$$\mathcal{AP}$$

ijpam.eu

EVALUATING THE IMPORTANCE OF MITOTIC ASYMMETRY IN CYTON-BASED MODELS FOR FSE-BASED FLOW CYTOMETRY DATA

H.T. Banks[1] [§], D.F. Kapraun[2], Cristina Peligero[3],
Jordi Argilaguet[4], Andreas Meyerhans[5]

[1,2]Center for Research in Scientific Computation
North Carolina State University
Raleigh, NC 27695-8212, USA
[3,4,5]ICREA Infection Biology Lab
Department of Experimental and Health Sciences
Universitat Pompeu Fabra
Barcelona, 08003, SPAIN

Abstract: We carry out computational inverse problem investigations to determine the importance of allowing for unequal label allocation to daughter cells during mitosis. Parameter estimations are performed using previously developed label-structured partial differential equation models that utilize cytons to account for variability in times between cell divisions as well as times until cell death. These models are augmented to allow for asymmetric cell divisions through inclusion of an additional model parameter. We employ well-known model comparison tests to analyze CFSE-based flow cytometry data with respect to the importance of asymmetric division during proliferation of human CD4+ and CD8+ T cells.

Key Words: cell proliferation, flow cytometry, CFSE, partial differential equations, label-structured population dynamics, cell division number, inverse problems, asymmetric cell division

[§]Correspondence author

1. Introduction

The adaptive immune response constitutes a major component of the mammalian immune system's defense against invading pathogens. B cells and T cells, which are the two classes of lymphocytes that comprise the adaptive immune system, recognize invaders by specific cell surface receptors and exert their responses through humoral and cellular effector mechanisms (i.e., through antibodies and cytotoxic T cells). As the success of the adaptive immune system in overcoming a perceived threat depends on the capacity of lymphocytes to proliferate, the ability to accurately predict cell proliferation dynamics in the presence of specific environmental stimuli has important implications for human health research in areas such as the treatment and prevention of infectious disease and immunosuppression for patients receiving organ and tissue transplants. Our recent efforts [2, 3, 4, 6, 7] have focused on estimating cell proliferation parameters and quantifying uncertainty in such estimates, but thus far we have always utilized mathematical models that are based upon the assumption of symmetric cell division. Here we consider how the incorporation of asymmetric cell division, which has been described as "almost an axiom of cell biology" [22], into our division- and label-structured partial differential equation models affects their ability to accurately describe data from CFSE dilution assays.

Over the last half-century, many mathematical models have been proposed that attempt to describe the dynamics of a proliferating population of lymphocytes, but until fairly recently it has been a challenge to validate such models. The discovery of the intracellular dye carboxyfluorescein succinimidyl ester (CFSE) was a major milestone in overcoming this obstacle. CFSE was originally developed as a tool for labeling lymphocytes so that their movements within animal subjects could be tracked over many months [24], but subsequently researchers determined that the dye could also be used to monitor lymphocyte proliferation [19]. Also, through the use of fluorescently labeled antibodies specific to various lymphocyte surface markers, it is now possible to follow the proliferative behavior of specific types of lymphocytes [18].

For the interested reader, Kapraun provides a detailed explanation of how CFSE and flow cytometry can be used to produce cell proliferation data [16], but the basic idea behind flow cytometry cell proliferation experiments can be described as follows. Suppose we are able to label all cells in a population with a "dye" that enters the cells and becomes covalently linked to molecules within their cytoplasm. If we then stimulate the cells to divide, mitosis causes each "mother" cell to produce two "daughter" cells. Naturally, each of the daughter

cells receives a portion of the cytoplasm that originally belonged to the mother, and therefore each of the daughter cells also receives a portion of the dye that was contained in the mother. Thus, as cell proliferation continues, those cells that have undergone more divisions tend to have less dye than those that have undergone fewer divisions. This allows one to use the total dye content of a cell as an indicator of the number of divisions it has undergone. In CFSE-based flow cytometry experiments, CFSE is the "dye" used to label the cells and flow cytometry is the means by which one can measure the total dye content of individual cells.

A flow cytometer is an instrument that is capable of quickly measuring various characteristics of large numbers of individual cells. Using hydrodynamic focusing, the instrument records the fluorescent intensity of individual cells (and possibly other particles, such as fluorescent count-scaling beads) into a single file. The cells (and in our case also some of the aforementioned beads) then pass one at a time through an interrogation point, where they are excited with a laser. During this process, the flow cytometer measures the fluorescence intensity (FI) at various wavelengths for each cell in the sample and counts the number of beads in the sample. FI observed at wavelengths in the range 515 to 545 nm corresponds to light emitted by CFSE [18, 24], while FI observed at other wavelengths can indicate the presence of another labeling agent and can therefore be used to identify the type of cell. Because FI induced by CFSE varies directly with the mass of CFSE within a cell, the FI observed in the relevant range of wavelengths can be used as a surrogate for CFSE mass contained in a particular cell [18, 24].

The output of a CFSE-based flow cytometry experiment can be summarized using histograms such as those depicted in Figure 1. To construct these histograms, cells are placed into bins defined by ranges of CFSE FI. Because the measured FI numbers tend to vary over orders of magnitude during the course of typical multi-day experiment, it is common to use the base 10 logarithm of FI as in the figure. Note that each curve in the figure provides a summary of the information collected on a given day, and the abscissa and ordinate for each point on the curve correspond to the lower limit of a histogram bin and the number of cells belonging to that bin, respectively. Therefore, the total number of cells present in a well at a given point in time can be visualized as the total area under the curve corresponding to that time.

