Math. Model. Nat. Phenom.
Vol. 7, No. 5, 2012, pp. 24–52
DOI: 10.1051/mmnp/20127504

Mathematical Models of Dividing Cell Populations: Application to CFSE Data

H.T. Banks[1] *, W. Clayton Thompson[12]

[1] Center for Research in Scientific Computation
Center for Quantitative Sciences in Biomedicine, N.C. State University
Raleigh, NC

[2] ICREA Infection Biology Lab
Department of Experimental and Health Sciences
Universitat Pompeu Fabra, Barcelona

Abstract. Flow cytometric analysis using intracellular dyes such as CFSE is a powerful experimental tool which can be used in conjunction with mathematical modeling to quantify the dynamic behavior of a population of lymphocytes. In this survey we begin by providing an overview of the mathematically relevant aspects of the data collection procedure. We then present an overview of the large body of mathematical models, along with their assumptions and uses, which have been proposed to describe the dynamics of proliferating cell populations. While much of this body of work has been aimed at modeling the generation structure (cells per generation) of the proliferating population, several recent models have considered the more fundamental task of modeling CFSE histogram data directly. Such models are analyzed and recent results are discussed. Finally, directions for future research are suggested.

Keywords and phrases: cell proliferation, cell division number, CFSE, ordinary differential equations, cytons, label structured population dynamics, partial differential equations, inverse problems

Mathematics Subject Classification: 00A71, 34A55, 92-02, 92C37, 92C99, 97M60

1. Overview

The human immune response is characterized by a large number of complex, interrelated steps in which the body recognizes an invading pathogen and (hopefully) neutralizes the threat posed by that pathogen. In the face of such complexity, quantitative modeling can play an active role in describing and organizing the vast array of biophysical processes underlying the cellular and subcellular mechanisms of the immune response. This information, in turn, has obvious implications for human health in areas such as allergen treatments, immunosuppression for tissue transplants, immunizations, pathogenesis, etc.

The mathematical analysis of lymphocyte activation and division can be performed on scales ranging from the molecular to the population level, and the behavior of individual cells can be highly variable, given the seemingly random intracellular and environmental factors to which a single cell might respond.

*Corresponding author. E-mail: htbanks@ncsu.edu

© EDP Sciences, 2012

Article published by EDP Sciences and available at http://www.mmnp-journal.org or http://dx.doi.org/10.1051/mmnp/20127504

However, the immune response as a whole (to be understood as the aggregate behavior of all cells in the population) is regular and predictable [77]. In a typical immune response, a small number of cells recognizes and responds to foreign antigen. These initial responders then rapidly produce identical copies of themselves by mitosis and begin to develop effector function in a process known as 'clonal expansion' [64, Sec. 1.17]. This clonal expansion–in particular, the rates at which cells activate, divide, and die immediately following antigen recognition–is a key component in characterizing the efficacy of an immune response. Thus, in this review, focus is placed on the processes of cell division responsible for the clonal expansion of lymphocytes following antigen presentation. In particular, a number of mathematical models are surveyed which attempt to describe the numbers of cells (starting from an initially 'undivided' population) which have divided a given number of times. They key idea is to link these cell numbers (in terms of the number of divisions undergone) back to the dynamic behavior of dividing cells in order to describe the 'cellular calculus' [43] by which cells receive and respond to environmental stimuli.

The mathematical modeling of cell division has long been of interest (see, e.g., [28,76]) but the many mathematical models proposed over the past 40 years have been hard to validate because of the difficulty of obtaining appropriate experimental data. Unlike many cell types, lymphocytes are inherently mobile. In the context of in vitro studies, the nonadherence of lymphocytes makes it difficult to determine the lineage of cells on a large scale [58]. While video microscopy techniques can be used for such a determination and significant information can be obtained from such studies (see, e.g., [49]) the number of cells which can be effectively monitored is comparatively small. Instead, flow cytometric analysis of a population of cells can be used to assess tens of thousands of individual cells (from a population of millions) accurately and efficiently. Thus one trades information on cell lineage for information from a larger population of cells. While cells are measured individually, individual cells (or their progeny) are not followed over the course of the experiment so that the measurement process is in effect an aggregate sampling. That is, one must use *aggregate* (or population level) data in an attempt to describe what are ultimately the dynamics of individual cells. This type of inverse problem is well known in mathematics, and successful mathematical models have been developed and fit to data in a variety of applications such as size-structured marine and insect population models [4, 7, 11, 12], wave propagation models for viscoelastic solids [19], electromagnetic wave propagation [13,14], physiologically-based pharmacokinetics models [6, 20], and HIV models [5]. In addition to these applications, theory for such inverse problems is well-developed [3, 6, 18]. In effect, the mathematical task is to model the processes involved in cell division with limited or no knowledge of familial relation between cells.

Intuitively, one cannot adequately model and estimate rates of cell division and death from knowledge only of the total number of cells in the population at a series of measurement times. For instance, even if one knows that the total number of cells in a population is unchanged between two measurement times, one still cannot say whether the cells are not dividing or are dividing and dying in such a way that the population size is unchanged. Instead, information is needed regarding the generational distribution of cells in the population (that is, the numbers of cells having not yet divided, divided once, twice, etc.). To this end, the relatively recent development of carboxyfluorescein succinimidyl ester (CFSE) by Lyons and Parish [60] as a division-tracking intracellular dye, in conjunction with flow cytometry, has made it possible to quickly and accurately assess generational structure within a population of cells. This is a powerful experimental tool which can be used to augment the formation, analysis, and validation of mathematical models of cell division. These models can then be used for the meaningful quantitative comparison of data sets in different experimental and biological conditions.

Since CFSE has become the de facto method for monitoring lymphocyte division over the past 15 years, a number of mathematical models of cell division have been proposed and tested against CFSE-based flow cytometry data. (While CFSE-based data remains the gold standard for model validation, competitive dyes are currently being developed and tested [70].) We begin this report with an overview of the experimental procedure for CFSE-based flow cytometric analysis of a lymphocyte culture, focusing on mathematically relevant aspects of the resulting data (Section 2). Then several mathematical models of dividing cell populations are surveyed (Section 3) and their salient features highlighted. (Readers who

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

find our overview of interest should find the recent manuscript by Miao, et al., [63] also of interest. In this presentation they provide a comparison of four of the type of models considered here: a branching process, the cyton, a Smith-Martin and linear birth-death ODE model, using simulated data. The branching process model, which outperforms the other three, is also applied to CFSE labeled data for CD4+ and CD8+ T cells proliferating after stimulation.) Our report is limited strictly to mathematical models for a dividing population of cells of a single type. In general, the cell dynamics models considered are phenomenological in nature, describing the dynamic behavior of a population of cells without regard for the underlying molecular pathways governing cell behavior. However, the work surveyed here represents an important initial step toward a more complete description of an immune response, and it seems certain that the framework(s) presented here can be generalized mathematically and experimentally to incorporate more complex information.

2. Flow Cytometry and CFSE Data

CFSE can be used on a wide variety of human and murine cell types, including T lymphocytes, B lymphocytes, NK cells, fibroblasts, hematopoietic stem cells, and smooth muscle cells, as well as bacteria [58,67,71,81]. Such CFSE-labeled cells are analyzed by flow cytometry using a standard fluorescein setup [67, 71]. While CFSE can be used both in vitro and in vivo (for animal studies), the models presented in Section 3 are limited to in vitro studies so that the total number of labeled cells can be assessed. An experiment begins with peripheral blood mononuclear cells (PBMCs) which are isolated either from whole blood or a buffy coat. CFSE is introduced into the culture of PBMCs as carboxyfluorescein diacetate succinimidyl ester (CFDA-SE) which is membrane permeable and thus freely diffuses into the cell. Once inside the cell, intracellular esterases remove the acetate groups, producing CFSE which is membrane impermeable and thus bound inside the cell [67, 71]. After the initial CFSE uptake, excess CFSE (or CFDA-SE) is removed from the culture. Inside the cell, further reactions with amino groups result in the rapid degradation or expulsion of some carboxyfluorescein conjugates, while a fraction of the carboxyfluorescein binds to stable intracellular proteins and remains within the cell for an extended period of time [67].

Carboxyfluorescein groups within the cell absorb light from an excitation laser and then release that energy with a peak emission wavelength of 517nm (green). Cells are passed through the flow cytometer one at a time via hydrodynamic focusing and the fluorescence intensity (FI) of the emission from each cell is quantified. It is known that FI varies directly with the mass of CFSE within a cell [58] so that measured FI is a useful surrogate for the mass of CFSE passing through the measurement apparatus. Because cells become fluorescently labeled via the free diffusion of CFDA-SE across the cell membrane, the uniformity (in the PBMC culture) of initial FI depends upon the heterogeneity of cell size in the PBMC culture [71], and to some extent upon the distribution of certain types of intracellular proteins within the cells [61]. While the various cell types present in a PBMC culture may exhibit a wide range of sizes and protein contents, cells of similar type should be sufficiently similar, and most CFSE labeling experiments are designed to monitor small subclasses of cells (e.g., CD4+ cells; see below). Thus CFSE labeling provides fairly uniform labeling of the initial population of cells [60, 80]. Upon mitosis, carboxyfluorescein is split approximately evenly between the resulting daughter cells; the consequent reduction in mass of CFSE (compared to the original mother cell) results in a decrease in measured FI. Thus measured FI (when compared to the FI of undivided cells) provides some indication of the number of divisions a cell has undergone [58, 60, 67, 71, 80, 83].

There are a number of advantages in using CFSE to study generational structure (and hence proliferation characteristics) in a population of dividing cells. At reasonable concentrations, CFSE is not believed to have any adverse effects on the functioning of the intracellular machinery [57, 59, 60]. Unlike tritiated thymidine (^{3}H-Tdr), CFSE is nonradioactive, and unlike ^{3}H-Tdr, bromodeoxyuridine (BrdU), and dimethylthiaxol (MTT), uptake of CFSE is not dependent upon the activation state of the cells being labeled [29, 60]. Measurements of DNA content can provide an indication of the position of a cell within its cell cycle, but not on the total number of divisions undergone [26]. The approximately

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

even partitioning of CFSE upon mitosis makes it easier to track a larger number of generations when compared to other measurements such as telomere length, and T-cell Receptor Excision Circles [31], and lipophilic dyes such as PHK26 [67, 80]. Given these advantages, experimental data in this report is limited to CFSE-based assays. The mathematical models, however, are not dependent upon any single experimental technique, and any labeling technique which can be used for the accurate determination of division history is acceptable. A survey of potential alternative techniques can be found in [67, 80].

In addition to its use in the determination of division history, CFSE is compatible with numerous other dyes. This makes it possible to measure quantities such as cell surface marker expression, cytokine content, and gene expression simultaneously with division number [58, 60, 80]. For instance, fluorescently tagged anti-CD4 antibodies can be introduced into a sample shortly before measurement by flow cytometry and used to identify the CD4 cells in the population of PBMCs (see Figure 1 and the discussion below). In principle, such techniques can be used to monitor cell differentiation or other division-linked changes as a population of cells divides in response to a stimulus (for example, the down-regulation of CD19 expression as B cells differentiate into plasma cells). As mentioned previously, the current report will focus on the proliferative characteristics of single classes of cells in terms of the number of divisions undergone. Thus cell surface markers are used simply to identify cells in culture; any considerations of differentiation, division-linked changes, or interactions between subpopulations of cells are not pursued here. It should be emphasized, however, that the models presented here can readily be generalized to such situations.

Numerous experimental protocols for CFSE-based proliferation assays are available and can generally be adapted to specific experimental needs [58, 60, 71, 80, 83]. After initial labeling the cells are stimulated to divide. For the data presented here, cells were stimulated at the beginning of the experiment with phytohaemagglutinin (PHA), which activates T cells nonspecifically so that the fraction of cells which respond and are activated is large; specific stimulation is also possible but will result in a much smaller number of dividing cells. Measurement times are given relative to the time at which PHA is introduced into the cell culture, so that initial stimulation is said to occur at $t = 0$ hours. The cells are then plated into separate wells (one for each sample to be measured) containing a nutrient medium. The use of separate wells prevents the disruption of the proliferating cell populations when cells are harvested for measurement. It is tacitly assumed that each well plate contains an identical population of cells at all times as well as identical nutrient.

At each sample time, cells from a single well are harvested and transferred to Trucount tubes containing a known number of calibration beads which can be used to estimate the total number of cells in the sample. Because of physical limitations, a flow cytometer will measure only a fraction of the total contents of a particular tube. The calibration beads can be easily detected in the flow cytometry output, and the number of beads measured by the flow cytometer can be assessed and compared to the total number of beads originally in the Trucount tube, thus indicating the fraction of the sample assessed during the measurement. In the Trucount tubes, cells are further stained with fluorescently labeled antibodies (e.g., anti-CD4) so that cells of a particular type may be identified in the PBMC culture. The flow cytometer is carefully calibrated so that the additional fluorescent labeling does not affect the measurement of FI resulting from intracellular CFSE. After the cells are fixed, washed, etc., (according to the specific experimental protocol) in the Trucount tubes, they are acquired and analyzed by the flow cytometer. The flow cytometer measures tens of thousands of cells (typically $1 - 5\%$ of the contents of the Trucount tube) in seconds, so that any changes to the population of cells during the measurement can be neglected.

Typical output from the measurement process is shown in the top left panel of Figure 1 where each dot represents a single cell and multiple cells with the same measurements are represented by colors for increasing density (blue, green, yellow, red, etc.). The cytometer measures multiple characteristics of the particles passing through the apparatus. Because the PBMC culture contains a variety of cells (T-cells, B-cells, monocytes, granulocytes, etc.) as well as cell fragments and other debris, care must be taken to isolate the cells of interest from the measured data. To do so, 'gates' are set by the experimenter which include (or exclude) cells meeting certain characteristics. In the gating procedure shown in Figure 1, the first gate (top left) is set to include all lymphocytes based upon their size (horizontal axis, measured

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

by forward scatter of the incident laser light, or FSC) and granularity (vertical axis, measured by side scatter of the incident laser light, or SSC). A gate is also set to identify the calibration beads from the Trucount tube (top center). Next, a gate is used to eliminate clumped cells by comparing the height (vertical axis) and area (horizontal axis) of the FSC signal (top right). CD4+ lymphocytes are identified by their expression of CD3 (bottom left) and CD4 (bottom right) as measured by the light emitted from the fluorochromes attached to anti-CD3 and anti-CD4.

