$$

Example of a CFSE dataset

		t (days)			
$X_n(t)$		0.5	1.25	3	8
n	0	7.38	7.07	1.77	0.0
	1	0.0	0.64	6.10	0.29
	2	0.0	0.0	6.58	5.71
	3	0.0	0.0	1.28	19.97
	4	0.0	0.0	0.0	18.83
	5	0.0	0.0	0.0	7.99
	6+	0.0	0.0	0.0	4.0
$X(t)$		7.38	7.71	15.73	56.79
$s(t)$		7.38	7.38	6.63	5.56
$\mu(t)$		0	0.08	1.46	3.71
$\sigma^2(t)$		0	0.08	0.64	1.14

The dynamics of P-14 Tg naive CD8 T cells after adoptive transfer into irradiated hosts (Murali-Krishna and Ahmed, 2000). The numbers above correspond to the number of cells per spleen divided by 10^4 . The generation numbers are denoted by n .

The Smith-Martin model



In this (Smith-Martin or SM) model the progression of cells through the cell cycle involves a stochastic recruitment of cells from an A-state (corresponding approximately to the G1 phase of the cell cycle) into the dividing B-phase (approximately equivalent to the S, G2, and M phases of the cycle). The B-phase has a fixed duration Δ . The recruitment of cells from the A-state into the B-phase occurs at the fixed rate λ (the waiting time in the A-state has exponential distribution with the parameter λ). As a first approximation, we assume that the death rates in the A-state and the B-phase are given by two constants d_A and d_B , respectively.

Equations of the SM model

The SM model is expressed as a system of differential equations of the form

$$A_n'(t) = 2b_{n-1}(t, \Delta) - (\lambda + d_A)A_n(t), \quad (3)$$

$$\frac{\partial b_n(t, s)}{\partial t} + \frac{\partial b_n(t, s)}{\partial s} = -d_B b_n(t, s), \quad (4)$$

$$b_n(t, 0) = \lambda A_n(t), \quad (5)$$

where $A_n(t)$ and $B_n(t) = \int_0^\Delta b_n(t, s) ds$ are the numbers of cells having undergone n divisions in the A-state and the B-phase, respectively, and $b_n(t, s)$ is the time density of cells which have spent time s in the B-phase ($0 \leq s \leq \Delta$). Note that we assume the time- and division-independence of λ , Δ , d_A and d_B , i.e., a homogeneous cell population.

Integrating along characteristics in (4-5), we find that

$$b_n(t, s) = \lambda e^{-d_B s} A_n(t - s), \quad b_{n-1}(t, \Delta) = \lambda e^{-d_B \Delta} A_{n-1}(t - \Delta), \quad (6)$$

and derive the following equation for the change of the total numbers of cells in the A-state $A_n(t)$:

$$A_n'(t) = 2\lambda e^{-d_B \Delta} A_{n-1}(t - \Delta) - (\lambda + d_A)A_n(t). \quad (7)$$

Fallacy of a simple random birth/death model

A simple random birth/death model is a limiting case of the SM model with $\Delta \rightarrow 0$. Unfortunately, there are many ways to consider this limit.

Model 1: $\Delta = 0$, $d_B = 0$, $d = d_A$;

$$\frac{dX_n(t)}{dt} = 2\lambda X_{n-1}(t) - (\lambda + d)X_n(t),$$
$$X_n(t) = \underbrace{\frac{(2\lambda t)^n}{n!} e^{-2\lambda t}}_{\text{distribution}} \underbrace{\left[X_0 e^{(\lambda-d)t} \right]}_{\text{total \#}}.$$

Predicted distribution: Poisson with $2\lambda t$ ($\lambda = d$).

Model 2: $\Delta \rightarrow 0$, $\exp(-d_B \Delta) = 1 - f$, $d_A = 0$;

$$\frac{dX_n}{dt} = 2\lambda(1-f)X_{n-1} - \lambda X_n,$$
$$X_n(t) = \underbrace{\frac{(2\lambda(1-f)t)^n}{n!} e^{-2\lambda(1-f)t}}_{\text{distribution}} \underbrace{\left[X_0 e^{\lambda(1-2f)t} \right]}_{\text{total \#}}.$$

Predicted distribution: Poisson with λt ($f = 0.5$).