In previous publications [2, 3, 4], the authors have described methods for using CFSE-based flow cytometry data to determine T cell division and death rates (for cells having completed various numbers of divisions). In this manuscript, however, we focus on a new model in which the apportioning of
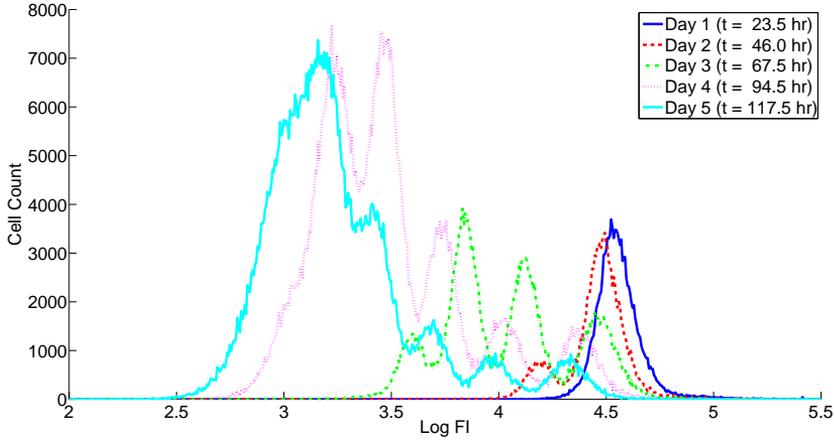
Figure 1: Histograms summarizing CFSE FI data at various time points.

CFSE during mitosis is *not* assumed to be even and consider whether such a model can describe one particular collection of data sets more accurately than a model based on symmetric (50-50) cell division. These data sets are comprised of CFSE-base flow cytometry information collected for two donors and two types of T cell (CD4+ and CD8+). In all cases, a special dye known as ViViD is utilized to exclude dead-but-not-disintegrated cells from cell counts. Data were collected at five time points corresponding to approximately 24, 48, 72, 96, and 120 hours after stimulation of the cells using a non-specific T cell mitogen. Full details of the experimental protocol and a precise explanation of the data we consider here are provided by Kapraun [16, Chapter 3].

2. Mathematical Model

Here, we summarize a new class of cell proliferation models that incorporates aspects of the cyton model of Hawkins et al. [14, 15] and the separable PDE models of Schittler et al. [20] and Hasenauer et al. [13], as well as the possibility of asymmetric cell divisions as proposed by Bocharov et al. [9] and Luzyanina et al. [17]. These models extend the cyton-based division- and label-structured PDE models previously used by Banks et al. [2, 3, 4] by including an additional parameter m that describes the degree to which cytoplasm apportionment during mitosis is asymmetric.

2.1. Basic Asymmetric Division- and Label-Structured Model for Cell Densities

Let $n_i(t, x)$ be a structured density (in cells per unit FI), where i is a whole number representing the number of divisions completed for a specific "genera-tion" of cells, t denotes the time elapsed (in hr) since some arbitrary starting time, and x denotes FI *induced by CFSE*. Also, let $\{\alpha_i(t)\}$, $\{\beta_i(t)\}$, and $v(t)$ denote exponential division rates, exponential death rates, and the CFSE ex-ponential decay rate, respectively (all in hr^{-1}). Finally, let $m \in (0, \frac{1}{2}]$ be an "asymmetric division parameter" so that the division of each mother cell re-sults in two daughter cells that receive proportions m and $1 - m$, respectively, of the cytoplasm (and CFSE) that was contained in the mother cell. Then the dynamics of a population of cells are described by

$$\frac{\partial}{\partial t}n_0(t, x) - v(t)\frac{\partial}{\partial x}\Big[xn_0(t, x)\Big] = -\Big(\alpha_0(t) + \beta_0(t)\Big)n_0(t, x) \qquad \text{(for } i = 0),$$

$$\frac{\partial}{\partial t}n_i(t, x) - v(t)\frac{\partial}{\partial x}\Big[xn_i(t, x)\Big] = -\Big(\alpha_i(t) + \beta_i(t)\Big)n_i(t, x) + R_i(t, x) \quad \text{for } i \geq 1,$$
$$(2.1)$$

where $x \geq 0$ and the "recruitment" terms are given by

$$R_i(t, x) = \frac{\alpha_{i-1}(t)}{m}n_{i-1}(t, \tfrac{1}{m}x) + \frac{\alpha_{i-1}(t)}{1 - m}n_{i-1}(t, \tfrac{1}{1-m}x) \qquad (2.2)$$

for $i \geq 1$. Note that when $m = \frac{1}{2}$, the expression for $R_i(t, x)$ given here simplifies such that the recruitment terms match those presented previously for the symmetric division models of Banks et al. [2, 3, 4]. The initial conditions are given by

$$n_i(t_0, x) = \begin{cases} \Phi(x) & \text{for } i = 0, \\ 0 & \text{for } i \geq 1, \end{cases} \qquad (2.3)$$

where t_0 indicates the time of the first observation and $\Phi(x)$ is the structured density for cells in the initial (undivided) population at that time. Kapraun [16] provides a detailed explanation of this model as well as a derivation from basic conservation principles. We remark here that t_0 typically coincides with the time at which the cells were stimulated to divide, but for the experimental data described in Section 1, the first observation actually occurred approximately 24 hours after stimulation.