After gating, the remaining cells can be counted into bins based upon their measured CFSE expression and then presented as a histogram. Because the mass of CFSE within a cell divides approximately in half with each division, it is most convenient to use a logarithmic scale for CFSE FI. Let z represent the (continuous) logarithmic axis on which FI is measured (hence units of z are log units of intensity, or log UI). Now consider some partition $\{z_k\}$ $(1 \leq k \leq K)$ of this axis. The histogram data consists of the ordered pairs (z_k^j, n_k^j), which represent the number of cells n_k^j with measured FI in the interval $[z_k, z_{k+1})$ at time t_j. Typical histogram data from a single experiment are shown in Figure 2. Each 'peak' in the data represents a distinct generation of cells (i.e., cells in each peak have divided the same number of times since the beginning of the experiment). Because all cells (even in the absence of labelling by CFSE) have a natural brightness or *autofluorescence*, cells which have divided a sufficiently large number of times (typically 8-10 [70]) can no longer be distinguished from background noise in the flow cytometry data. In general, the effect of this detection limit is small as few cells are capable of completing this many divisions during a typical experiment.

This overview of the experimental procedure highlights several features of the resulting data sets which are of mathematical interest. Most importantly, data obtained over the course of multiple measurement times does not represent a repeated census of the population of cells (in which every cell in a single population is measured at each measurement time). Each measurement is obtained from a sample of a population of cells, and in fact, these samples are not drawn from the same population. (While all cells in the experiment are ultimately taken from the same donor, cells are harvested from physically distinct wells at each measurement time.) For almost all analysis and especially for mathematical modeling investigations, it is routinely and tacitly assumed that each well contains a population of cells which is identical to the populations in all other wells at all times. Moreover, only a fraction of the contents of each well is assessed by the flow cytometer. The measured fraction is estimated as the fraction of counted calibration beads to the known total number of beads. It is tacitly assumed that the sample drawn from each well is representative of the total contents of that well, and that the fraction of the well measured is accurately estimated by the bead scaling factor. In the context of a histogram representation of the measured cells, one can obtain an approximate population level histogram by scaling the numbers of cells counted into each bin by the reciprocal of the fraction of beads counted.

Finally, it should be noted that the identification of a particular cell type by setting a series of gates is ultimately a probabilistic process. That is, the measurements consist of samples of cells randomly drawn from a larger population, some of which are counted and some of which are discarded based upon physical characteristics (i.e., the gating criteria). It is typically assumed that for a given cell type of interest (e.g., CD4+ cells), the histogram data (Figure 2) displays counts for CD4+ cells only, and that all CD4+ cells in the population are represented in the data. In fact, the cells represented in the histogram data are all those cells (of any type) which satisfy the gating criteria set by the experimenter. Of course, such criteria are designed to identify particular types of cells and are for the most part effective in doing so. It should be acknowledged, however, that the gating process itself is not exact. Hence some variability in the selected populations of cells is to be expected, particularly considering that the measurements are actually obtained using separate populations of cells. The accurate characterization of this variability is an open problem. Some preliminary work has been considered in [79, Ch. 4].

Given the assumptions of the previous two paragraphs (wells contain identical populations of cells, the fraction of measured cells is representative of the well population, that fraction is exactly knowable, and gates accurately select a population of interest at each measurement time) one is able to treat a series of histograms resulting from flow cytometry measurements as if those histograms represent a complete

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

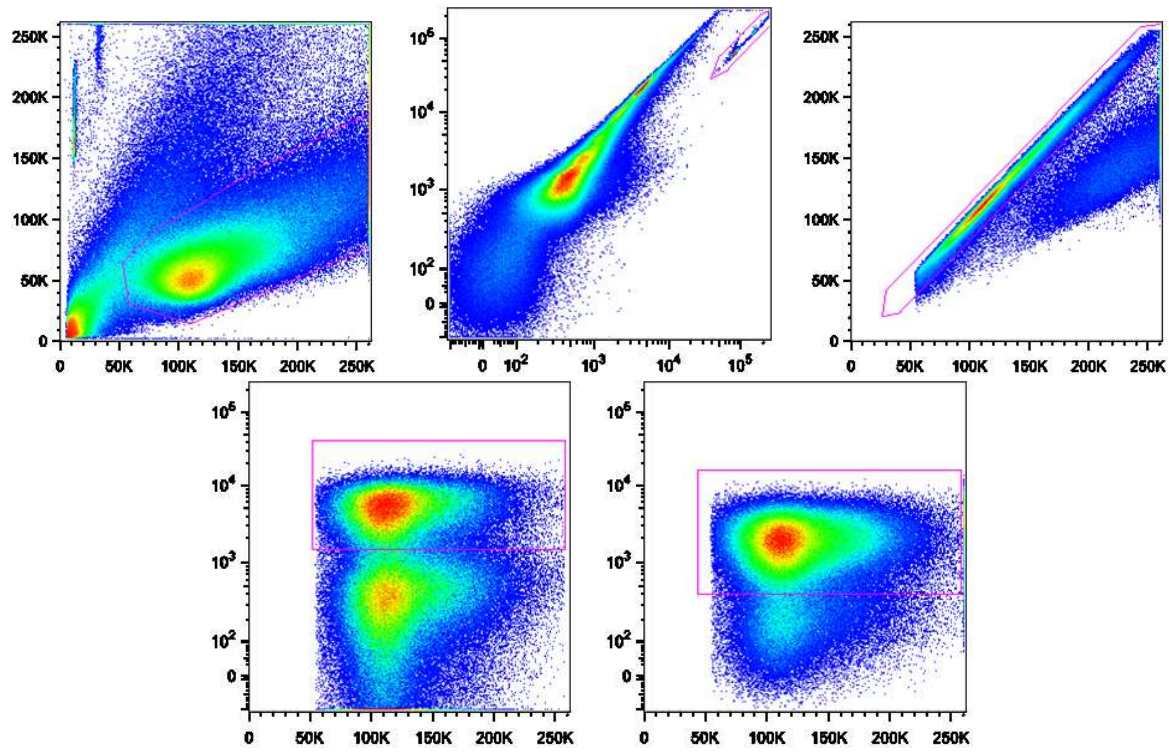


FIGURE 1. Typical gating procedure for flow cytometry data. By rows from top left: Lymphocytes are identified based upon size (forward scattered light, horizontal axis) and granularity (side scattered light, vertical axis); beads are identified by their fluorescence properties in the PE (horizontal axis) and FITC (vertical axis) channels; cell duplets are excluded by comparing the area (horizontal axis) and height (vertical axis) of the forward scattered light; CD3+ cells are identified; CD4+ cells are identified. Graphics prepared by C. Peligero.

census of a single population of cells at each measurement time. Of course, such a treatment is only accurate to the extent that the assumptions are valid. In actual fact, there are small differences (e.g., initial cell number) between populations of cells in distinct wells. Both the scaling factors (computed from the numbers of beads counted) and the gating procedures are subject to some error. Typically, all such errors are small and make minimal difference in the outcome of the experiment (and, in particular, in the calibration of a mathematical model). We emphasize that the assumptions which must be made in order to obtain histogram data are not indicative of any experimental shortcoming, but represent necessary steps to obtain meaningful data from a complex experimental setup. For the moment, the standard assumption will be made that histogram data (such as that in Figure 2) can be treated as census data. Shortcomings resulting from this assumption have been addressed in [79, Ch. 4] and will be highlighted at the end of this report.

Given the histogram data in Figure 2, we return to the original goal of determining the number of cells having divided a given number of times. One simple method of obtaining this information is interval gating, in which the horizontal axis (FI) is partitioned into intervals which are assumed to correspond to particular numbers of divisions undergone. For instance, the data at $t = 48$ hours in Figure 2 might be partitioned at $z = 2.55$ so that cells with FI $z \leq 2.55$ are assumed to have divided once while cells with FI $z > 2.55$ are assumed to be undivided. While this technique is quite simple, there is a clear

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

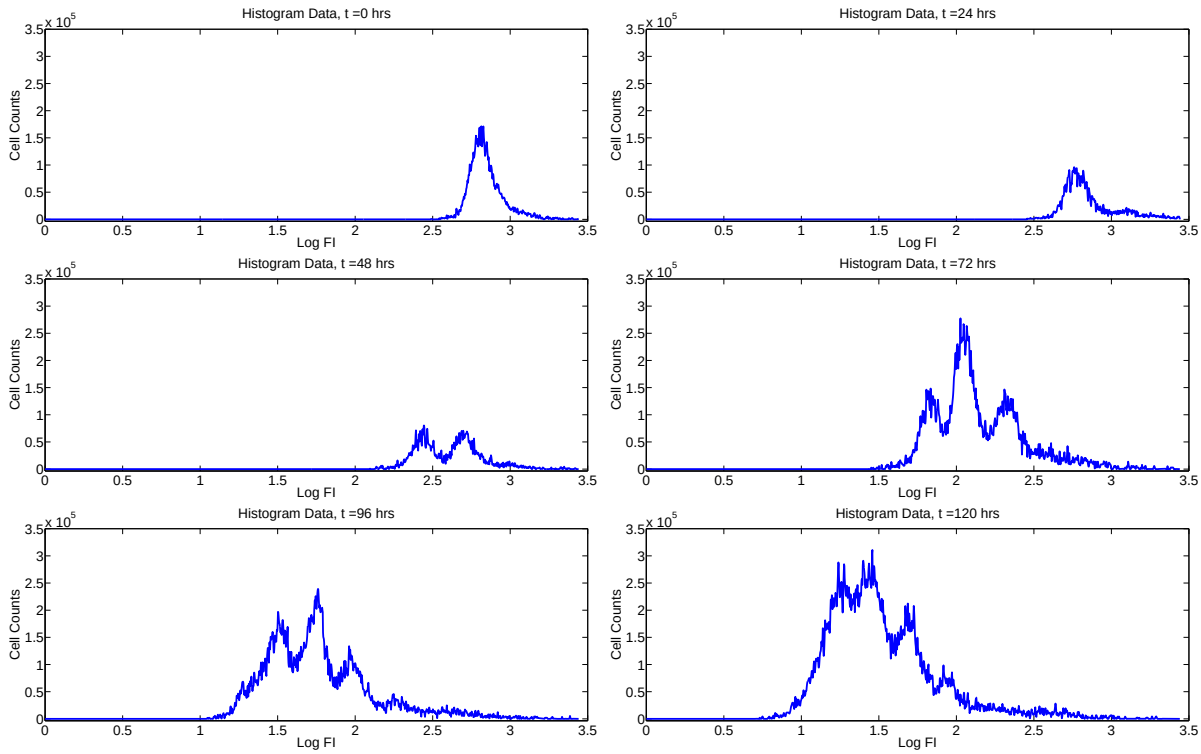


FIGURE 2. Typical histogram representation of data from a CFSE-based proliferation assay. Originally from [56].

shortcoming in that peaks corresponding to distinct generations of cells slightly overlap as a result of inhomogeneity in the initial labeling process. A less arbitrary and hopefully more accurate method to compute cell numbers is to fit each histogram with a series of gaussian-type curves. Consider the family of functions

$$\phi_i^j(z) = C_i^j \exp\left(\frac{(z-\mu_i^j)^2}{2(\sigma_i^j)^2}\right).$$

The parameters C_i^j, μ_i^j, and σ_i^j correspond (approximately) to the height, mean, and variance in the histogram data of the cohort of cells having completed i divisions at time t_j. These parameters can be determined by minimizing the least squares cost functional

$$O^j = \sum_i \sum_{k,j} \frac{1}{w_k^j} \left(\phi_i^j(z_k) - n_k^j\right)^2 \tag{2.1}$$

where w_k^j are weighting terms to account for the variance of the noisy histogram data points. While some work has focused on these weighting terms [79,84], a complete statistical model is not known so that one typically uses $w_k^j = 1$ for all j and k (ordinary least squares). Any similar parameter estimation technique (weighted least squares, likelihood estimation) would also suffice. Alternatively, approximate locations of each peak of cells can be determined using two experimental controls: one for unstimulated cells (which remain undivided) and one for unlabeled cells. See [58, Secs. B,C] for details. Typical results are shown (for the data from Figure 2) in Figure 3. One can then determine the number of cells $\hat{N}_i^j$ having divided

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

	Time (hrs)							
Divisions	0	24	32	48	56	66	75	85
0	50000	14807	9160	1777	429	35	2	0
1	0	527	1861	3920	2351	541	65	3
2	0	75	560	5191	6501	3762	1045	104
3	0	3	66	2731	7161	10401	6627	1672
4	0	0	1	566	3129	11457	16710	10603
5	0	0	0	44	537	5006	16787	26736
6	0	0	0	0	33	859	6684	26859
7	0	0	0	0	0	52	1044	10695
8	0	0	0	0	0	0	54	1671

TABLE 1. Typical data for cell numbers computed from a histogram of CFSE data. Originally from [47].

i times at measurement time t_j by computing

$$\hat{N}_i^j = \sum_k \phi_i^j(z_k).$$

Cell numbers in terms of the number of divisions undergone are shown for a typical experiment in Table 1 (originally from [47]). It is these numbers with which mathematical models of cell division may be calibrated. (Later, in Section 3.5, mathematical models are considered which can be fit directly to histogram data, and the calibrated models themselves used to compute cell numbers. However, the mathematical modeling of cell division dynamics contained in such models is sufficiently similar). In the following section, several mathematical models of dividing cell populations are reviewed, with their assumptions and salient features discussed. Parameter estimation methods are also briefly surveyed, and the various models are compared in terms of their accuracy in describing data from CFSE-based proliferation assays.

3. Mathematical Models

One can clearly see in Figure 2 and in Table 1 that cell division is an asynchronous process. That is, there is a high degree of variability in the times at which cells divide so that multiple generations of cells are present in the population at any given time. The largest source of this asynchrony is generally considered to be the length of time taken for cells to enter the first division [43, 71], which is typically much longer than the time required for subsequent divisions as a result of the time required for initial activation of the cells. It has been shown that many mathematical models of cell division can be placed into one of two classes based upon how the source of this asynchrony is identified and mathematized [54]. The first class of models are consistent with the assumption that asynchrony is the result of inherent stochasticity in the intracellular machinery governing cell growth and division, so that cell division time is distributed in the population of cells under study. The second class of models can be derived from an assumption that population heterogeneity is the result of a deterministic cell interacting randomly with its environment. In the present context, the accuracy of either class of models can only be assessed in terms of accuracy in describing cell counts (e.g., Table 1). Unfortunately, such data alone is insufficient to reach any conclusions regarding which of the two classes of models is more accurate [54]. As such, the discussion below groups mathematical models not by underlying assumptions but rather by the manner in which cell cycle dynamics are modeled.