The two models would produce equally good fits to the data, yet the estimates of λ would differ by a factor of 2.

Analysis of the SM model

Exponential rate of change r . Summing over n in (7), we obtain

$$A'(t) = 2\lambda e^{-d_B \Delta} A(t - \Delta) - (\lambda + d_A)A(t). \quad (8)$$

The principal eigenvalue r is the unique real root of the characteristic equation

$$r = 2\lambda e^{-(d_B+r)\Delta} - (\lambda + d_A). \quad (9)$$

Normalizing the cells in the n -th generation by 2^n , we obtain the equation for the total number of precursors in the A-state

$$A'_N(t) = \lambda e^{-d_B \Delta} A_N(t - \Delta) - (\lambda + d_A)A_N(t), \quad (10)$$

and the corresponding characteristic equation

$$d = \lambda e^{-(d_B+d)\Delta} - (\lambda + d_A), \quad (11)$$

where $d \leq 0$ is the rate of decline of precursors. In fact,

$$-\max(d_a, d_B) \leq d \leq -\min(d_A, d_B).$$

Proliferating fraction P

In the exponential phase, $A(t) = A(0) \exp(rt)$. The total number of cells in the B-phase can be calculated from (6):

$$B(t) = \int_0^\Delta \lambda A(t - \tau) \exp(-d_B \tau) d\tau = A(t) \int_0^\Delta F(\tau) d\tau,$$

where $F(\tau) = \lambda e^{-(r+d_B)\tau}$. Since the total population size, $X(t) = A(t) + B(t)$, is also changing exponentially, we can calculate the fraction of cells in division (i.e., in the B-phase):

$$P = \frac{\int_0^\Delta F(\tau) d\tau}{1 + \int_0^\Delta F(\tau) d\tau} = \frac{\lambda(1 - e^{-(r+d_B)\Delta})}{r + d_B + \lambda(1 - e^{-(r+d_B)\Delta})}. \quad (12)$$

In what follows, we will also refer to P as the *proliferating fraction*.

Increase in the mean number of division

We define $p_n^A(t) = \frac{A_n(t)}{A(t)}$, and let $\mu^A(t)$ and $(\sigma^2)^A(t)$ be the mean and the variance of the distribution generated by $p_n^A(t)$. In the exponential phase, the rate of change of $p_n^A(t)$ is given by

$$(p_n^A)'(t) = 2\lambda e^{-(d_B+r)\Delta} \left(p_{n-1}^A(t-\Delta) - p_n^A(t) \right). \quad (13)$$

From (13), we obtain the equation for $\mu^A = \sum_n n p_n^A$:

$$(\mu^A)'(t) = k[\mu^A(t-\Delta) - \mu^A(t)] + k. \quad (14)$$

Letting $\mu^A(t) = at + a_1$, we substitute this expression into (14) and solve for a ,

$$(\mu^A)'(t) = a = \frac{k}{1 + k\Delta}. \quad (15)$$

A similar argument shows that $(\sigma^2)^A(t)$ also grows linearly but at a different rate

$$((\sigma^2)^A)'(t) = b = a(1 - \Delta a)^2 = \frac{k}{(1 + k\Delta)^3}, \quad (16)$$

and therefore $(\sigma^2)^A(t) = a(1 - \Delta a)^2 t + a_2$. The rates of increase for μ^{A+B} and $(\sigma^2)^{A+B}$ are the same, i.e., $\mu'(t) = a$ and $(\sigma^2)'(t) = b$. Similar to (15), the rate of increase of $\mu_N(t)$ is given by

$$(\mu_N)'(t) = \frac{k_N}{1 + k_N\Delta},$$

where $k_N = \lambda + d_A + d$.