As in the model of Schittler et al. [20], the solutions to the partial differential equations (PDEs) given in (2.1) can be factored as

$$n_i(t, x) = N_i(t)\bar{n}_i(t, x), \qquad (2.4)$$

where $N_i(t)$ indicates the *number* of cells having completed i divisions at time t and $\bar{n}_i(t,x)$ describes the *distribution* of CFSE FI within that generation of cells at time t; that is, $\bar{n}_i(t,x)$ is a probability density function (pdf) in the variable x, so that for any fixed t, $\bar{n}_i(t,x) \geq 0$ for all x and

$$\int_0^\infty \bar{n}_i(t,x)\,dx = 1.$$

We offer the following formal proposition of this factorability of solutions below. Kapraun [16] provides a proof of this proposition following the general approach of Schittler et al. [20].

Proposition 2.1. *Let $\{N_i(t)\}_{i=0}^\infty$ be a set of functions satisfying the system of weakly coupled ordinary differential equations (ODEs) given by*

$$
\begin{aligned}
\frac{dN_0(t)}{dt} &= -\Big(\alpha_0(t)+\beta_0(t)\Big)N_0(t) && \text{(for } i=0\text{),}\\
\frac{dN_i(t)}{dt} &= -\Big(\alpha_i(t)+\beta_i(t)\Big)N_i(t) + 2\alpha_{i-1}(t)N_{i-1}(t) && \text{for } i \geq 1,
\end{aligned}
\tag{2.5}
$$

and the initial conditions given by

$$
N_i(t_0) = \begin{cases} N_0 = \int_0^\infty \Phi(x)\,dx & \text{for } i=0,\\ 0 & \text{for } i \geq 1. \end{cases}
\tag{2.6}
$$

Also, let $\{\bar{n}_i(t,x)\}_{i=0}^\infty$ be a set of functions such that each $\bar{n}_i$ satisfies the PDE

$$
\frac{\partial \bar{n}_i(t,x)}{\partial t} - v(t)\frac{\partial[x\bar{n}_i(t,x)]}{\partial x} = 0
\tag{2.7}
$$

and the initial condition

$$
\bar{n}_i(t_0,x) = \begin{cases} \dfrac{\Phi(x)}{N_0} & \text{for } i=0,\\[2mm] \dfrac{1}{2m}\cdot\bar{n}_{i-1}(t_0,\tfrac{1}{m}x) + \dfrac{1}{2(1-m)}\cdot\bar{n}_{i-1}(t_0,\tfrac{1}{1-m}x) & \text{for } i \geq 1, \end{cases}
\tag{2.8}
$$

for all $x \geq 0$. Then the solution to (2.1) and (2.3) is given by (2.4) for $i \in \{0,1,2,\ldots\}$.

2.2. Autofluorescence

Thus far, we have described a model that accounts only for FI *induced by CFSE*, but as noted by Thompson [23], the experimentally measured FI of a cell is actually the sum of CFSE-induced FI and the cell's natural "autofluorescence". Therefore, following the work of Hasenauer et al. [13], we let $\tilde{n}_i(t, \tilde{x})$ be a structured density (in cells per unit FI), where i again denotes a specific generation of cells, t denotes time elapsed (in hr), and $\tilde{x}$ denotes *measured* FI. Here

$$\tilde{x} = x + x_a,$$

where x and x_a represent the FI due to CFSE content and cellular autofluorescence, respectively.

If we assume solutions $n_i(t, x)$ to (2.1) and (2.3) have already been computed and that x_a is a realization of a random variable X_a with pdf $f_{X_a}(x_a; t)$, then the densities $\tilde{n}_i(t, \tilde{x})$ can be computed using the convolution formula [10]

$$\tilde{n}_i(t, \tilde{x}) = \int_{-\infty}^{\infty} n_i(t, x) f_{X_a}(\tilde{x} - x; t) \, dx = \int_0^{\tilde{x}} n_i(t, x) f_{X_a}(\tilde{x} - x; t) \, dx \qquad (2.9)$$

as proposed by Hasenauer, Schittler, et al. [13, 20]. These authors demonstrate that, under certain assumptions, this convolution can be computed quickly and efficiently [13].

2.3. Division- and Label-Structured Cyton Model for Cell Densities

The cyton model [14, 15] is an alternative to (2.5) that arises from two simple assumptions. The first, which is self-evident, is that any given cell must eventually either divide or die. The second, which is based upon experimental evidence, is that the processes of cell division and death operate independently of one another [15]. Thus, we can assume that the destiny of any particular cell is governed by two fixed numbers: a "time until division" and a "time until death". In particular, the actual fate of the cell (division or death) can be determined by observing which of these two numbers is smaller. For an individual cell within a population of cells sharing similar characteristics (e.g., cells of the same type having undergone the same number of divisions), it is reasonable to assume that the "time until division" and "time until death" are realizations of two independent random variables. These random variables are described by probability distributions, and so the cyton model requires parameters that can be used to uniquely determine the probability distributions for

times until division and death of cells in a given population (e.g., CD4+ T cells having undergone 1 division). Hawkins et al. chose the term "cyton" to denote the "combination of independent cellular machines governing times to divide and die" and represented a cyton mathematically using a pair of probability density functions [15] . For example, if ϕ_i and ψ_i are the pdfs for time until division and time until death, respectively, of cells in generation i, then the cyton for generation i can be denoted (ϕ_i, ψ_i). One additional consideration is that, in reality, not all cells in a given population will divide if they avoid death (at least not within the time frame of a typical *in vitro* cell culturing experiment) [15]. Therefore, the cyton model includes the notion of "progressor fraction": for a given generation of cells, only a certain proportion have the potential to "progress" to the next generation via cell division.