We remark that the dilemma of how to incorporate asynchrony and variability into "growth" models is not unique to cell proliferation; how to mathematically formulate variability in other growth processes such as marine populations (mosquitofish, shrimp) involving individual deterministic mechanisms in a

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

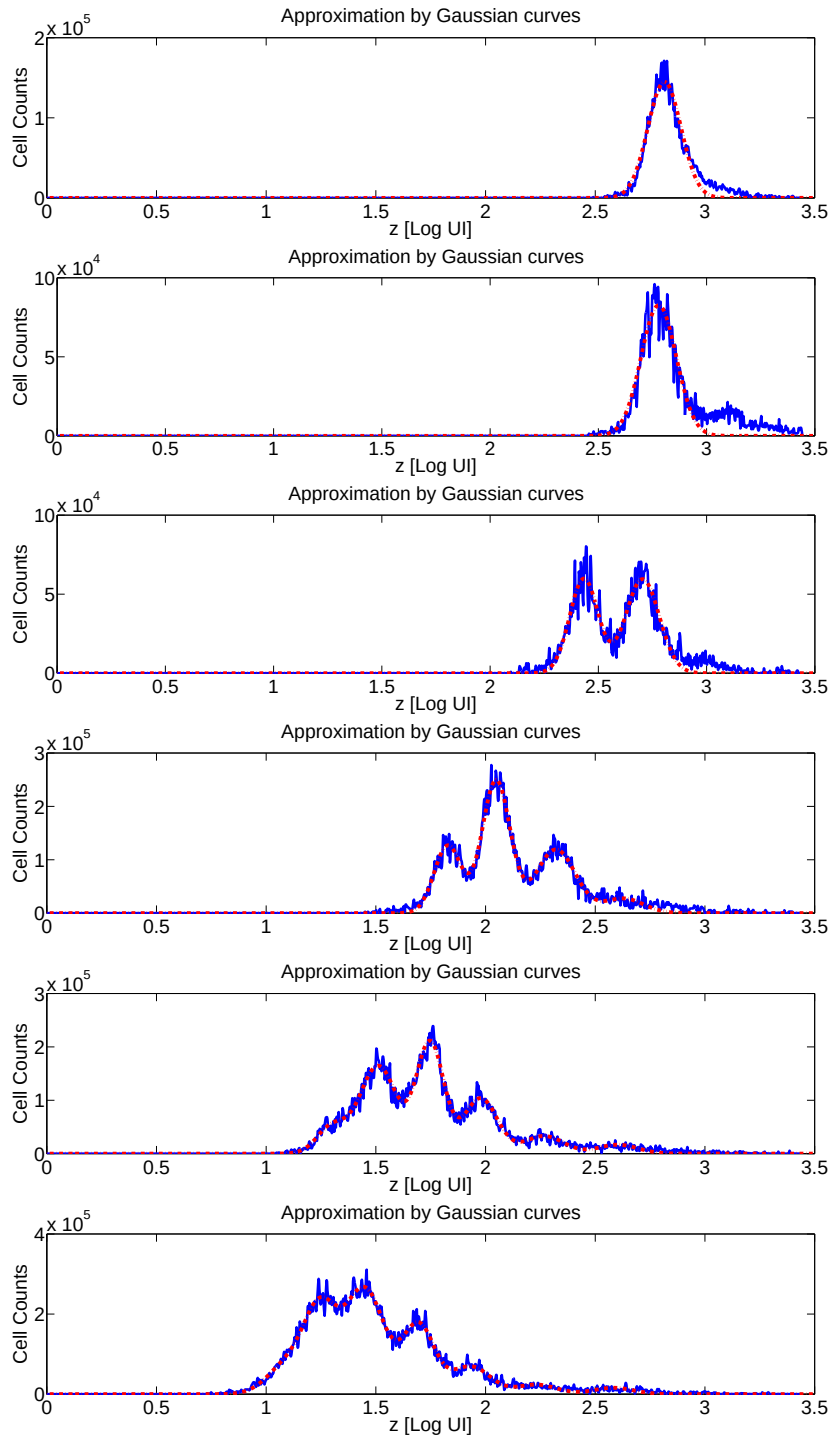


FIGURE 3. Data from Figure 2 (blue, solid) after being fitted with a series of gaussian curves (red, dashed). The fitted curves can be used to approximate the numbers of cells having divided a given number of times.

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

variable environment vs. individual stochastic mechanisms has been treated specifically in [4, 11, 12, 25] as well as more generically in [9, 10, 16, 17]. While probability and stochasticity is fundamental to all of these models, the mathematical constructs used to incorporate asynchrony/variability into the modeling is strikingly different conceptually and computationally.

The division dependence of rates of cell division and death is well-documented in the literature [35–37, 52, 53], though the underlying mechanisms (e.g., senesence, exhaustion) are not known. Ultimately, each group of models can be identified by an underlying probability distribution (over the population of cells) placed upon times to divide and times to die for each generation of cells. First, 'random birth-death' (RBD) models [42], in which times to divide and die are treated as independent exponential processes, are considered. These are followed by a discussion of fixed-cycle models (in which the length of the cell cycle is held constant). Next the Smith-Martin model and various extensions are presented. Finally, more general probabilistic models are addressed. For each class of models, solutions are highlighted and issues associated with data fitting are reviewed. Finally, we turn our focus to modeling CFSE histogram data directly (as opposed to relying upon cell numbers approximately computed from the histograms). As will be shown, this more fundamental level of mathematical modeling involves significantly more information from the CFSE histograms, and should prove quite useful for future modeling efforts.

3.1. Random Birth-Death Models

The simplest class of models used to describe cellular division dynamics are random birth-death models which describe the processes of division and death with exponential rates. Such models follow from the assumption that times to divide and die are independent, exponential random variables. Let $N_i(t)$ be the total number of cells having completed i divisions at time t. Let α be the (exponential) rate at which cells divide, and let β be the (exponential) rate at which cells die. Then a simple RBD model for cell numbers per generation is an autonomous system of ordinary differential equations (ODEs),

$$\begin{aligned}\frac{dN_0}{dt} &= -(\alpha+\beta)N_0(t)\\ \frac{dN_i}{dt} &= -(\alpha+\beta)N_i + 2\alpha N_{i-1}(t),\end{aligned} \tag{3.1}$$

with solution

$$\begin{aligned}N_0(t) &= N_0(0)e^{-(\alpha+\beta)t}\\ N_i(t) &= \frac{(2\alpha t)^i}{i!}N_0(t).\end{aligned}$$

Such a model has been applied to cell numbers obtained from CFSE data by Revy et al., [72]. Given the values of $N_i(t_j)$ at measurement times t_j and the data $\hat{N}_i^j$ (cf. Table 1), one can estimate the parameters (α, β) in an ordinary least squares (OLS) framework similar to Equation (2.1). It follows that the average times to division and death are $1/\alpha$ and $1/\beta$, respectively.

Interestingly, the model (3.1) is also amenable to an alternative method of parameter estimation originally defined by the Hodgkin lab [43]. The discussion below follows the analysis of De Boer and Perelson [36]. Define

$$P_i(t) = N_i(t)/2^i$$

to be the number of *precursors* (this terminology is widely used and perhaps sometimes without sufficient care in the literature–see [73] for a cogent discussion!) for the cells in generation i at time t (allowing for fractional precursors). It follows that

$$\begin{aligned}P_0(t) &= N_0(t)\\ P_i(t) &= \frac{(\alpha t)^i}{i!}N_0(t).\end{aligned}$$

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

Note that the precursors at time t indicate the lines of proliferating cells that originated from a given cell originally in the population at time $t = 0$. The total number of precursors at time t is

$$P(t) = \sum_i P_i(t) = N_0(0)e^{-(\alpha+\beta)t}\left(\sum_i \frac{(\alpha t)^i}{i!}\right) = N_0(0)e^{-\beta t}. \tag{3.2}$$

Finally, the *normalized precursor frequency* for generation i is

$$\frac{P_i(t)}{P(t)} = \frac{(\alpha t)^i}{i!}e^{-\alpha t},$$

which is a Poisson process with mean αt [36]. It follows that a graph of the Poisson process $P_i(t)/P(t)$ versus i will appear normal for sufficiently large t [35] and the mean of this distribution will increase linearly in time with rate α. Given data $\hat{N}_i^j$, one can compute precursor data $\hat{P}_i^j = \hat{N}_i^j/2^i$. It is shown by Hodgkin et al., that, after one division, normalized precursor frequencies estimated from CD4+ cell data can be well-fit by a normal density function. The mean of the gaussian curve (a measure of the average number of divisions undergone) was indeed seen to increase linearly with time. This rate of linear increase could then be used to provide a measure of the proliferation rate parameter α. The parameter β follows from the change in the total numbers of precursors (Equation (3.2)). The authors also compute biological parameters such as mean time to first division [43].

Of course, the model (3.1) is generally too simple to accurately describe most proliferation assay data. However, several extensions and generalizations have been proposed [36]. For instance, it has been observed that experimental data typically exhibit a delay between activation and the onset of proliferation [33]. This can easily be modeled as an initial transient of length τ,

$$\begin{aligned}
\frac{dN_0}{dt} &= \begin{cases} -\beta N_0(t), & 0 \le t \le \tau \\ -(\alpha+\beta)N_0(t), & t > \tau \end{cases} \\
\frac{dN_i}{dt} &= \begin{cases} 0, & 0 \le t \le \tau \\ -(\alpha+\beta)N_0(t) + 2\alpha N_{i-1}(t), & t > \tau \end{cases}.
\end{aligned} \tag{3.3}$$

Additionally, one can explicitly incorporate the fraction, ϕ, of cells which divides by writing

$$\begin{aligned}
\frac{dN_0}{dt} &= \begin{cases} -\beta N_0(t), & 0 \le t \le \tau \\ -(\alpha\phi+\beta)N_0(t), & t > \tau \end{cases} \\
\frac{dN_1}{dt} &= \begin{cases} 0, & 0 \le t \le \tau \\ -(\alpha+\beta)N_1(t) + 2\alpha\phi N_0(t), & t > \tau \end{cases} \\
\frac{dN_i}{dt} &= \begin{cases} 0, & 0 \le t \le \tau \\ -(\alpha+\beta)N_0(t) + 2\alpha\phi N_{i-1}(t), & t > \tau \end{cases}.
\end{aligned} \tag{3.4}$$

Alternatively, studies of cell proliferation and death have long recognized that the gross behavior of a population differs greatly between divided and undivided cells [54]. Thus one could also consider the heterogeneous model [36]

$$\begin{aligned}
\frac{dN_0}{dt} &= -(\alpha_0+\beta_0)N_0(t) \\
\frac{dN_1}{dt} &= -(\alpha+\beta)N_1(t) + 2\alpha_0 N_0(t) \\
\frac{dN_i}{dt} &= -(\alpha+\beta)N_i(t) + 2\alpha N_{i-1}(t).
\end{aligned} \tag{3.5}$$

Mathematical analysis [36] has shown how the parameters of the models (3.3)–(3.5) can be identified from graphical data in a manner similar to that proposed by Gett and Hodgkin [43]. The primary advantage

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

of such a graphical fitting method is that it does not require any complex mathematics or nonlinear optimization software. Biologically relevant parameters are readily determined from simple computations and graphs of the data. In general, the primary shortcomings of such a parameter estimation procedure are tied to the shortcomings of the ODE system which motivates it.

In Figure 4 the models (3.1), (3.4) and (3.5) (the best-fit solution for the model (3.3) is quite similar to the solution for model (3.4)) are shown in comparison to the data from Table 1 after parameter estimation via ordinary least squares. For both models which lack an initial transient ((3.1) and (3.5)) a rapid onset of proliferation is observed in the model solution resulting in the appearance of two many cells (when compared with the data) with high division number at early measurement times. Though the time to first division is more accurately modeled by the initial transient in model (3.4), a rapid onset of proliferation is again observed for higher division numbers. In all three models, the distribution of cells among generations exhibits far greater variance in the model solutions than in the data. In fact, these problems are well-known when using ODE models to describe cell proliferation data. The assumption that times to division and death are exponentially distributed does not impose any minimum cell cycle time [36], which is typically believed to be an important feature of an accurate mathematical model of cell division [36,42,69]. Models lacking this feature may be able to accurately describe the mean number of divisions undergone, particularly for cells which divide slowly [35, 36, 42], but do not appropriately account for the distribution of cells among generations and thus do not typically estimate mean time to first division with accuracy [36, 42, 69]. Still, it should be emphasized that ODE models have been used successfully in some situations [2, 72, 78], particularly when CFSE histogram data is poorly resolved so that generations of cells are hard to distinguish [2]. A more detailed treatment of the assumptions, uses, and limitations of ODE models can be found in [33].