Explicit solutions of the SM model

Assuming that all cells are initially in the A-state of the 0-th generation, that is, $A_0(0) = A_0 > 0$ and $b_0(s, 0) = 0$ ($A_n(0) = b_n(s, 0) = 0$ for $n \geq 1$), we can compute the Laplace transforms of $A_n(t)$:

$$\mathcal{L}\{A_n(t)\} = \frac{A_0 2^n \lambda^n e^{-(s+d_B)n\Delta}}{(s + \lambda + d_A)^{n+1}}, \quad n \geq 0.$$

Therefore,

$$A_n(t) = A_0 (2\lambda e^{(\lambda+d_A-d_B)\Delta})^n e^{-(\lambda+d_A)t} \frac{(t - n\Delta)^n}{n!} H(t - n\Delta). \quad (17)$$

The number of cells in the B-phase of each generation is then:

$$B_n(t) = A_0 \frac{2^n \lambda^{n+1} e^{-d_B t}}{n! (\lambda + d_A - d_B)^{n+1}} \cdot \Gamma(n + 1, \max(0, t - (n + 1)\Delta) \cdot (\lambda + d_A - d_B)), \quad (18)$$

where $\Gamma(n + 1, x, y) = \int_x^y z^n e^{-z} dz$ is the incomplete generalized gamma function. The total number of cells $X_n(t)$ is the sum of (17) and (18).

The number of undivided cells $X_0(t)$ is given by

$$X_0(t) = A_0 e^{-(\lambda+d_A)t} \frac{\lambda e^{(\lambda+d_A-d_B)\tau} + d_A - d_B}{\lambda + d_A - d_B},$$

where $\tau = \min(t, \Delta)$.

Estimating parameters of the SM model

Method of direct fitting

Using expressions (17) and (18), we can fit the explicit analytic solution of the SM model to the CSFE dataset. This approach has been used independently by Bernard *et al* (Bernard, 2003). There are three major difficulties in this approach. First, we have to artificially impose a set of specific initial conditions. Second, we have to fit an infinite dimensional model to a very scarce dataset. Third, this method produces very broad confidence intervals when the data is very noisy. In fact, we tested this method on a sample generated numerically from the SM model with 5 % added noise and found its performance inadequate.

Estimating parameters of the SM model (cont-d)

Method of indirect fitting

If we neglect the transients, only three measured quantities are independent in the SM model (for example, r , $a = \mu'$ and $b = (\sigma^2)'$, but other combinations may be used as well). Thus, given the CFSE data alone, we can only estimate the duration of the B-phase Δ and the accessory parameter k :

$$\Delta = \frac{1}{a} \left(1 - \sqrt{\frac{b}{a}} \right), \quad k = a \sqrt{\frac{a}{b}}. \quad (19)$$

By assuming that death during the cell cycle is restricted only to the A-state (i.e., $d_B = 0$) or the B-phase (i.e., $d_A = 0$), we can also estimate the range for the average division time $T = \Delta + \lambda^{-1}$ of CFSE labeled cells:

$$T \in \left(\Delta + \frac{2e^{-r\Delta}}{k}, \Delta + \frac{1}{k-r} \right). \quad (20)$$

If, in addition, we could measure the proliferation fraction P

$$P = \frac{k(e^{(d_B+r)\Delta} - 1)}{2(d_B + r) + k(e^{(d_B+r)\Delta} - 1)}, \quad (21)$$

then we could obtain d_B from (21) and estimate

$$\lambda = \frac{k}{2} e^{(d_B+r)\Delta}, \quad d_A = k - \lambda - r.$$

A more general model: assumptions

- (i) Cells proliferate by binary fission, which is viewed as an event when one mother cell leaves its generation and at the same time two identical daughter cells enter the next generation.
- (ii) The cell population is homogeneous, that is, cells in different age classes (generations) exhibit identical behavior which is independent of the behavior of other cells or a given cell's genealogy.
- (iii) The generation time (defined as the time required for a cell to complete the cell cycle) is a random variable that depends only on the time since the cell entered the generation as a newborn daughter.
- (iv) Cell death (the removal of cells from the population) is a random event whose probability of occurrence depends only on the time since the birth of a given cell.
- (v) The probability that division and death events occur simultaneously is negligibly small.
- (vi) The system is closed, so that new cells enter the population only through division.