In the cyton model, $n_i^{div}(t)$ and $n_i^{die}(t)$ are rates (in cells/hr) at which cells in generation i divide and die, respectively. The cytons $\{(\phi_i, \psi_i)\}$ are used to compute $\{n_i^{div}(t)\}$ and $\{n_i^{die}(t)\}$, which are in turn used to compute $\{N_i(t)\}$, the numbers of cells present in each generation i at any given time t. Banks et al. [2, 3, 4] and Kapraun [16] provide details of the cyton model, and use the same notation and symbols we use here.

As in earlier work [2, 3, 4], we incorporate the cyton model for cell numbers into the division- and label-structured model framework described previously by replacing the sink and source terms in the right-hand sides of (2.1) with terms involving the cyton-based rates $\{n_i^{div}(t)\}$ and $\{n_i^{die}(t)\}$ to obtain

$$\frac{\partial n_0(t,x)}{\partial t} - v(t)\frac{\partial[xn_0(t,x)]}{\partial x} = -\left(n_0^{div}(t) + n_0^{die}(t)\right)\bar{n}_0(t,x) \qquad \text{(for } i = 0\text{)},$$

$$\frac{\partial n_i(t,x)}{\partial t} - v(t)\frac{\partial[xn_i(t,x)]}{\partial x} = \left(2n_{i-1}^{div}(t) - n_i^{div}(t) - n_i^{die}(t)\right)\bar{n}_i(t,x) \quad \text{for } i \geq 1.$$
$$\tag{2.10}$$

Solutions of this system are then given by $n_i(t,x) = N_i(t)\bar{n}_i(t,x)$, where the $N_i(t)$'s satisfy the cyton model equations [16, Equation 2.11] and initial conditions (2.6) and each $\bar{n}_i(t,x)$ satisfies (2.7) and (2.8) as before. We state this claim as a proposition. As with the previous proposition, Kapraun [16] provides a proof of this proposition following the general approach of Schittler et al. [20].

Proposition 2.2.   *Let $\{N_i(t)\}_{i=0}^{\infty}$ be a set of functions satisfying the cyton model [16, Equation 2.11], where the initial condition N_0 is given by the relation shown in (2.6). Also, let $\{\bar{n}_i(t,x)\}_{i=0}^{\infty}$ be a set of functions such that each $\bar{n}_i$ satisfies the PDE (2.7) and the initial condition (2.8) for all $x \geq 0$. Then the solution to (2.10) with initial conditions (2.3) is given by (2.4) for $i \in \{0, 1, 2, \ldots\}$.*

It is worth noting that cells will only divide a finite number of times in the time frame of a typical *in vitro* cell culturing experiment. Therefore, one typically computes solutions to (2.10) only for $i \in \{0, 1, \ldots, i_{\max}\}$, where $i_{\max}$ is the largest number of divisions we expect a cell from the initial population to undergo during the period of observation. For a five-day experiment such as that described in Section 1, it is rare for cells to undergo more than 12 divisions. Because of the compounding expense of the recursive function calls that are required when computing numerical solutions for the asymmetric division model [16, Appendix B], we only perform computations for cells that have divided 8 or fewer times; i.e., we let $i_{\max} = 8$. Further explanation and justification of this choice are provided by Kapraun [16, Chapter 5].

Finally, we remark that (2.10) actually describes an entire class of models. In order to specify a particular model for further investigation, we must decide on forms for the distribution of the autofluorescence X_a, the (exponential) label decay rate $v(t)$, the cytons $\{(\phi_i(t), \psi_i(t))\}$, and the progressor fractions $\{F_i\}$.

2.4. Assumptions and Parameterization for a Specific Mathematical Model

Here, we list the assumptions for the specific cyton-based mathematical model we consider and describe the parameters that we use to designate this model. All of the parameters for our specific mathematical model are provided in Table 1.

First, we assume that the random variable X_a is time-independent and has a lognormal distribution with mean and variance denoted $\mathrm{E}[X_a]$ and $\mathrm{Var}[X_a]$, respectively. Although experiments indicate that the distribution of autofluorescence does, in fact, vary somewhat with time, ignoring this time-dependence greatly reduces the number of parameters required to designate the model and still allows for a reasonable fit to summary histogram data [4]. We therefore have two parameters that completely describe the distribution of autofluorescence: $\mathrm{E}[X_a]$ and $\mathrm{SD}[X_a] = (\mathrm{Var}[X_a])^{1/2}$, which are listed as parameters 1 and 2, respectively, in Table 1.

Next, we assume that the rate of decay in CFSE-induced FI is given by $v(t) = c$, where $c > 0$ is some constant. This follows the convention established in our earlier work [4], in which we assume that an exponential decay model is sufficient to describe decay of CFSE in experiments for which *the first data collection time occurs after approximately 24 hours*. Therefore, we require only one parameter to completely describe the decay of CFSE: c, which is listed as parameter 3 in the table.

As previously asserted, we assume that $m \in (0, \frac{1}{2}]$ is a constant such that the division of each mother cell results in two daughter cells that receive proportions m and $1-m$, respectively, of the cytoplasm (and CFSE) that was contained in the mother cell. This is listed as parameter 4.

We assume that each T_i^{div} has a lognormal distribution with mean and variance denoted $\mathrm{E}\left[T_i^{div}\right]$ and $\mathrm{Var}\left[T_i^{div}\right]$, respectively. We further assume that, for $i \geq 1$, all T_i^{div} are independent and identically distributed (*i.i.d.*). Such assumptions are consistent with earlier work using the cyton model [4, 14, 15], as well as experimental evidence [12]. We therefore have four parameters that

Table 1: Parameters for specific mathematical model with asymmetric division.