3.2. Fixed-Cycle Models

One alternative to assuming an exponential distribution of times to divide and die (as is implicit in RBD models) is to assume a fixed cell cycle time for all cells in the population. Equivalently, one assumes that the time at which cells divide in any given generation is an atomic measure with a single unit mass (that is, it is a Dirac measure, so that all cells in a given generation divide after spending a fixed length of time in that generation). Of course, as noted previously, a population of activated cells does not undergo completely synchronous division. In keeping with the observation that the primary source of asynchrony within a dividing population is the time required to complete the first division, Deenick et al., [37] proposed a model in which the time to first division is normally distributed. Subsequent generations divide with a fixed cell cycle duration Δ. Cell death is again assumed to be independent of division and is described by an exponential distribution, with a parameter β_0 for undivided cells and a parameter β for divided cells. A derivation of the resulting model can be found in [35] and can be written as an ODE-algebraic system

$$\begin{aligned}
\frac{dN_0}{dt} &= -\frac{R(t)}{2} - \beta_0 N_0(t) \\
\frac{dN_1}{dt} &= R(t) - R(t-\Delta)e^{-\beta\Delta} - \beta N_1(t) \\
N_i(t) &= \left(2e^{-\beta\Delta}\right)^{i-1} N_1(t-(i-1)\Delta).
\end{aligned} \tag{3.6}$$

The function $R(t)$ is a recruitment function describing the entry of cells into the first division. While Deenick et al. [37] propose a normal distribution, subsequent experiments using tritiated thymidine uptake have shown a lognormal distribution to be more accurate [53]. Thus the recruitment function is

$$R(t;\mu,\sigma,\Delta_0,C) = \frac{C}{\sqrt{2\pi}\sigma(t-\Delta_0)}\exp\left(-\frac{(\log(t-\Delta_0)-\log(\mu))^2}{2\sigma^2}\right). \tag{3.7}$$

The parameters μ and σ describe the mean and variance of the lognormal recruitment function, while Δ_0 is an initial transient (so that the mean time to first division is $\mu + \Delta_0$. The parameter $C = 2\phi N_0$,

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

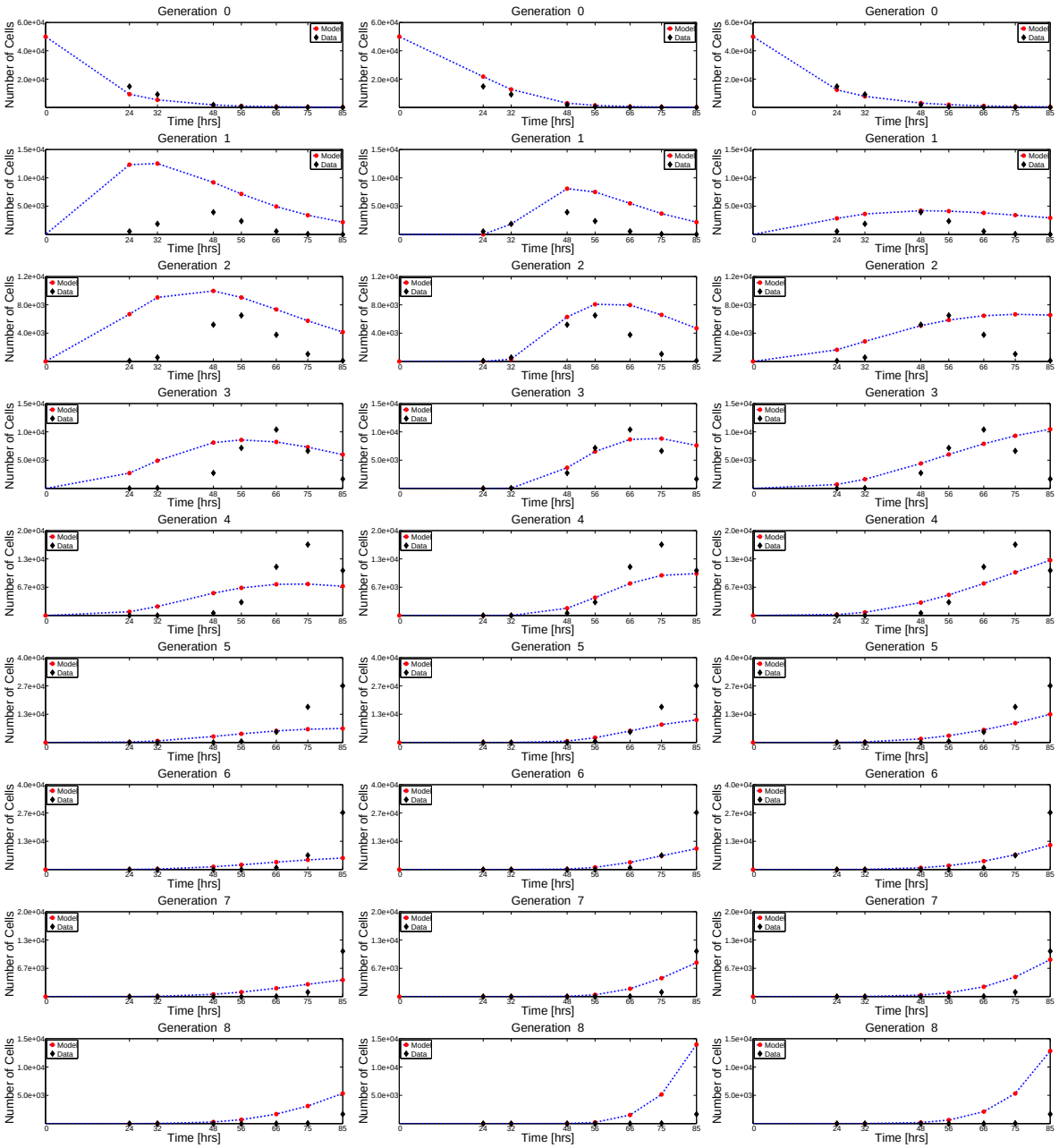


FIGURE 4. OLS best-fit results for several RBD models. Left: Homogeneous ODE model (3.1) [72]. Middle: Homogeneous ODE model with initial transient and precursor fraction (3.4) [36]. Right: Heterogeneous ODE model (3.5) [36]. The assumption of exponential times to divide and die leads to an unrealistically rapid onset of proliferation and a wide distribution of cells in terms of the number of divisions undergone. OLS costs are summarized in Table 2.

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

where N_0 is the initial number of (undivided) cells in the population and ϕ is the precursor fraction which indicates the number of cells which would have divided in the absence of cell death. The model (3.6) has been generalized to account for cell death which is heterogeneous with respect to the number of divisions undergone [52, 53] (for example, as a result of cell exhaustion). For generations $i \geq 1$, define the death rate parameters $\beta_i = \beta + m(i-1)$. Then a generalization of (3.6) is

$$\frac{dN_i}{dt} = 2^{i-1}R(t-(i-1)\Delta)\exp\left(-\sum_{j=1}^{i-1}\beta_j\Delta\right) - 2^{i-1}R(t-i\Delta)\exp\left(-\sum_{j-1}^{i}\beta_j\Delta\right) - \beta_i N_i, \tag{3.8}$$

for $i \geq 2$. The equations for $i = 0$ and $i = 1$ are unchanged.

The two models (3.6) and (3.8) are shown in comparison to the data from Table 1 after parameter estimation via ordinary least squares in Figure 5. Both models are significantly more accurate than RBD models for the given data set. The distribution of cells among various generations is accurately captured, as is the time to the first division. While the solutions to the two models are quite similar as shown in Figure 5, the heterogeneous model (3.8) reduces the OLS cost by approximately 25% and thus seems to indicate that an increase in death with division number would be an important feature of an accurate mathematical model. However, subsequent analysis has indicated that the role of linearly increasing death cannot be distinguished from a linearly increasing cell cycle duration for some data sets in the absence of additional data (e.g., numbers of dead cells) [53]. The most notable systematic inaccuracy of the fixed-cycle models occurs at large time ($t \geq 56$ hours) for cells having completed 1-4 divisions, in which the model consistently overestimates the total number of cells. Since cell death is modeled exponentially (and is thus independent of time) and the model is accurate for $t < 56$ hours, this discrepancy cannot be explained (in the current model formulation) by heterogeneity of cell death with respect to division number.

3.3. Smith-Martin Models

Rather than consider a fixed cell cycle time Δ which is constant for all cells in the population, one could instead assume some level of variability in cell cycle time. In many cases, mathematical formalisms of such cell cycle dynamics begin with the Smith-Martin [76] model, in which the cell cycle is divided into an A state and a B state. The A phase of the Smith-Martin model (which is approximately the G1 phase of the cell cycle) is assumed to have a stochastic duration so that cells may remain in the A phase indefinitely, but exit into the B phase with a fixed transition probability. Meanwhile the B phase (which is approximately the S, G2, and M phases of the cell cycle) is assumed to have a deterministic length after which cells divide and return to the A state. The Smith-Martin model thus imposes a minimum cell cycle time (the duration of the B state) while also allowing for variability in cycle lengths in the population of cells.

Nordon et al., [66] used a Smith-Martin model to account for precursor numbers for CFSE-based cytometry data and derived a set of biologically meaningful parameters which could be determined from graphical representations of the data (cf., the method of Gett and Hodgkin [43]). A complete Smith-Martin model which accounts for total cell numbers computed from CFSE data can be derived from a coupled set of equations [29, 69]. Assume the stochastic A phase has a duration which can be modeled as an exponential random variable. Then the transition of cells from the A state to the B state is modeled with an ordinary differential equation. The fixed duration of the B phase is then modeled with an age-structured partial differential equation in which all cells must progress from an initial age $a = 0$ (when cells enter the B phase from the A phase) to a final age $a = \Delta$ (the duration of the B phase, at which time cells return to the A state). Let $A_i(t)$ represent the total number of cells having completed i divisions at time t and currently in the 'A' state. Let $b_i(t, s)$ be the age-structured density (number per unit age) of cells having completed i divisions at time t and having spent time s in the 'B' state. Then

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

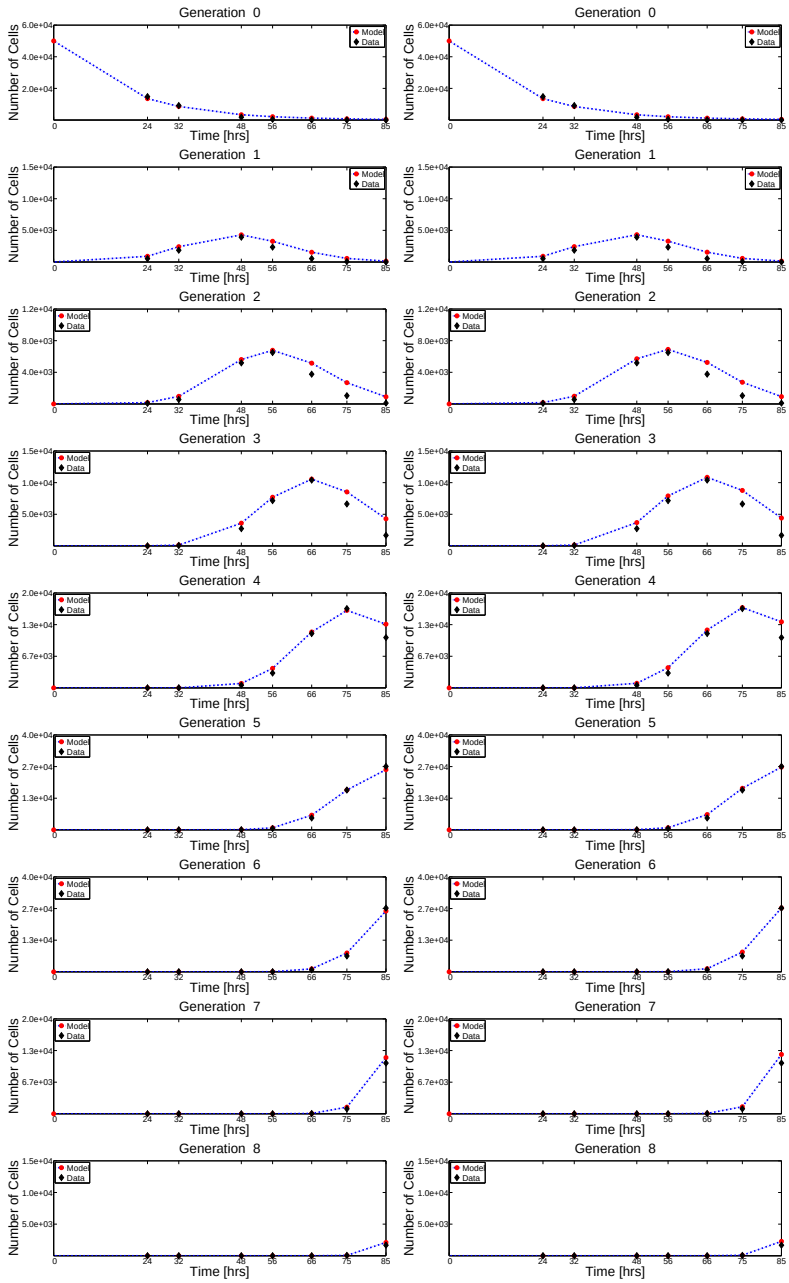


FIGURE 5. OLS best-fit results for two fixed-cycle models. Left: Homogeneous Deenick model (3.6) [37] with lognormal recruitment function (3.7). Right: Heterogeneous Deenick model (3.8) with linearly increasing cell death [35]. OLS costs are summarized in Table 2.

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

the Smith-Martin model is

$$\begin{aligned} \frac{dA_i(t)}{dt} =& 2b_{i-1}(t,\Delta) - (\alpha - \beta_A)A_i(t) \\ \frac{\partial b_i}{\partial t} + \frac{\partial b_i}{\partial s} =& -\beta_B b_i(t,s) \quad 0 \le s \le \Delta \\ b_i(t,0) =& \alpha A_i(t), \end{aligned}$$

where α is the rate of transition from the A state to the B state and β_A and β_B are the (exponential) probabilities of death in the two states. By convention, $b_{-1}(t,s) = 0$ for all t, s. Trivially, the solution for the B phase is

$$b_i(t,s) = \alpha e^{-\beta_B s} A_i(t-s), \quad i \ge 0.$$

Hence it follows that the total number of cells having undergone i divisions is $N_i(t) = A_i(t) + B_i(t)$ where

$$\begin{aligned} \frac{dA_i}{dt} &= -(\alpha + \beta_A)A_i(t) + 2\alpha e^{-\beta_B \Delta} A_{i-1}(t-\Delta), \quad i \ge 0 \\ B_i(t) &= \int_0^\Delta b_i(t,s)ds = \alpha \int_0^\Delta e^{-\beta_B s} A_i(t-s)ds, \end{aligned} \tag{3.9}$$

where it is assumed that $A_{-1}(t) = 0$ for all t. It should be noted that one may alternatively compute the number of cells in the B phase as [41]

$$\frac{dB_i(t)}{dt} = \alpha A_i(t) - \alpha e^{-\beta_B \Delta} A_i(t-\Delta) - \beta_B B_i(t).$$

It is assumed that all cells are in state A_0 at $t = 0$ hours. Given the nature of the two phases of the Smith-Martin model (an A phase with an exponentially distributed duration, and a B phase with a constant duration), it is accurate to think of the Smith-Martin model as a generalization of both RBD models and fixed-cycle models. In fact, it can be shown [69, Appendix] that the homogeneous fixed cycle model (3.6) is equivalent to a Smith Martin model with no A state. The age-structured PDE then has the new boundary condition $b_i(t,s) = 2b_{i-1}(t,\Delta)$. Similarly, the homogeneous ODE model (3.1) is obtained in the limit as $\Delta \to 0$, provided $\beta_A = \beta_B = \beta$. However, under the alternative assumption that $\beta_A = 0$ while $\beta_B \neq 0$, a different ODE model is obtained. Thus it is clear that the manner in which death is modeled within the cell cycle is of fundamental importance for the accurate estimation of parameters. Unfortunately, the parameters β_A and β_B cannot be simultaneously estimated (uniquely) without additional data such as the fraction of cells in division [42]. This lack of identifiability is common to both direct (least squares) estimation and indirect/graphical methods for parameter estimation. In this summary it will be assumed that $\beta_A = \beta_B$ as this is most closely associated with RBD models for purposes of comparison. However, it should be carefully noted that such an assumption may not necessarily accurately reflect the actual underlying dynamics [69].