General model: equations

We let $x_n(t, s)$ denote the density of cells in the n -th generation at time t which have already spent s time units in this age class. We call s the age of cells inside the generation. We let $\lambda(s)$ denote the probability rate of cell division at age s and $d(s)$ denote the probability rate of cell death at age s inside the generation.

$$\frac{\partial x_n(t, s)}{\partial t} + \frac{\partial x_n(t, s)}{\partial s} = -(\lambda(s) + d(s))x_n(t, s), \quad n \geq 0. \quad (22)$$

The total number of cells that divide anywhere between the times t and $t + dt$ is given by

$$\left(\int_0^\infty \lambda(s)x_n(t, s) ds \right) dt,$$

and therefore twice the number of cells enter the next generation between t and $t + dt$. The dynamics of consecutive generations are coupled through the boundary condition

$$x_n(t, 0) = 2 \int_0^\infty \lambda(s)x_{n-1}(t, s) ds, \quad n \geq 1. \quad (23)$$

We let $X_n(t) = \int_0^\infty x_n(t, s)ds$ denote the total number of cells in n -th generation at time t . A typical data set is a table of values

$$X_n(t_m), \quad n = 0, 1, \dots, N_{\max}, \quad t_m \in \{t_1, t_2, \dots, t_k\}.$$

Rescaled model: equations

The boundary condition given by equation (23) can be considered as the rate of external input of cells into the n -th generation. Equation (22) is linear, therefore rescaling the external input by a factor of $a \geq 0$ will result in the identical rescaling of the cell density $x_n(t, s)$. The dynamics of the rescaled cell densities $x_n(t, s, a) = a^n x_n(t, s)$ satisfy the equations

$$\frac{\partial x_n(t, s, a)}{\partial t} + \frac{\partial x_n(t, s, a)}{\partial s} = -(\lambda(s) + d(s))x_n(t, s, a), \quad (24)$$

$$x_n(t, 0, a) = 2a \int_0^\infty \lambda(s)x_{n-1}(t, s, a) ds. \quad (25)$$

We let $X_n(t, a) = a^n X_n(t) = \int_0^\infty x_n(t, s, a) ds$. The total number of cells in the rescaled population is

$$X(t, a) = \sum_{n=0}^{\infty} X_n(t, a).$$

Rescaled model: characteristic equation

Summing up equations in (24-25), we obtain

$$\frac{\partial x(t, s, a)}{\partial t} + \frac{\partial x(t, s, a)}{\partial s} = -(\lambda(s) + d(s))x(t, s, a),$$

$$x(t, 0, a) = 2a \int_0^\infty \lambda(s)x(t, s, a) ds.$$

Substitution $x = e^{r(a)t}U(s)$ yields the eigenfunctions

$$U(s) = \exp\left(-\int_0^s (\lambda(z) + d(z)) dz\right) = \exp(-\Lambda(s) - D(s)),$$

and the characteristic equation

$$1 = 2a \int_0^\infty \lambda(s)e^{-\Lambda(s)-D(s)}e^{-r(a)s} ds. \quad (26)$$

Method of rescaling:

For a given experimental time series $X_n(t)$, we generate a family of rescaled time series $X_n(t, a)$, calculate the change in the total population size $X(t, a)$ with time, and evaluate the exponential proliferation rate $r(a)$ for each value of a . Theoretically, we can obtain the function $r(a)$ by manipulating a single time series.

Which parameters can we estimate?

(1) In theory, we can find the generating kernel $\lambda(s)e^{-\Lambda(s)-D(s)}$ by inverting the Laplace transform given by (26). Nevertheless, we *cannot* infer $\Lambda(s)$ or $D(s)$ without some additional assumptions. Roughly speaking, we cannot estimate birth or death rate *within* the cell cycle.

(2) The parameters that we can estimate must therefore describe the kinetics of a complete cell cycle. For example, we can estimate the probability that a cell dies before dividing (\mathcal{D}), or the mean generation time for surviving cells (\mathcal{T}) and its higher order moments.