Number	Parameter	Description
1	$\mathrm{E}\left[X_a\right]$	mean autofluorescence
2	$\mathrm{SD}\left[X_a\right]$	std. dev. of autofluorescence
3	c	exponential decay rate for CFSE
4	m	asymmetric division constant
5	$\mathrm{E}\left[T_0^{div}\right]$	mean time to divide for cells in generation $i = 0$
6	$\mathrm{SD}\left[T_0^{div}\right]$	std. dev. of time to divide for cells in generation $i = 0$
7	$\mathrm{E}\left[T^{div}\right]$	mean time to divide for cells in later generations $(i \geq 1)$
8	$\mathrm{SD}\left[T^{div}\right]$	std. dev. of time to divide for cells in later generations $(i \geq 1)$
9	$\mathrm{E}\left[T^{die}\right]$	mean time to die for cells in later generations $(i \geq 1)$
10	$\mathrm{SD}\left[T^{die}\right]$	std. dev. of time to die for cells in later generations $(i \geq 1)$
11	F_0	progressor fraction for cells in generation $i = 0$
12	D_μ	mean of discrete normal distribution (used to compute F_i for $i \geq 1$)
13	D_σ	std. dev. of discrete normal distribution (used to compute F_i for $i \geq 1$)

completely describe $\{T_i^{div}\}$: $\mathrm{E}\left[T_0^{div}\right]$, $\mathrm{SD}\left[T_0^{div}\right] = (\mathrm{Var}\left[T_0^{div}\right])^{1/2}$, $\mathrm{E}\left[T^{div}\right] =$ $\mathrm{E}\left[T_i^{div}\right]$ for $i \geq 1$, and $\mathrm{SD}\left[T^{div}\right] = \mathrm{SD}\left[T_i^{div}\right] = (\mathrm{Var}\left[T_i^{div}\right])^{1/2}$ for $i \geq 1$. These are listed as parameters 5 through 8, respectively.

We also assume that the random variables T_i^{die} for $i \geq 1$ are *i.i.d.* with a lognormal distribution, in this case with mean and variance denoted $\mathrm{E}\left[T^{die}\right]$ and $\mathrm{Var}\left[T^{die}\right]$, respectively. Again such assumptions are consistent with earlier modeling work [4, 14, 15]. We further assume that there is no death for undivided cells (i.e., those cells in generation $i = 0$). In reality, there tends to be a large die-off of cells following stimulation with PHA [7], but we reiterate that for the data analyzed in this manuscript the first measurements were made approximately 24 hours post-stimulation. As a result, *the initial conditions for our mathematical model only reflect those cells that have not died in the first 24 hours after stimulation.* Therefore, as in our earlier work [4], we assume that the cells in our "initial population" (consisting of those undivided cells that are still alive 24 hours after stimulation) that do not go on to divide experience essentially no death for the duration of the experiment. Hence, we have two parameters that completely describe $\{T_i^{die}\}$: $\mathrm{E}\left[T^{die}\right] = \mathrm{E}\left[T_i^{die}\right]$ and $\mathrm{SD}\left[T^{die}\right] = \mathrm{SD}\left[T_i^{die}\right] = (\mathrm{Var}\left[T_i^{die}\right])^{1/2}$ for $i \geq 1$, which are listed as parameters 9 and 10, respectively.

The only remaining parameters that are required to characterize our model are the progressor fractions $\{F_i\}$. We allow F_0 to be one of our required model parameters and assume that each progressor fraction F_i with $i \geq 1$ is uniquely determined by the mean and standard deviation of a "discrete normal distribution", denoted D_μ and D_σ, respectively. This is consistent with the "division destiny" approach for determining progressor fractions that has been employed by Hawkins et al. [15] and Banks et al. [4], and we refer the interested reader to the latter reference for a complete discussion of the method by which the progressor fractions are computed. We therefore have three parameters that can be used to determine all the progressor fractions: F_0, D_μ, and D_σ, which are listed as parameters 11 through 13, respectively.

Thus, our specific model depends on exactly 13 parameters. We remark that parameters 1 through 3, while important for describing the data, are not considered to be "biologically relevant" parameters in the sense that they do not have any bearing on the proliferative behavior of a population of cells. Note that the specific cyton-based model described here closely resembles the model denoted "Model 6" (with exponential label decay) in our previous work [4], except that it allows for asymmetric division. Kapraun [16] provides precise details of the numerical methods we use for computing solutions for this model in his Appendix B.

3. Parameter Estimation

In order to estimate the parameters in our specific mathematical model, we must first describe a statistical model that relates observable data to the mathematical model. As was previously explained, CFSE-based flow cytometry data are typically summarized in the form of histograms, and furthermore, measured FI is commonly represented on a logarithmic scale (cf. Figure 1). Therefore, we begin by describing how our model can be used to obtain information on cell numbers in a form that can be compared directly with such summarized experimental data.

If we define the structured densities $\tilde{n}_i(t, \tilde{x})$ (in terms of *measured* FI) as in Section 2.2, then the structured density for the entire population of cells is

$$\tilde{n}(t, \tilde{x}) = \sum_{i=0}^{\infty} \tilde{n}_i(t, \tilde{x}) \approx \sum_{i=0}^{i_{\max}} \tilde{n}_i(t, \tilde{x}).$$

Now, because CFSE histogram data are most commonly reported using a base 10 logarithmic scale, we make the change of variables $z = \log_{10}(\tilde{x})$ to obtain

$$\hat{n}(t, z) = 10^z \log(10)\, \tilde{n}(t, 10^z)$$

as the structured density in cells per base 10 log unit FI.