The Smith-Martin model has been shown to much more accurately describe cell counts obtained from CFSE data when compared to RBD models, perhaps as the result of the minimum cell cycle time [36] (as was observed with the fixed-cycle models). In keeping with the observation that the first division after activation typically takes much longer than subsequent divisions, a generalization of the Smith-Martin model can allow for heterogeneous (with respect to the number of divisions undergone) transition rates

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

and cell cycle times. The resulting model is

$$\begin{aligned}
\frac{dA_0(t)}{dt} =& -(\alpha_0+\beta_0)A_0(t)\\
\frac{dA_1(t)}{dt} =& -(\alpha+\beta_A)A_1(t)+2\alpha_0 A_0(t-\Delta_0)e^{-\beta_B\Delta_0}\\
\frac{dA_i(t)}{dt} =& -(\alpha+\beta_A)A_i(t)+2\alpha A_{i-1}(t-\Delta)e^{\beta_B\Delta}.\\
B_0(t) =& \alpha_0\int_0^{\Delta_0} A_0(t-s)e^{-\beta_B s}ds\\
B_i(t) =& \alpha\int_0^{\Delta} A_i(t-s)e^{-\beta_B s}ds.
\end{aligned} \tag{3.10}$$

As before, the total number of cells in each generation is $N_i(t) = A_i(t) + B_i(t)$. Another possible generalization of the Smith-Martin model uses a recruitment function (cf. the fixed-cycle models (3.6) and (3.8)) to describe the transition of cells from the initial A_0 state to the B_0 state. The resulting model is

$$\begin{aligned}
\frac{dA_0(t)}{dt} =& -\beta_0 A_0(t)-\frac{R(t)}{2}\\
\frac{dA_1(t)}{dt} =& -(\alpha+\beta)A_1(t)+R(t).
\end{aligned} \tag{3.11}$$

The $A_i(t)$, $i \geq 2$ and $B_i(t)$, $i \geq 0$ are computed exactly as in (3.10). Lee and Perelson discuss both a lognormal distribution and a delayed gamma distribution to describe the recruitment function $R(t)$. While the delayed gamma distribution can be derived from assumptions regarding a two-step activation process and permits the analytic solution of the model, the authors find that a lognormal distribution (3.7) is a more accurate description of cellular activation as measured by thymidine incorporation [53] and that is what is used here. The Smith-Martin model can be generalized further by considering a cell death rate which increases with division number, $\beta_i = \beta + (i-1)m$, for $i \geq 1$ [41,53] or by considering a B phase duration which increases with division number, $\Delta_n = \Delta + (i-1)m$ [53].

For the estimation of model parameters, Pilyugin et al., [69] discuss the similarities between the Smith-Martin model above and renewal equations familiar in demographic modeling. They then propose graphical techniques for estimating these parameters, in a manner similar to that of Gett and Hodgkin [43] for RBD models. As above, our summary comparisons focus on least squares estimation. Additional methods of parameter estimation for the Smith-Martin model are surveyed in [42]. Figure 6 depicts two Smith-Martin models after OLS parameter estimation from the data in Table 1. Both the classical homogeneous Smith-Martin model (3.9) (results not shown) and the generalization of the classical model to allow different cycle times for undivided cells (3.10) (left column of Figure 6) fail to accurately describe the generation structure of the data set. Generalization to a recruitment function to describe the first division (3.11) results in a much more accurate fit to the data. Results for additional generalizations such as division-dependent death rates or cell cycle times are not shown as the fit of the model to the data is similar. These generalizations do provide an improved fit to data (in terms of decreasing the OLS cost–see Table 2). The results obtained here confirm the observation of Lee and Perelson, that the effects of increasing cell death with division number or increasing division time improve the model similarly [53], with slightly more improvement arising from increasing cell death [52].

3.4. Probabilistic Models

As discussed above, the classical Smith-Martin model is consistent with the assumption that the probability of division is exponentially distributed with a delay while time to death is exponentially distributed. Yet it is also clear from Figure 6 that a generalized Smith-Martin model, in which the first division is

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

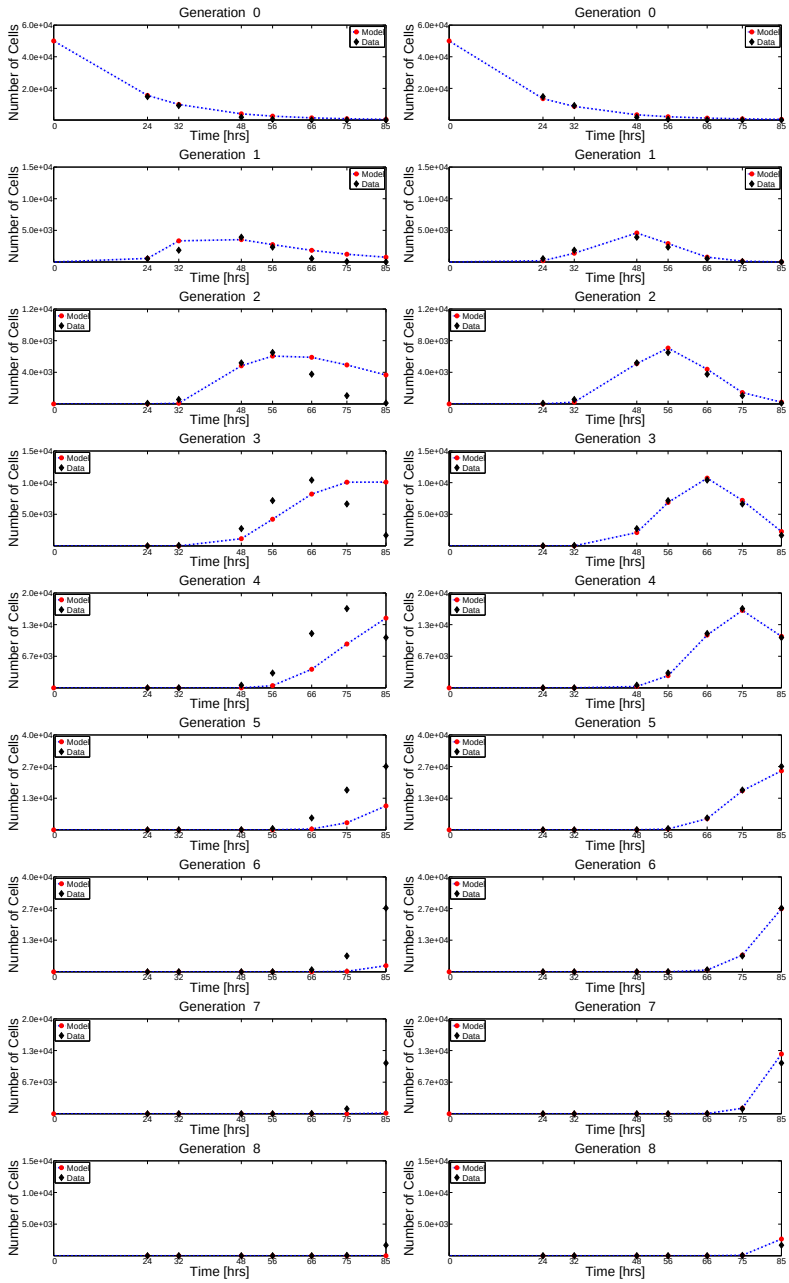


FIGURE 6. OLS best-fit results for two Smith-Martin models. Left: Traditional Smith-Martin model with separate cycle parameters for divided and undivided cells (3.10) [36]. Right: Heterogeneous Smith-Martin model (3.11) with lognormal recruitment function (3.7) [35]. OLS costs are summarized in Table 2.

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

described by a lognormal recruitment process, is much more accurate in describing the data from Table 1. An obvious question, then, is whether such probabilistic structures can be used more generally to describe rates of division and death within a population of cells. Using observations obtained from several experiments, Hawkins et al., [48] show that the apparently random nature of the mechanisms of cell division and death is consistent with the hypothesis that the cellular controls of those processes operate independently of one another. Thus the time to divide and time to die of an individual cell can be thought of as independent random variables; after each division, a single cell receives one realization of each random variable, and the fate of the cell is determined by the minimum of the two realizations. Hawkins et al., propose the name 'cyton' for this regulatory mechanism [47]. It is assumed that division times are not inherited from the previous generation, and that sibling cells are independent of one another. It follows that the expected number of cells having undergone a specified number of divisions at a given time can be determined from the probability distributions from which the times to divide and die are drawn, given the initial number of cells in the population.

In most general terms, the cyton model computes the number of cells having undergone i divisions at time t as

$$N_i(t) = 2^i N_0 \left(\int_0^t r_i(\tau)d\tau - \int_0^t d_i(\tau)d\tau - \int_0^t r_{i+1}(\tau)d\tau \right),$$

where $r_i(t)$ is a density function describing the probability of cells entering the i^{th} generation at time t. Similarly, $d_i(t)$ is a density function describing the probability cells in generation i dying at time t. Of course, the functions $r_i(t)$ and $d_i(t)$ will depend upon the corresponding functions for all prior generations of cells [52]. As a result, these density functions are typically defined via convolution integrals [52, 54]. With appropriate choice of the functions $r_i(t)$ and $d_i(t)$, it can be shown that the cyton model is a generalization of both fixed-cycle models and Smith-Martin models [52, 54].

While the formulation above is quite general, the original formulation of the cyton model by Hawkins et al. [33,47] is more intuitive. For cells having undergone i divisions, consider the numbers of cells $n_i^{div}(t)$ and $n_i^{die}(t)$ having undergone i divisions that divide and die, respectively, at time t. For undivided cells, these functions can be specified as

$$\begin{aligned} n_0^{div}(t) &= F_0 N_0 \left(1 - \int_0^t \psi_0(s)ds\right) \phi_0(t) \\ n_0^{die}(t) &= N_0 \left(1 - F_0 \int_0^t \phi_0(s)ds\right) \psi_0(t), \end{aligned} \tag{3.12}$$

where $\phi_0(t)$ and $\psi_0(t)$ are density functions for the probability that undivided cells divide and die, respectively. F_0 is the fraction of undivided cells which would (in the absence of cell death) progress to the next generation. For the i^{th} division class, let the functions $\phi_i(t)$ and $\psi_i(t)$ be density functions which characterize the probability of division and death, respectively, and let F_i be the progressor fraction. Then for subsequent generations we have

$$\begin{aligned} n_i^{div}(t) &= 2F_i \int_0^t n_{i-1}^{div}(s) \left(1 - \int_0^{t-s} \psi_i(\xi)d\xi\right) \phi_i(t-s)ds \\ n_i^{die}(t) &= 2 \int_0^t n_{i-1}^{div}(s) \left(1 - F_i \int_0^{t-s} \phi_i(\xi)d\xi\right) \psi_i(t-s)ds. \end{aligned} \tag{3.13}$$

Finally, one can compute the cell numbers

$$\begin{aligned} N_0(t) &= N_0 - \int_0^t \left(n_0^{div}(s) - n_0^{die}(s)\right) ds \\ N_i(t) &= \int_0^t \left(2n_{i-1}^{div}(s) - n_i^{div}(s) - n_i^{die}(s)\right) ds. \end{aligned} \tag{3.14}$$

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

Thus the progression of a population of cells through several rounds of division, with the possibility of dying, as well as the variability inherent in that progression, is described by the 'cytons' $\{\phi_i(t), \psi_i(t)\}$ which specify the probability (in time) of division and death for cells having undergone i divisions, as well as the progressor fractions F_i.

Typically, it is assumed that functions $\phi_i(t)$ and $\psi_i(t)$ are lognormal density functions and thus can be described by two parameters, a mean and a variance. In our summary here it will be assumed that undivided cells are characterized by a cyton $\{\phi_0(t), \psi_0(t)\}$ and divided cells are characterized by a separate cyton $\{\phi(t), \psi(t)\}$ for all $i \geq 1$. For the simplest model, it is assumed that $F_i = 1$ for all i and that death is negligible for divided cells. (This can be accomplished either by setting $\psi(t) = 0$, or assigning $\psi(t)$ with sufficiently large mean and sufficiently small variance so that essentially all cells have divided before there is any probability of dying.) This simple model can be generalized by including an initial progressor fraction $F_0 \leq 1$ (but $F_i = 1$ still for all $i \geq 1$). Each of these two models can in turn be generalized by including a nonnegligible probability of death for divided cells. The best fit results of these cyton models are summarized in Table 2. It seems clear from that table that death cannot be ignored among divided cells, a hypothesis further support by the fits to data in Figure 7.

It is clear from Figure 7 that the cyton model is an accurate description of the given data set. In fact, it outperforms all other models tested and does not require a significantly larger number of parameters (see Table 2). Moreover, the probability density functions $\phi_i(t)$ and $\psi_i(t)$ allow for intuitive interpretations of the calibrated model (mean time to division, mean time to death, etc.). Several additional generalizations of the cyton model have been implemented which account for such additional biological and experimental features as division destiny and mechanical death, and software implementing the method is publicly available [34, 47].

3.5. Label-Structured PDE Models

Each of the models discussed thus far has been used effectively to provide various measures of the proliferative capacity of a population of cells. The cyton model in particular is capable of describing with considerable accuracy the time evolution of the generation structure of a population of dividing cells. Of course, validation of such models is based upon cell numbers computed from CFSE histogram data as discussed in Section 2. While such approaches are common and easy to implement, the particular functions (e.g., gaussian densities) might lead to biased computation of cell numbers, particularly if the division peaks in the histogram data are not well-resolved. Thus, there seems to be an advantage in the direct mathematical modeling of CFSE histogram data. To this end, structured population models have been proposed [21, 22, 55, 56, 79]. Such partial differential equation (PDE) models have long been discussed in the context of cell populations [28] and can be structured by age [1, 27, 39, 44], cyclin content [27], size [40, 44, 45, 68] or DNA-content [26], or any number of other physiological variables [62].