(3) The probability that a cell divides before dying is

$$1 - \mathcal{D} = \int_0^\infty \lambda(s)e^{-\Lambda(s)-D(s)} ds. \quad (27)$$

Here we assume that all cells eventually divide, i.e. $\Lambda(s) \rightarrow \infty$.

(4) The mean generation time for surviving cells is

$$\mathcal{T} = \frac{1}{1 - \mathcal{D}} \int_0^\infty s\lambda(s)e^{-\Lambda(s)-D(s)} ds. \quad (28)$$

(5) The variance of generation times for surviving cells is

$$\sigma_{\mathcal{T}}^2 = \frac{1}{1 - \mathcal{D}} \int_0^\infty s^2\lambda(s)e^{-\Lambda(s)-D(s)} ds - \mathcal{T}^2. \quad (29)$$

Evaluating \mathcal{D} :

Let a^* be such that $r(a^*) = 0$. Equations (21) and (22) imply that

$$1 - \mathcal{D} = \frac{1}{2a^*}, \quad \mathcal{D} = 1 - \frac{1}{2a^*}.$$

Evaluating \mathcal{T} :

We implicitly differentiate (21) with respect to a at the point $a = a^*$ and substitute $r(a^*) = 0$ to obtain

$$0 = 2 \int_0^\infty \lambda(s) e^{-\Lambda(s) - D(s)} ds - 2a^* r'(a^*) \int_0^\infty s \lambda(s) e^{-\Lambda(s) - D(s)} ds.$$

This equation can be simply written as

$$\frac{1}{a^*} - 2a^* r'(a^*) \mathcal{T} (1 - \mathcal{D}) = 0,$$

and since $2a^*(1 - \mathcal{D}) = 1$, we derive that

$$\mathcal{T} = \frac{1}{a^* r'(a^*)}. \tag{30}$$

Evaluating $\sigma_{\mathcal{T}}^2$:

Repeated implicit differentiation of (21) yields

$$\sigma_{\mathcal{T}}^2 = \mathcal{T}^2 \left(1 + (a^*)^2 r''(a^*) \mathcal{T} \right). \tag{31}$$

Rescaling method applied to SM model

The characteristic equation of the rescaled SM model is given by

$$r(a) = 2a\lambda e^{-(r(a)+d_B)\Delta} - (d_A + \lambda).$$

Using $r(a)$, one can only estimate three parameter combinations which can be written as model-invariant parameters:

$$\mathcal{D} = 1 - \frac{\lambda e^{-d_B\Delta}}{\lambda + d_A}, \quad \mathcal{T} = \Delta + \frac{1}{\lambda + d_A}, \quad \sigma_{\mathcal{T}}^2 = \frac{1}{(\lambda + d_A)^2}. \quad (32)$$

Note that only $\Delta = \mathcal{T} - \sigma_{\mathcal{T}}$ can be estimated from the CFSE data alone. Assuming that death is restricted only to the A-state ($d_B = 0$) or the B-phase ($d_A = 0$), we find the range for the average division time T :

$$T \in \left(\mathcal{T}, \mathcal{T} + \frac{\mathcal{D}}{1 - \mathcal{D}} \sigma_{\mathcal{T}} \right). \quad (33)$$