In the discussion that follows, we let $\vec{q}_0$ denote a hypothetical "true" parameter vector (so that, in the case of our specific mathematical model, $\vec{q}_0 \in \mathbb{R}^{13}$) and let

$$I[\hat{n}](t_j, z_k; \vec{q}_0) = \int_{z_k}^{z_{k+1}} \hat{n}(t_j, z; \vec{q}_0)\, dz \tag{3.1}$$

denote the total number of cells with log (base 10) FI in the interval $[z_k, z_{k+1}]$ at time t_j. Also, we let B denote the (fixed) total number of beads in each sample tube and b_j denote the number of beads counted for the sample measured at time t_j.

Now, let N_k^j be a random variable representing the number of cells with log FI in the interval $[z_k, z_{k+1})$ measured at time t_j. Then it has been argued [4] that

$$N_k^j \sim \mathcal{N}\left(I[\hat{n}](t_j, z_k; \vec{q}_0), \frac{B}{b_j} I[\hat{n}](t_j, z_k; \vec{q}_0)\right); \tag{3.2}$$

i.e., each N_k^j is normally distributed with mean $I[\hat{n}](t_j, z_k; \vec{q}_0)$ and variance $\frac{B}{b_j} I[\hat{n}](t_j, z_k; \vec{q}_0)$. Note that this does not lead to either (1) a constant variance model or (2) a constant coefficient of variance model. Although these latter

two types of statistical models are commonly assumed to underly data-collection processes [8, 11, 21], modified residual plots indicate that (3.2) is a better choice in this case [5, 23].

Define the generalized least squares (GLS) parameter *estimator* [1, 11]

$$\vec{q}_{GLS} = \operatorname*{argmin}_{\vec{q} \in \mathcal{Q}} J(\vec{q}; \{N_k^j\}),$$

where

$$J(\vec{q}; \{N_k^j\}) = \sum_{j,k} \frac{(I[\hat{n}](t_j, z_k; \vec{q}) - N_k^j)^2}{w_k^j}, \tag{3.3}$$

$\mathcal{Q}$ denotes the set of allowable parameter vectors, and the *weights* (selected to match the variance of the N_k^j's) are given by

$$w_k^j = \begin{cases} \frac{B}{b_j} I[\hat{n}](t_j, z_k; \vec{q}_0) & \text{for } I[\hat{n}](t_j, z_k; \vec{q}_0) > I^*, \\ \frac{B}{b_j} I^* & \text{for } I[\hat{n}](t_j, z_k; \vec{q}_0) \le I^*. \end{cases} \tag{3.4}$$

The value of I^* is positive to prevent division by zero, and in practice it is selected such that the modified residual plots produce uniform random patterns. We once again follow the convention established in our earlier work [4] and set $I^* = 200$.

If we consider the measured data to be a set of realizations $\{n_k^j\}$ of the random variables $\{N_k^j\}$, we can obtain the GLS parameter *estimate*

$$\hat{q}_{GLS} = \operatorname*{argmin}_{\vec{q} \in \mathcal{Q}} J(\vec{q}; \{n_k^j\}). \tag{3.5}$$

Note that to compute the weights $\{w_k^j\}$ we need $\vec{q}_0$, but to estimate $\vec{q}_0$ we need the weights. In order to overcome this obstacle, we use a conventional generalized least squares iterative estimation procedure [1, 8] as described in Algorithm 3.1. In this algorithm, note that ε is a threshold tolerance that allows the user to specify a termination criterion, $\vec{q}_{typ}$ is a vector with elements that reflect the relative sizes of the parameters to be estimated, and "./" denotes element-wise division. Kapraun [16] provides specific details of the implementation we employ for this algorithm in his Appendix B.

Algorithm 3.1 Parameter Estimation Procedure

---

1. Obtain initial estimate $\hat{q}^{(0)}$ using (3.5) with $w_k^j = 1$ for all j, k.

2. Compute weights w_k^j using (3.4) with $\vec{q}_0$ replaced by $\hat{q}^{(0)}$.

3. Initialize the iteration counter ℓ with the value 1.

4. Do each of the following:

    - Compute $\hat{q}^{(\ell)}$ using (3.5) with current weights w_k^j.
    - Update the weights using (3.4) with $\vec{q}_0$ replaced by $\hat{q}^{(\ell)}$.
    - Store the value of $\left|\left|[\hat{q}^{(\ell)} - \hat{q}^{(\ell-1)}]./[\vec{q}_{typ}]\right|\right|$ in Δ.
    - Increment ℓ by 1.

5. If $\Delta > \varepsilon$, return to Step 4. Otherwise, terminate the algorithm.

---

4. Results and Discussion

We are ultimately interested in determining whether or not refining our cell proliferation model to allow for asymmetric division is worthwhile. That is, we wish to learn whether the asymmetric division model given by (2.1) – (2.3) produces better fits with values of $m < 0.5$ than with $m = 0.5$. To make this determination, we perform parameter estimations as described in Section 3. As we mentioned in Section 2.3, the algorithm for generating solutions to the 13-parameter asymmetric model is considerably more computationally complex than that previously used by the authors to generate solutions to the 12-parameter symmetric division model [2, 3, 4]. In fact, parameter estimations for the asymmetric division model require an average of 851.3 minutes per inverse problem even when only considering cells that have divided $i_{\max} = 8$ or fewer times, whereas parameter estimations for the symmetric division model require and average of only 6.41 minutes per inverse problem when using $i_{\max} = 16$ [4]. (Timings provided here are based upon runs using MATLAB Release 2012a on a Dell Optiplex 990 with eight (8) 3.4 GHz Intel Core i7-2600 processors and 8 GB of 1333 MHz memory.)