Rather than using such physiologically-structured models, the histogram presentation of a cell population in terms of measured fluorescence intensity (Figure 2) makes fluorescence intensity a natural structure variable with which to model. The use of such a nonphysiological structure variable was suggested as early as 2000 for BrdU-based assays [30], and was first explored in the context of CFSE-based assays in [56]. More recent work [8, 21, 22, 55, 79] has consistently demonstrated that the label-structured PDE framework can accurately model the observed histogram data from a CFSE-based proliferation assay. The primary benefit of using such a model is the ability to treat CFSE histogram data directly. Though the reconstruction of CFSE profiles from computed cell numbers has been suggested [47, SI Text], a label-structured model is a more fundamental and more complex effort, as one must account for the intracellular dynamics of label dilution and turnover while simultaneously estimating proliferation and death dynamics at the population level. In return for this added difficulty, one can potentially avoid some issues of biased cell counts, particularly as this method should be less reliant on distinct peak separations in the CFSE histogram data [23, 79].

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

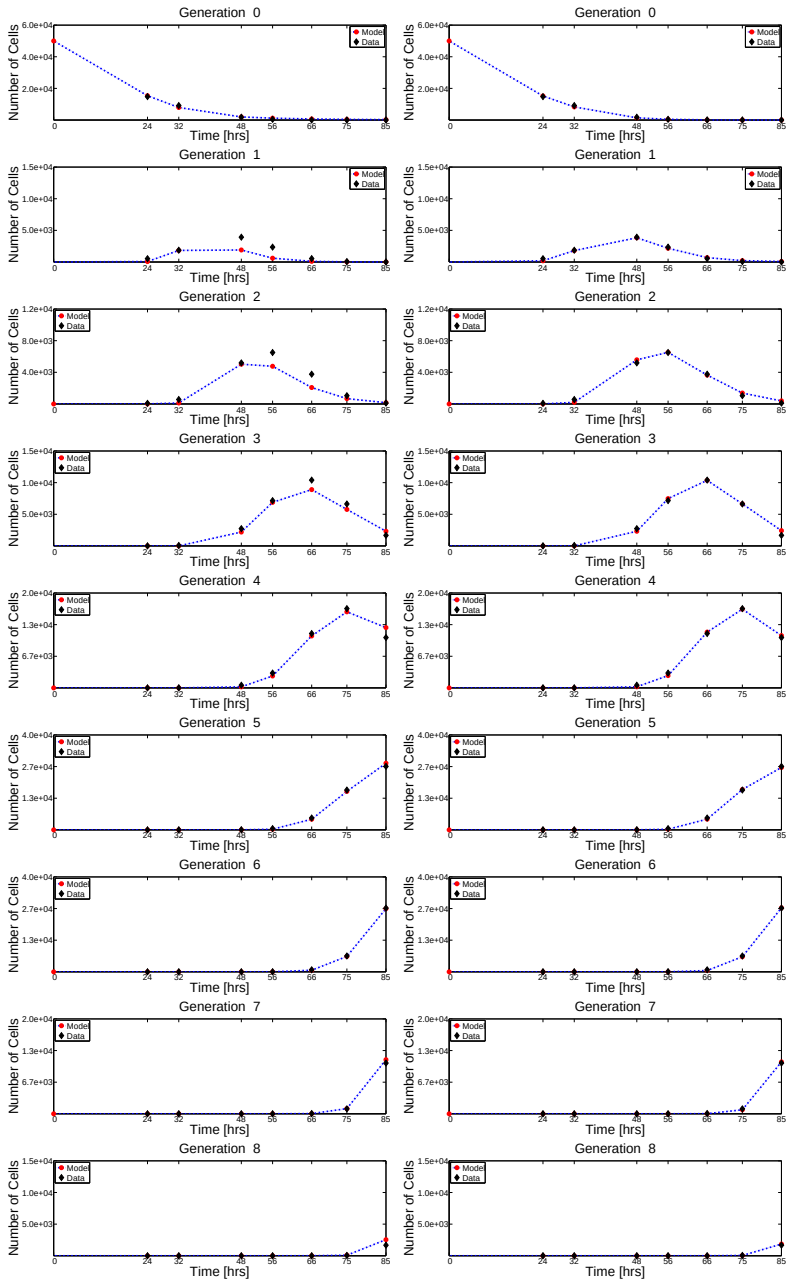


FIGURE 7. OLS best-fit results for two cyton models. Left: 7-parameter model–2 parameters each for lognormal densities $\phi_0(t)$, $\psi_0(t)$, and $\phi_i(t)$, plus a progressor fraction F_0. Right: 9-parameter model–2 parameters each for lognormal densities $\phi_0(t)$, $\psi_0(t)$, $\phi_i(t)$, and $\psi_i(t)$, plus a progressor fraction F_0. All cyton densities are assumed to be log normal, following [47]. OLS costs are summarized in Table 2.

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

Model	Description	# Params	OLS Cost	AIC	AIC Diff
RBD1	Homogeneous ODE; (3.1)	2	1.9159×10^9	1091.51	374.70
RBD2	RBD1 with initial transient; (3.3)	3	1.1667×10^9	1062.26	345.45
RBD3	RBD2 with progressor fraction; (3.4)	4	1.1667×10^9	1064.26	347.45
RBD4	Heterogeneous ODE; (3.5)	4	9.9388×10^8	1054.16	337.35
Fixed-Cycle1	Lognormal recruitment, fixed cycle time; (3.6)	7	4.6761×10^7	867.60	150.79
Fixed-Cycle2	Fixed-Cycle1 with increasing death; (3.8)	8	3.3363×10^7	848.33	131.52
SM1	Classical Smith-Martin model; (3.9)	3	2.3748×10^9	1107.04	390.23
SM2	SM1 with different divided/undivided rates; (3.10)	6	1.5218×10^9	1085.00	368.19
SM3	SM2 with lognormal recruitment into first division; (3.11)	8	2.1772×10^7	821.44	104.63
SM4	SM3 with linearly increasing death	9	2.0506×10^7	819.66	102.85
SM5	SM3 with linearly increasing cycle length	9	2.1116×10^7	821.51	104.70
Cyton1	Lognormal distributions, no death for divided cells	6	3.8969×10^7	854.11	137.30
Cyton2	Cyton1 with a progressor fraction	7	3.0204×10^7	840.06	123.25
Cyton3	Cyton1 with lognormal death for undivided cells	8	4.1363×10^6	716.81	–
Cyton4	Cyton 3 with a progressor fraction	9	4.1363×10^6	718.81	2.00

TABLE 2. Summary of mathematical models for describing cell numbers per generation. The OLS cost and Akaike Information Criteria (AIC) [32] are provided, as well as the AIC differences (difference between an AIC value and the smallest AIC value of all the models tested). The model with the minimum AIC value is considered the information-theoretic best model. Models with AIC differences greater than 10 (hence all models other than the cyton model) are considered to be significantly inferior based upon the given data set. The results above are consistent with those of [47].

Let $n(t,x)$ be the structured density (cell per unit of fluorescence intensity) of a population of cells at time t and with measured FI x. Then this population density can be described by the PDE

$$\begin{aligned} \frac{\partial n(t,x)}{\partial t} - ce^{-kt}\frac{\partial[(x-x_a)n(t,x)]}{\partial x} &= -(\alpha(t,x)+\beta(t,x))n(t,x)\chi_{[x_a,x^*]}4\alpha(t,2x-x_a)n(t,2x-x_a) \\ n(0,x) &= \Phi(x) \\ n(t,x_{\max}) &= 0 \\ v(t,x_a)n(t,x_a) &= 0. \end{aligned} \tag{3.15}$$

The functions $\alpha(t,x)$ and $\beta(t,x)$ are the rates of cell division and death, respectively, in units 1/hr. The motivating idea for the structural dependence of these rates is that the dilution of CFSE by division allows one to relate measured FI to the number of divisions undergone. Thus the structural dependence of the division and death rate functions is a surrogate for division dependence. The parameter x_a accounts for the autofluorescence or natural brightness of unlabeled cells. The advection term in the equation above (with parameters c and k) models the slow decay of CFSE FI (as a result of the natural turnover of intracellular proteins within the cell) using a Gompertz decay process [21]. These two features (autofluorescence and label decay) are important mathematical structures describing the manner in which the intracellular dye is processed by cells. The density $n(t,x)$ can then be used to compute cell counts in terms of measured FI, to be compared directly to histogram data such as Figure 2.

While the model (3.15) has been shown to very accurately fit CFSE histogram data [21,22], it cannot be used to compute quantities such as the number of cells having undergone a given number of divisions. As such, it is not (in the form above) directly comparable to the methods previously discussed. Also,

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

though the structural dependence of the proliferation and death rate functions can be used to infer division-dependent characteristics [21,79], this interpretation is neither straightforward nor intuitive [74]. In order to alleviate these difficulties, a simple reformulation [23, 46, 74, 79] of the model (3.15) can be used to describe the structured densities of subpopulations of cells corresponding to distinct generations, which interact through division. Let $n_i(t,x)$ be the structured density of a subpopulation of cells at time t and measured FI x and having undergone i divisions. Then the subpopulations are described by the system

$$\begin{aligned}
\frac{\partial n_0}{\partial t} - ce^{-kt}(x-x_a)\frac{\partial n_0}{\partial x} &= -(\alpha_0(t)+\beta_0(t)-ce^{-kt})n_0(t,x)\\
\frac{\partial n_1}{\partial t} - ce^{-kt}(x-x_a)\frac{\partial n_1}{\partial x} &= -(\alpha_1(t)+\beta_1(t)-ce^{-kt})n_1(t,x)+R_1(t,x)\\
&\vdots\\
\frac{\partial n_{i_{\max}}}{\partial t} - ce^{-kt}(x-x_a)\frac{\partial n_{i_{\max}}}{\partial x} &= -(\beta_{i_{\max}}(t)-ce^{-kt})n_{i_{\max}}(t,x)+R_{i_{\max}}(t,x),
\end{aligned} \tag{3.16}$$

with boundary conditions given as in (3.15). The recruitment term, which as above assumes symmetric label division at mitosis, is given by $R_i(t,x) = 4\alpha_{i-1}(t)n_{i-1}(t,2x-x_a)$. It is assumed that all cells are undivided at $t=0$, with some initial FI $n_0(0,x)=\Phi(x)$. Several parameterizations of the division and death rate functions $\alpha_i(t)$ and $\beta_i(t)$ have been proposed [23], with best-fit results obtained when the division rates are piecewise-linear functions of time and the death rates are constants, with both division and death rates depending upon the number of divisions undergone. It has also been shown [23, 79] that variability (among all cells in the population) in the autofluorescence parameter x_a is a significant physical feature of an accurate mathematical model and can be accurately modeled with a lognormal distribution [23, 46, 79]. The best-fit solution to model (3.16) for a particular data set (data originally from [56]) is shown in Figure 8 and demonstrates the suitability of the given model in describing data from CFSE-based proliferation assays.

A slightly different form of the model (3.16) has also been formulated in [46, 74] and this leads to significant computational benefits. While the model (3.16) uses the structure variable x as measured fluorescence intensity (which is the variable observed by a flow cytometer), it is also possible to derive a model in terms of the fluorescence intensity *resulting from CFSE* (that is, in the absence of autofluorescence). In this case, x is a mass-like quantity. Again, assuming symmetric division, the resulting model is

$$\begin{aligned}
\frac{\partial n_0}{\partial t} - ce^{-kt}\frac{\partial x n_0}{\partial x} &= -(\alpha_0(t)+\beta_0(t))n_0(t,x)\\
\frac{\partial n_1}{\partial t} - ce^{-kt}\frac{\partial x n_1}{\partial x} &= -(\alpha_1(t)+\beta_1(t))n_1(t,x)+4\alpha_{i-1}(t)n_{i-1}(t,2x)\\
&\vdots
\end{aligned} \tag{3.17}$$

The boundary conditions and initial condition are again given as in (3.15) and (3.16). The major advantage of this slight reformulation is the manner of solution. In keeping with the notation used in previous sections, define the initial and subpopulation cell numbers,

$$\begin{aligned}
N_0 &= \int_0^\infty \Phi(x)dx\\
N_0(t) &= \int_0^\infty n_0(t,x)dx\\
N_1(t) &= \int_0^\infty n_1(t,x)dx\\
&\vdots
\end{aligned}$$

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

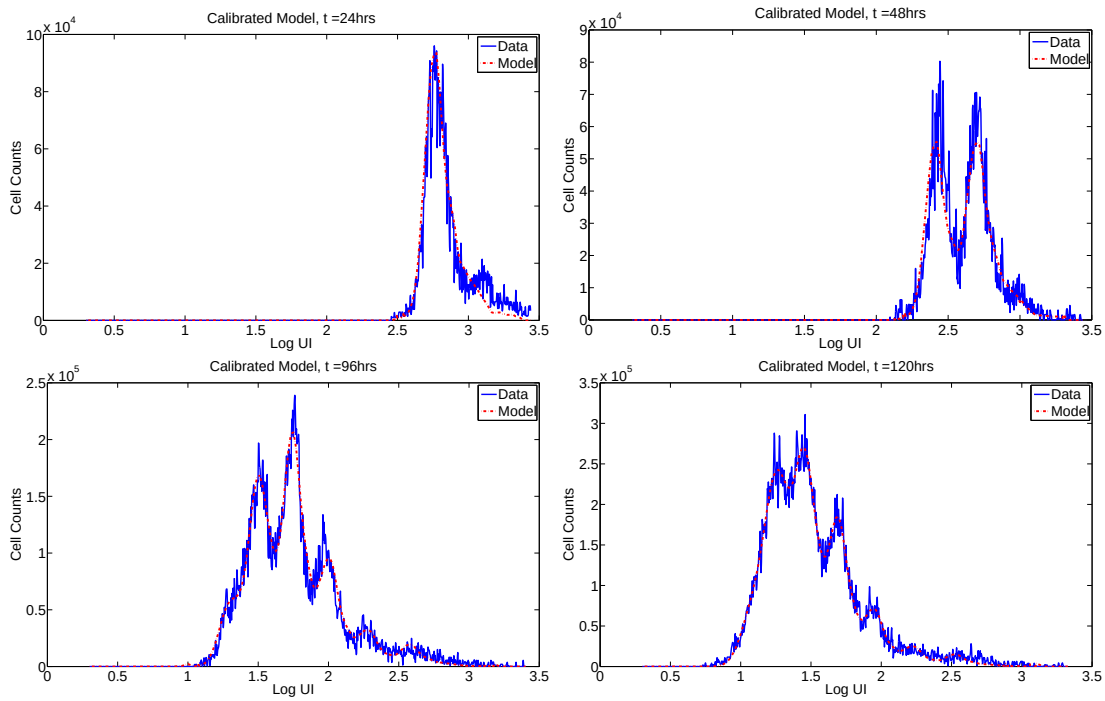


FIGURE 8. Best-fit solution for model (3.16) to a particular data set. Data originally from [56].