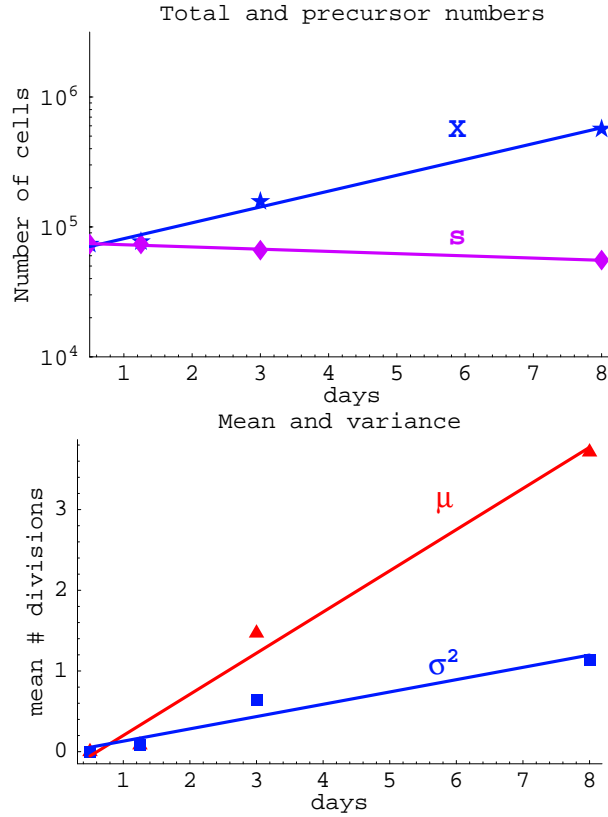
If P is known, all four parameters of the SM model can be estimated by finding d_B from

$$r + d_B = e^{-r\Delta} \frac{1 - \mathcal{D}}{\mathcal{T} - \Delta} (1 - P)(e^{d_B\Delta} - 1)$$

with $r = r(1)$. Using the estimate of d_B , we can then estimate λ and d_A :

$$\lambda = \frac{1 - \mathcal{D}}{\mathcal{T} - \Delta} e^{d_B\Delta}, \quad d_A = \frac{1 - (1 - \mathcal{D})e^{d_B\Delta}}{\mathcal{T} - \Delta}.$$

Analysis of a CFSE dataset



Panel A shows the change in the total number of Tg cells X (stars) and precursors s (diamonds). Panel B shows the change in the mean number of divisions μ (triangles) and the variance σ^2 (boxes). Solid lines represent the best regression lines, the slopes of the regression lines are $r = 0.28 \text{ day}^{-1}$, $d = -0.04 \text{ day}^{-1}$, $a = \mu' = 0.51 \text{ day}^{-1}$, $b = (\sigma^2)' = 0.15 \text{ day}^{-1}$.

Analysis of a CFSE dataset (cont-d)

Parameter	Direct fitting	Indirect fitting	Rescaling method
λ , day^{-1}	0.58(0.19, 1.75)	(-)	(-)
Δ , days	0.94(0.84, 1.04)	0.89(0.63, 1.20)	0.97(0.67, 1.46)
d_A , day^{-1}	0.02(-)	(-)	(-)
d_B , day^{-1}	0.01(-)	(-)	(-)
T , days	2.66(1.44, 6.23)	(2.42, 2.56)*	(2.33, 2.47)**

The estimates for the parameters of the SM model obtained by direct fitting, indirect fitting, and method of rescaling. The 67 % confidence intervals are shown in parentheses. The range for the average division time T was calculated from (28) for rescaling method** and (30) for indirect fitting*.

Discussion and conclusions

1. The CFSE labeling experiments provide a significant amount of information about cell division and death *in vivo*. A major limitation of CFSE data is that it says nothing about the distribution of cells within one generation. Another limitation is the cutoff of the CFSE dataset at 5-10 divisions.

2. There are two major approaches to analyze the CFSE data.
 - (a) The first approach is to use a general model with rescaling method to estimate the invariant parameters of the cell cycle. Downside: these parameters provide limited insight into the specifics of cell division and death.
 - (b) The second approach is to use a specific model (e.g. SM model) and estimate its parameters. Downside: these parameter estimates would be conditional on the particular model and/or the method of fitting used for estimation. Some models are degenerate, and we cannot estimate all parameters. Additional measurements (e.g., the proliferating fraction P) may be needed.
 - (c) So far, we found no single robust method to estimate all parameters of the SM model. Generally, one can obtain reliable estimates for the parameters describing cell division (λ , Δ) but not cell death (d_A , d_B).

3. Our analysis is restricted to the turnover homogeneous cell populations. When naive cells are stimulated with specific antigen, the first division time may be longer than subsequent divisions, and the analysis has to be modified.