Because of the computational time required, we choose to perform parameter estimations only for a select subset of the data sets that were presented and analyzed previously by Kapraun [16]. Whereas he considered $4 \times 162 = 648$

Table 2: Data sets selected for determining the importance of the model parameter m. All data sets include data from five time points as described in Section 1. In all cases, the actual times for data collection coincided with 23.5, 46.0, 67.5, 94.5, and 117.5 hours after stimulation with PHA.

Data Set	Experiment	Data ID	Cost
1	Donor1/Vivid/CD4	10012	12415.5
2	Donor1/Vivid/CD4	10011	12642.6
3	Donor1/Vivid/CD4	11012	12710.0
4	Donor1/Vivid/CD4	21211	32661.5
5	Donor1/Vivid/CD4	11211	32698.6
6	Donor1/Vivid/CD4	11111	33047.9
7	Donor1/Vivid/CD8	21111	8446.42
8	Donor1/Vivid/CD8	20111	8451.07
9	Donor1/Vivid/CD8	10111	8774.17
10	Donor1/Vivid/CD8	01200	21334.6
11	Donor1/Vivid/CD8	02000	21607.9
12	Donor1/Vivid/CD8	02200	22621.2
12	Donor2/Vivid/CD4	00222	10210.0
14	Donor2/Vivid/CD4	10222	10214.6
15	Donor2/Vivid/CD4	00212	10526.7
16	Donor2/Vivid/CD4	00100	16099.4
17	Donor2/Vivid/CD4	11110	16290.7
18	Donor2/Vivid/CD4	01100	16435.0
19	Donor2/Vivid/CD8	02001	10256.4
20	Donor2/Vivid/CD8	00002	10419.4
21	Donor2/Vivid/CD8	02020	10471.5
22	Donor2/Vivid/CD8	20211	15565.6
23	Donor2/Vivid/CD8	21211	16008.1
24	Donor2/Vivid/CD8	21111	16099.3

data sets for Donor 1 and $4 \times 243 = 972$ data sets for Donor 2, we narrow the list of data sets by identifying the three lowest cost and three highest cost data sets for each donor and cell type. Note that the "cost" of a data set is determined by the value of the minimized GLS cost functional similar to that given in (3.3), but based on a symmetric division model. The 24 data sets satisfying these criteria are listed in Table 2.

Applying our parameter estimation scheme to the the 24 data sets listed in Table 2 leads to the results shown in Table 3. Since we are primarily interested in the asymmetric division parameter m, the table only shows estimates for that parameter. We see that in 8 of 24 cases (33.3% of the data sets), the parameter estimation scheme returns a value of *precisely* 0.5 for the parameter m. Recall that a value of $m = 0.5$ in the asymmetric division model yields model output equivalent to that of a *symmetric* division model. Moreover, in another 6 cases the value of m returned does not differ from 0.5 by more than 4 percent. That is, in $8+6 = 14$ cases (58.3% of the data sets), the parameter estimation scheme returns a value of m in the interval $[0.48, 0.5]$. This immediately suggests that asymmetry does *not* play an important role in the majority of the data sets considered.

To assess the degree to which the parameter m is influential in describing the 16 data sets for which m is not precisely 0.5, we perform statistically based model comparison tests. The approach we use is based on analysis of variance (ANOVA) hypothesis testing as outlined by Banks and Tran in their text on mathematical modeling [8]. We consider two distinct mathematical models, both of which can be evaluated using the cost functional J given in (3.3). The first is the 13-parameter model described in Section 2.4 and the second is the nested model that results when the parameter m in that model is fixed at the value 0.5 (which corresponds to perfectly symmetric cell divisions).

To formulate our statistical hypotheses, we let $\mathcal{Q} \subset \mathbb{R}^{13}$ denote the set of all admissible parameters for the 13-parameter model and $\mathcal{Q}^H = \{\vec{q} \in \mathcal{Q} : H\vec{q} = \vec{c}\} \subset \mathcal{Q}$ be the set of admissible parameters for the nested model, where $H \in \mathbb{R}^{1 \times 13}$ and $\vec{c} \in \mathbb{R}$. Using the parameter ordering suggested by Table 1 and setting

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and $\vec{c} = 0.5$ results in the nested model just described. We wish to test the null hypothesis that the "true" parameter vector $\vec{q}_0$ is in the restricted set $\mathcal{Q}^H$; i.e.,

$$H_0 : \vec{q}_0 \in \mathcal{Q}^H.$$

Therefore, we follow the usual inferential statistics practice of defining a test statistic. Let $\{N_k^j\}$ be a set of random variables as in (3.2) with corresponding realizations $\{n_k^j\}$ constituting observed data so that we can define the GLS estimators

$$\vec{q}_{GLS} = \underset{\vec{q} \in \mathcal{Q}}{\operatorname{argmin}} J(\vec{q}; \{N_k^j\}) \quad \text{and} \quad \vec{q}_{GLS}^H = \underset{\vec{q} \in \mathcal{Q}^H}{\operatorname{argmin}} J(\vec{q}; \{N_k^j\}) \qquad (4.1)$$

Table 3: Parameter estimates for m for 24 selected data sets. Data sets for which $m \neq 0.5$ are emphasized in boldface. The actual times for data collection coincided with 23.5, 46.0, 67.5, 94.5, and 117.5 hours after stimulation with PHA.