Then it can be shown [46, 74] that the solution to model (3.17) is given by

$$n_i(t,x) = N_i(t)\bar{n}_i(t,x) \tag{3.18}$$

for all i, where the functions $N_i(t)$ satisfy the ODE system

$$\begin{aligned} \frac{dN_0}{dt} &= -(\alpha_0(t)+\beta_0(t))N_0(t) \\ \frac{dN_1}{dt} &= -(\alpha_1(t)+\beta_1(t))N_1(t) + 2\alpha_{i-1}(t)N_{i-1}(t) \\ &\vdots \end{aligned} \tag{3.19}$$

with initial conditions $N_0(0) = N_0$, $N_i(0) = 0$ for all $i \geq 1$, and the functions $\bar{n}_i(t,x)$ satisfy the PDE

$$\frac{\partial \bar{n}_i(t,x)}{\partial t} - ce^{-kt}\frac{\partial [x\bar{n}_i(t,x)]}{\partial x} = 0 \tag{3.20}$$

with initial condition

$$\bar{n}_i(0,x) = \frac{2^i \Phi(2^i x)}{N_0}.$$

Equation (3.20) can be solved analytically by the method of characteristics to obtain

$$\bar{n}_i(t,x) = 2^i \exp\left(-\frac{c}{k}(1-e^{-kt})\right)\frac{\Phi\left(2^i \exp\left(\frac{c}{k}(1-e^{-kt})\right)x\right)}{N_0}.$$

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

A detailed proof that (3.18) is a solution of (3.17) can be found in [46, 74]. While the equations above assume a Gompertz decay process for intracellular label turnover, more general results are given in [46]. This remarkable result significantly reduces the computational burden of computing a forward solution by replacing a coupled system of PDEs (describing both cell numbers and label dynamics) with a coupled system of ODEs (describing only cell numbers) and an uncoupled system of PDEs (describing label dynamics). In many cases (as above) the system of PDEs can be solved analytically [46], so that the computational burden is reduced to the much simpler task of computing a numeric solution for an ODE system. Of course, one must relate the model solutions $n_i(t, x)$ (in which x is fluorescence intensity *resulting from CFSE*) to the actual measured CFSE profiles (which include the contribution of autofluorescence); this can be done via convolution with a density kernel representing the distribution of AutoFI in the population, and fast approximation techniques have been established. Appropriate numerical approximation and convergence results have also been demonstrated in [46].

In addition to the speed with which a solution can be computed, the model formulation (3.17) and its associated method of solution are premised upon mathematically decoupling the mechanisms of cell division, intracellular processing of label mass, and measurement by flow cytometry. The result of [46] is a more intuitive model formulation which can be viewed as a unifying framework linking models of cell numbers (described in Sections 3.1–3.4) and models of label dynamics (Section 3.5). Ideas from [23, 46] have recently been combined with the cyton formulation [33, 47] (see also Section 3.4) to yield a conservation-based probabilistic model that provides an excellent fit to histogram data [24].

4. Concluding Remarks

This review began with an overview of the complex experimental procedure which is used to study the manner in which cells of the immune system respond to stimulation. Several significant issues regarding mathematical and statistical aspects of the data collection procedure were addressed there, but were subsequently given little attention as several mathematical models of dividing cell dynamics were discussed. The single assumption common to every mathematical model of cell dynamics is the treatment of flow cytometry measurements as a census of a single population of cells. As discussed previously, this assumption is reliable for most data sets and is perhaps unavoidable. Somewhat surprisingly there have been few detailed analyses of the mathematical and statistical implications of this most fundamental assumption.

The most significant implication of assuming a single population is being measured repeatedly is that subtle variations in the populations of distinct wells are indistinguishable from fluctuations in the total number of cells in the population at a given time. The same problem arises from using a small sample of each well and then scaling (which is subject to some error) to estimate the total population. For instance, one can end up with more cells in a population at a given measurement time than would be physically possible given the preceding measurement [21, 35, 51, 79]. Of equal importance, parameters of a mathematical model (rates of cell death, in particular) may be incorrectly inferred, or confidence intervals on parameters may be too small if such issues are not taken into account. Some effort has been directed toward formulating an accurate statistical model of data from CFSE based proliferation assays, both for cell numbers computed from histogram data [84] and for the histogram data itself [50, 79].

In Sections 3.1–3.4, numerous mathematical models are summarized which describe the dynamics of a division-structured population of cells. Each of these models has been fit to a typical cell count data set (Table 1) and the results are summarized in Table 2, along with the associated AIC values for model comparison. This was done with recognition that the AIC is only a suggestive statistic and conditions [32] for AIC to be a rigorous approximation are not met. Nonetheless, based upon this table, the cyton model is strongly supported as providing the best description of cell division dynamics. On one hand, this is not surprising as the cyton model can be considered a generalization of the other models tested. In spite of this generality, the cyton model is based upon a simple premise (times to division and death) which is readily relatable to biologically meaningful parameters. On the other hand, care must be taken to

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

avoid overreaching conclusions given only a single data set, in addition to the uncertainties just discussed regarding the statistical model and the AIC criteria.

In Section 3.5, an alternative approach is discussed in which the flow cytometry histogram data is modeled directly. The label-structured model (3.15) has been shown to provide an excellent fit to histogram data [21, 79] and the model has been generalized (3.16) to permit the direct computation of cell numbers and division-specific division and death rates [23, 46, 74, 79]. Though some comparison of the two different modeling approaches (describing cell numbers and fitting histogram data) has been done [56], it has not been immediately clear how to directly compare the distinct approaches in an impartial manner. Fortunately, the recent work on a 'division and label-structured population model' [46, 74] has demonstrated how model (3.17) can be viewed as a unifying link between models of cell numbers and models of label dynamics (and thus histogram data). The incorporation of alternative mathematical treatments of cell division rates (such as those summarized in Table 2) into a label-structured framework for direct comparison to histogram data is an open problem which may help to improve the biological interpretation of existing label structured models such as [21, 23]. For instance, time-dependence of the exponential rates of division (the $\alpha_i(t)$ in model (3.16)) has been found to be an important mathematical feature of an accurate model of cell dynamics [23, 79]. However, these time-dependent exponential rates are not nearly as intuitive as, say, the cyton model of cell division. These cyton ideas have been combined in [24] with the formulations in [23, 46] to produce a probabilistic model for use with histogram data. The results indicate that such a combination is an excellent framework in which to consider the future studies mentioned below. Additional model generalizations might also be considered. For instance, the Smith-Martin model has also been generalized to allow for division-linked differentiation and the inheritance of division times [65]. Branching process models (which can be thought of as generalizations of the probabilistic cyton model [51, 77]) have been used to study the effects of correlations between the division characteristics of closely related cells (siblings, cousins, etc.) [38, 51, 77, 82] and the inheritance of parameters [65], both of which have been observed in experimental settings [49].

Because the statistical model is of vital importance for meaningful model comparison, it will be difficult to reach strong conclusions regarding mathematical models of cell cycle dynamics until an accurate statistical model is developed and validated. Given such numerous possibilities for alternative mathematical descriptions of the underlying physical process, an accurate statistical model of the data is a necessity for an information-theoretic comparison (such as the AIC, the TIC, etc., [32]) of mathematical models. In the end, the ultimate goal of any mathematical model of cell dynamics is to link experimental data with model parameters which are not only *consistently* identifiable from the data but which are also biologically intuitive. Such models might then be used to examine how changes in experimental conditions can be linked to cell behavior, thus building an understanding of stimulatory and costimulatory processes at the cellular level (cf., [37, 43]). Significantly, the models (3.16) and (3.17) are not limited to a specific cell type or stimulation condition. A validated mathematical model might also be used to aid [15] in the design of experiments (e.g., to determine when to take measurements in a manner that maximizes parameter information while minimizing the amount of blood needed for the experiment).

It is the authors' belief that mathematics, by organizing information in an insightful quantitative fashion, has a significant and positive role to play when properly used in developing a comprehensive understanding of the immune system. It is hoped that the issues raised here and the mathematical models surveyed can be used to organize and direct future efforts in mathematical modeling of populations of dividing cells for eventual use in pathogenesis modeling, immunosuppression/control, etc.

Acknowledgements. This research was supported in part by Grant Number NIAID R01AI071915-09 from the National Institute of Allergy and Infectious Disease. The authors are most grateful to their collaborators from Barcelona, Andreas Meyerhans, Jordi Argilaguet and Cristina Peligero for their patience during numerous stimulating and instructive conversations which have helped us understand to some degree the complicated data resulting from CFSE-labeled cells in proliferation as analyzed by flow cytometry. As acknowledged above, they have also freely provided us with material for this article. We are also grateful to Gena Bocharov, for stimulat-

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

ing comments during earlier collaborations, and indeed for suggesting that we should attempt to put together a modeling "review" that resulted in this manuscript.

References

[1] O. Arino, E. Sanchez, G.F. Webb. *Necessary and sufficient conditions for asynchronous exponential growth in age structured cell populations with quiescence*. Mathematical Analysis and Applications, 215(1997), 499–513.

[2] B. Asquith, C. Debacq, A. Florins, N. Gillet, T. Sanchez-Alcaraz, A. Mosley, L. Willems. *Quantifying lymphocyte kinetics in vivo using carboxyfluorein diacetate succinimidyl ester*. Proc. R. Soc. B, 273 (2006), 1165–1171.

[3] H.T. Banks. *A Functional Analysis Framework for Modeling, Estimation and Control in Science and Engineering*. CRC Press/Taylor-Francis, Boca Raton London New York, 2012.

[4] H.T.Banks, V. A. Bokil, S. Hu, F.C.T. Allnutt, R. Bullis, A.K. Dhar, C.L. Browdy, *Shrimp biomass and viral infection for production of biological countermeasures, CRSC-TR05-45*. North Carolina State University, December 2005; Mathematical Biosciences and Engineering, 3 (2006), 635–660.

[5] H.T. Banks, D.M. Bortz, S.E. Holte, *Incorporation of variability into the mathematical modeling of viral delays in HIV infection dynamics*, Math. Biosciences. 183 (2003), 63–91.

[6] H.T. Banks, D. M. Bortz, G.A. Pinter, L.K. Potter. *Modeling and imaging techniques with potential for application in bioterrorism*. CRSC-TR03-02, North Carolina State University, January 2003; Chapter 6 in Bioterrorism: Mathematical Modeling Applications in Homeland Security, (H.T. Banks and C. Castillo-Chavez, eds.), Frontiers in Applied Math, FR28, SIAM, Philadelphia, PA, 2003, 129–154.

[7] H.T. Banks, L.W. Botsford, F. Kappel, C. Wang, *Modeling and estimation in size structured population models*. LCDS/CSS Report 87-13, Brown University, March 1987; Proc. 2nd Course on Math. Ecology (Trieste, December 8-12, 1986), World Scientific Press, Singapore, 1988, 521–541.

[8] H.T. Banks, F. Charles, M. Doumic, K. L. Sutton, W. C. Thompson. *Label structured cell proliferation models*. CRSC-TR10-10, North Carolina State University, June 2010; Appl. Math. Letters, 23 (2010), 1412–1415.

[9] H.T. Banks, J.L. Davis, S.L. Ernstberger, S. Hu, E. Artimovich, A.K. Dhar, C.L. Browdy. *A comparison of probabilistic and stochastic differential equations in modeling growth uncertainty and variability*. CRSC-TR08-03, North Carolina State University, February 2008; *Journal of Biological Dynamics*, 3 (2009), 130–148.

[10] H.T. Banks, J.L. Davis, S. Hu, *A computational comparison of alternatives to including uncertainty in structured population models*. CRSC-TR09-14, North Carolina State University June 2009; in *Three Decades of Progress in Systems and Control*, X. Hu, U. Jonsson, B. Wahlberg, B. Ghosh (Eds.), Springer, 2010, 19–33.

[11] H.T. Banks, B.F. Fitzpatrick. *Estimation of growth rate distributions in size-structured population models*. CAMS Tech. Rep. 90-2, Univ. of Southern California, January 1990; Quart. Appl. Math., 49 (1991), 215–235.

[12] H.T. Banks, B.G. Fitzpatrick, L.K. Potter, Y. Zhang, *Estimation of probability distributions for individual parameters using aggregate population observations*. CRSC-TR98-06, North Carolina State University, January 1998; Stochastic Analysis, Control, Optimization and Applications (W.McEneaney, G. Yin, and Q. Zhang, eds.), Birkhauser, 1998, 353–371.

[13] H.T. Banks, N.L. Gibson, *Well-posedness in Maxwell systems with distributions of polarization relaxation parameters*. CRSC-TR04-01, North Carolina State University, January 2004; Applied Math. Letters, 18 (2005), 423–430.

[14] H.T. Banks, N.L. Gibson, *Electromagnetic inverse problems involving distributions of dielectric mechanisms and parameters*. CRSC-TR05-29, North Carolina State University, August2005; Quarterly of Applied Mathematics, 64 (2006), 749–795.

[15] H.T. Banks, K. Holm, F. Kappel, *Comparison of optimal design methods in inverse problems*. CRSC-TR10-11, North Carolina State University, May 2011; Inverse Problems, 27 (2011), 075002.

[16] H.T. Banks, S. Hu. *Nonlinear stochastic Markov processes and modeling uncertainty in populations*. CRSC-TR11-02, North Carolina State University, January 2011; *Mathematical Bioscience and Engineering*, 9 (2012), 1–25.

[17] H.T. Banks, S. Hu. *Uncertainty propagation in physiologically structured population models*. CRSC-TR12-08, North Carolina State University, Raleigh, NC, March 2012; Journal on Mathematical Modelling of Natural Phenomena, submitted.

[18] H.T. Banks, K. Kunisch. *Estimation Techniques for Distributed Parameter Systems*, Birkhauser, Boston, 1989.

[19] H.T. Banks and G.A. Pinter. *A probabilistic multiscale approach to hysteresis in shear wave propagation in biotissue*. CRSC-TR04-03, North Carolina State University, January 2004; IAM J. Multiscale Modeling and Simulation, 3 (2005), 395–412.

[20] H.T. Banks, L.K. Potter. *Probabilistic methods for addressing uncertainty and variability in biological models: Application to a toxicokinetic model*. CRSC-TR02-27, North Carolina State University, September 2002; Math. Biosci., 192 (2004), 193–225.

[21] H.T. Banks, Karyn L. Sutton, W. Clayton Thompson, G. Bocharov, Marie Doumic, Tim Schenkel, Jordi Argilaguet, Sandra Giest, Cristina Peligero, Andreas Meyerhans. *A New Model for the Estimation of Cell Proliferation Dynamics Using CFSE Data*. CRSC-TR11-05, North Carolina State University, Revised July 2011; J. Immunological Methods, 373 (2011), 143–160; DOI:10.1016/j.jim.2011.08.014.

[22] H.T. Banks, Karyn L. Sutton, W. Clayton Thompson, G. Bocharov, D. Roose, T. Schenkel, A. Meyerhans. *Estimation of cell proliferation dynamics using CFSE data*. CRSC-TR09-17, North Carolina State University, August 2009; Bull. Math. Biol., 70 (2011), 116–150.

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

[23] H.T. Banks, W. Clayton Thompson, Cristina Peligero, Sandra Giest, Jordi Argilaguet, Andreas Meyerhans. *A Division-Dependent Compartmental Model for Computing Cell Numbers in CFSE-based Lymphocyte Proliferation Assays*. CRSC-TR12-03, North Carolina State University, January 2012; Math Biosci. Eng., to appear.

[24] H.T. Banks, W. Clayton Thompson. *A division-dependent compartmental model with cyton and intracellular label dynamics*. CRSC-TR12-12, North Carolina State University, May 2012; Intl. J. Pure and Appl. Math 77 (2012), 119–147.

[25] H.T. Banks, H.T. Tran, D.E. Woodward. *Estimation of variable coefficients in the Fokker-Planck equations using moving node finite elements*. *SIAM J. Numer. Anal.*, 30 (1993), 1574–1602.

[26] B. Basse, B. Baguley, E. Marshall, G. Wake, D. Wall. *Modelling the flow cytometric data obtained from unperturbed human tumour cell lines: Parameter fitting and comparison*. Bull. Math. Biol., 67 (2005), 815–830.

[27] F. Bekkal Brikci, J. Clairambault, B. Ribba, B. Perthame. *An age-and-cyclin-structured cell population model for healthy and tumoral tissues*. Math. Biol., 57 (2008), 91–110.

[28] G. Bell, E. Anderson. *Cell Growth and Division I. A Mathematical Model with Applications to Cell Volume Distributions in Mammalian Suspension Cultures*, Biophysical Journal, 7 (1967), 329–351.

[29] S. Bernard, L. Pujo-Menjouet, M.C. Mackey. *Analysis of cell kinetics using a cell division marker: Mathematical modeling of experimental data*. *Biophysical Journal*, 84 (2003), 3414–3424.

[30] S. Bonhoeffer, H. Mohri, D. Ho, A.S. Perelson. *Quantification of cell turnover kinetics using 5-Bromo-2'-deoxyuridine*. Immunology, 64 (2000), 5049–5054.

[31] Jose A. M. Borghans, R.J. de Boer. *Quantification of T-cell dynamics: from telomeres to DNA labeling*. Immunological Reviews, 216 (2007), 35–47.

[32] K.P. Burnham, D.R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd Edition), Springer, New York, 2002.

[33] R. Callard, P.D. Hodgkin. *Modeling T- and B-cell growth and differentiation*. Immunological Reviews, 216 (2007), 119–129.

[34] "Cyton Calculator", Walter and Eliza Ball Institute of Medical Research. Available Online. Accessed 16 March 2012. http://www.wehi.edu.au/faculty_members/research_projects/cyton_calculator

[35] R.J. DeBoer, V.V. Ganusov, D. Milutinovic, P.D. Hodgkin, A.S. Perelson. *Estimating lymphocyte division and death rates from CFSE data*. *Bull. Math. Biol.*, 68 (2006), 1011–1031.

[36] R.J. DeBoer, A. S. Perelson. *Estimating division and death rates from CFSE data*. Comp. and Appl. Mathematics, 184 (2005), 140–164.

[37] E.K. Deenick, A.V. Gett, P.D. Hodgkin. *Stochastic model of T cell proliferation: a calculus revealing IL-2 regulation of precursor frequencies, cell cycle time, and survival*. Immunology, 170 (2003), 4963–4972.

[38] K. Duffy, V. Subramanian. *On the impact of correlation between collaterally consanguineous cells on lymphocyte population dynamics*. Math. Biol., 59 (2009), 255–285.

[39] J.Z. Farkas. *Stability conditions for the non-linear McKendrick equations*. Appl. Math. and Comp., 156 (2004), 771–777.

[40] J.Z. Farkas. *Stability conditions for a non-linear size-structured model* Nonlinear Analysis: Real World Applications, 6 (2005), 962–969.

[41] V. V. Ganusov, D. Milutinovi, R. J. De Boer. *IL-2 regulates expansion of CD4+ T cell populations by affecting cell death: insights from modeling CFSE data*. Immunology, 179 (2007), 950–957.

[42] V.V. Ganusov, S.S. Pilyugin, R.J. De Boer, K. Murali-Krishna, R. Ahmed, R. Antia. *Quantifying cell turnover using CFSE data. Immunological Methods*, 298 (2005), 183–200.

[43] A.V. Gett, P.D. Hodgkin. *A cellular calculus for signal integration by T cells*. Nature Immunology, 1 (2000), 239–244.

[44] M. Gyllenberg, G.F. Webb. *Age-size structure in populations with quiescence*. Mathematical Biosciences, 86 (1987), 67–95.

[45] M. Gyllenberg, G.F. Webb. *A nonlinear structured population model of tumor growth with quiescence*. J. Math. Biol., 28 (1990), 671–694.

[46] J. Hasenauer, D. Schittler, F. Allgöwer. *A computational model for proliferation dynamics of division- and label-structured populations*. `arXive.org`, arXiv:1202.4923v1,22Feb,2012.

[47] E.D. Hawkins, Mirja Hommel, M.L Turner, F. Battye, J Markham, P.D Hodgkin. *Measuring lymphocyte proliferation, survival and differentiation using CFSE time-series data*. Nature Protocols, 2 (2007), 2057–2067.

[48] E.D. Hawkins, M.L. Turner, M.R. Dowling, C. van Gend, P.D. Hodgkin. *A model of immune regulation as a consequence of randomized lymphocyte division and death times*. Proc. Natl. Acad. Sci., 104 (2007), 5032–5037.

[49] E.D. Hawkins, J.F. Markham, L.P. McGuinness, P.D. Hodgkin. *A single-cell pedigree analysis of alternative stochastic lymphocyte fates. Proc. Natl. Acad. Sci.*, 106 (2009), 13457–13462.

[50] O. Hyrien, M.S. Zand. *A mixture model with dependent observations for the analysis of CFSE-labeling experiments*. American Statistical Association, 103 (2008), 222–239.

[51] O. Hyrien, R. Chen, M.S. Zand. *An age-dependent branching process model for the analysis of CFSE-labeling experiments*. Biology Direct, 5 (2010), Published Online.

[52] H.Y. Lee, E.D. Hawkins, M.S. Zand, T. Mosmann, H. Wu, P.D. Hodgkin, A.S. Perelson. *Interpreting CFSE obtained division histories of B cells in vitro with Smith-Martin and Cyton type models*. Bull. Math. Biol., 71 (2009), 1649–1670.

[53] H.Y. Lee, A.S. Perelson. *Modeling T cell proliferation and death in vitro based on labeling data: generalizations of the Smith-Martin cell cycle model*. Bull. Math. Biol., 70 (2008), 21–44.

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press

[54] K. Leon, J. Faro, J. Carneiro. *A general mathematical framework to model generation structure in a population of asynchronously dividing cells*. Theoretical Biology, 229 (2004), 455–476.

[55] T. Luzyanina, D. Roose, G. Bocharov. *Distributed parameter identification for a label-structured cell population dynamics model using CFSE histogram time-series data*. Math. Biol., 59 (2009), 581–603.

[56] T. Luzyanina, D. Roose, T. Schenkel, M. Sester, S. Ehl, A. Meyerhans, G. Bocharov. *Numerical modelling of label-structured cell population growth using CFSE distribution data*. Theoretical Biology and Medical Modelling, 4 (2007), Published Online.

[57] A.B. Lyons. *Divided we stand: tracking cell proliferation with carboxyfluorescein diacetate succinimidyl ester*. *Immunology and Cell Biology*, 77 (1999), 509–515.

[58] A.B. Lyons, J. Hasbold, P.D. Hodgkin. *Flow cytometric analysis of cell division history using diluation of carboxyfluorescein diacetate succinimidyl ester, a stably integrated fluorescent probe*. Methods in Cell Biology, 63 (2001), 375–398.

[59] A.B. Lyons, K. V. Doherty. *Flow cytometric analysis of cell division by dye dilution*. Current Protocols in Cytometry, (2004), 9.11.1-9.11.10.

[60] A.B. Lyons, C.R. Parish. *Determination of lymphocyte division by flow cytometry*. Immunol. Methods, 171 (1994), 131–137.

[61] G. Matera, M. Lupi, P. Ubezio. *Heterogeneous cell response to topotecan in a CFSE-based proliferative test*. Cytometry A, 62 (2004), 118–128.

[62] J.A. Metz, O. Diekmann. *The Dynamics of Physiologically Structured Populations*. Springer Lecture Notes in Biomathematics 68, Heidelberg, 1986.

[63] H. Miao, X. Jin, A. Perelson, H. Wu. *Evaluation of multitype mathemathematical modelsfor CFSE-labeling experimental data*. Bull. Math. Biol., 74 (2012), 300–326; DOI 10.1007/s11538-011-9668-y.

[64] K. Murphy, *aneway's Immunobiology*, 8^{th} Edition. Garland Science, London New York, 2012.

[65] R. E. Nordon, Kap-Hyoun Ko, R. Odell, T. Schroeder. *Multi-type branching models to describe cell differentiation programs*. *Theoretical Biology*, 277 (2011), 7–18.

[66] R.E. Nordon, M. Nakamura, C. Ramirez, R. Odell. *Analysis of growth kinetics by division tracking*. Immunology and Cell Biology, 77 (1999), 523–529.

[67] C. Parish. *Fluorescent dyes for lymphocyte migration and proliferation studies*. Immunology and Cell Biol., 77 (1999), 499–508.

[68] B. Perthame. *Transport Equations in Biology*. Birkhauser Frontiers in Mathematics, Basel, 2007.

[69] S. S. Pilyugin, V. V. Ganusov, K. Murali-Krishnac, R. Ahmed, R. Antia. *The rescaling method for quantifying the turnover of cell populations*. Theoretical Biology, 225 (2003), 275–283.

[70] B.J.C. Quah, C.R. Parish. *New and improved methods for measuring lymphocyte proliferation* in vitro *and* in vivo *using CFSE-like fluorescent dyes*. Immunological Methods, (2012), to appear.

[71] B. Quah, H. Warren, C. Parish. *Monitoring lymphocyte proliferation in vitro and in vivo with the intracellular fluorescent dye carboxyfluorescein diacetate succinimidyl ester*. Nature Protocols, 2 (2007), 2049–2056.

[72] P. Revy, M. Sospedra, B. Barbour, A. Trautmann. *Functional antigen-independent synapses formed between T cells and dendritic cells*. Nature Immunology, 2 (2001), 925–931.

[73] M. Roederer. *Interpretation of cellular proliferation data: Avoid the panglossian, Cytometry A*, 79 (2011), 95–101.

[74] D. Schittler, J. Hasenauer, F. Allgöwer. *A generalized model for cell proliferation: Integrating division numbers and label dynamics*. Proc. Eighth International Workshop on Computational Systems Biology (WCSB 2011), June 2001, Zurich, Switzerland, p. 165–168.

[75] J. Sinko, W. Streifer. *A New Model for Age-Size Structure of a Population*. Ecology, 48 (1967), 910–918.

[76] J.A. Smith, L. Martin. *Do Cells Cycle?* Proc. Natl. Acad. Sci., 70 (1973), 1263–1267.

[77] V.G. Subramanian, K.R. Duffy, M.L. Turner, P.D. Hodgkin. *Determining the expected variability of immune responses using the cyton model*. Math. Biol., 56 (2008), 861–892.

[78] H. Veiga-Fernandez, U. Walter, C. Bourgeois, A. McLean, B. Rocha. *Response of naive and memory CD8+ T cells to antigen stimulation in vivo*, *Nature Immunology*. 1 (2000), 47–53.

[79] W. C. Thompson. *Partial Differential Equation Modeling of Flow Cytometry Data from CFSE-based Proliferation Assays*. Ph.D. Dissertation, Dept. of Mathematics, North Carolina State University, Raleigh, December, 2011.

[80] P.K. Wallace, J.D. Tario, Jr., J.L. Fisher, S.S. Wallace, M.S. Ernstoff, K.A. Muirhead. *Tracking antigen-driven responses by flow cytometry: monitoring proliferation by dye dilution*. Cytometry A, 73 (2008), 1019–1034.

[81] H. S. Warren. *Using carboxyfluorescein diacetate succinimidyl ester to monitor human NK cell division: Analysis of the effect of activating and inhibitory class I MHC receptors*. Immunology and Cell Biology, 77 (1999), 544–551.

[82] C. Wellard, J. Markham, E.D. Hawkins, P.D. Hodgkin. *The effect of correlations on the population dynamics of lymphocytes*. Theoretical Biology, 264 (2010), 443–449.

[83] J.M. Witkowski. *Advanced application of CFSE for cellular tracking*. Current Protocols in Cytometry, 44 (2008), 9.25.1–9.25.8.

[84] A. Yates, C. Chan, J. Strid, S. Moon, R. Callard, A.J.T. George, J. Stark. *Reconstruction of cell population dynamics using CFSE*. BMC Bioinformatics, 8 (2007), Published Online.

https://doi.org/10.1051/mmnp/20127504 Published online by Cambridge University Press