Data Set	Experiment	Data ID	m
1	Donor1/Vivid/CD4	10012	0.5000
2	Donor1/Vivid/CD4	10011	0.5000
3	Donor1/Vivid/CD4	11012	0.5000
4	Donor1/Vivid/CD4	21211	0.5000
5	**Donor1/Vivid/CD4**	**11211**	**0.4916**
6	Donor1/Vivid/CD4	11111	0.5000
7	Donor1/Vivid/CD8	21111	0.5000
8	Donor1/Vivid/CD8	20111	0.5000
9	Donor1/Vivid/CD8	10111	0.5000
10	**Donor1/Vivid/CD8**	**01200**	**0.4271**
11	**Donor1/Vivid/CD8**	**02000**	**0.4219**
12	**Donor1/Vivid/CD8**	**02200**	**0.4198**
13	**Donor2/Vivid/CD4**	**00222**	**0.4839**
14	**Donor2/Vivid/CD4**	**10222**	**0.4821**
15	**Donor2/Vivid/CD4**	**00212**	**0.4732**
16	**Donor2/Vivid/CD4**	**00100**	**0.4842**
17	**Donor2/Vivid/CD4**	**11110**	**0.4857**
18	**Donor2/Vivid/CD4**	**01100**	**0.4857**
19	**Donor2/Vivid/CD8**	**02001**	**0.4421**
20	**Donor2/Vivid/CD8**	**00002**	**0.4529**
21	**Donor2/Vivid/CD8**	**02020**	**0.4525**
22	**Donor2/Vivid/CD8**	**20211**	**0.4381**
23	**Donor2/Vivid/CD8**	**21211**	**0.4398**
24	**Donor2/Vivid/CD8**	**21111**	**0.4379**

and GLS estimates

$$\hat{q}_{GLS} = \operatorname*{argmin}_{\vec{q} \in \mathcal{Q}} J(\vec{q}; \{n_k^j\}) \quad \text{and} \quad \hat{q}_{GLS}^H = \operatorname*{argmin}_{\vec{q} \in \mathcal{Q}^H} J(\vec{q}; \{n_k^j\}). \tag{4.2}$$

We note here that the inequality $J(\hat{q}_{GLS}^H; \{n_k^j\}) \geq J(\hat{q}_{GLS}; \{n_k^j\})$ should always hold because the estimate $\hat{q}_{GLS}^H$ is obtained by optimizing over a subset of $\mathcal{Q}$, while $\hat{q}_{GLS}$ is obtained by optimizing over all of $\mathcal{Q}$. Using the GLS estimators

and estimates, we can define the test statistic

$$U(\{N_k^j\}) = \frac{n\left(J(\vec{q}_{GLS}^H; \{N_k^j\}) - J(\vec{q}_{GLS}; \{N_k^j\})\right)}{J(\vec{q}_{GLS}; \{N_k^j\})} \tag{4.3}$$

with corresponding realization

$$\hat{U}(\{n_k^j\}) = \frac{n\left(J(\hat{q}_{GLS}^H; \{n_k^j\}) - J(\hat{q}_{GLS}; \{n_k^j\})\right)}{J(\hat{q}_{GLS}; \{n_k^j\})}, \tag{4.4}$$

where

$$n = [\text{number of histogram bins}] \times [\text{number of time points}]$$

is the number of observations in a data set. We remark that, for the five-day time series data sets considered in this study, we use 1024 bins and $5-1 = 4$ time points (because the first time point is used to construct an initial condition and is not considered to be a data point), so $n = 1024 \times 4 = 5096$. As discussed by Banks and Tran [8], the test statistic U converges in distribution to a χ^2 distribution with $r = 1$ degree of freedom (where r is the number of constraints defined by the system $H\vec{q} = \vec{c}$) as $n \to \infty$. We use this statistic to test our null hypothesis.

As shown in Table 4, the costs associated with the optimal parameter vector $\hat{q}_{GLS}^H$ on the restricted set $\mathcal{Q}^H$ tend to be significantly greater than those associated with with the optimal parameter vector $\hat{q}_{GLS}$ on the set $\mathcal{Q}$. Note that the results shown in the table were obtained only for those data sets from Table 3 for which m was *not* equal to 0.5. We remark that for 3 of the selected data sets $J(\hat{q}_{GLS}^H; \{n_k^j\}) < J(\hat{q}_{GLS}; \{n_k^j\})$, which leads to a negative value for $\hat{U}(\{n_k^j\})$. While this *theoretically* should not occur, we point out that all estimates are obtained through numerical optimization (cf. Section 3). The optimization algorithm involved in parameter estimation apparently failed to find a global minimum on $\mathcal{Q}$ because it first arrived at a local minimum in these cases. Based on the very low p-values (in some cases these were below the machine precision of approximately 10^{-16} and therefore appear as zeros) resulting from the model comparison tests in the majority of cases, we should reject the null hypothesis and infer that the asymmetric division parameter m *is* important for describing the behavior of this population of proliferating T cells.

Thus, we have arrived at two seemingly contradictory results concerning the importance of incorporating asymmetric division into our model. In one third of the data sets we examined, our parameter estimation algorithm returned

Table 4: Results of the model comparison test described in Section 4.

Data Set	$J(\hat{q}_{GLS}; \{n_k^j\})$	$J(\hat{q}_{GLS}^H; \{n_k^j\})$	$\hat{U}(\{n_k^j\})$	p-value
5	33235.4	13380.5	-2446.96	1
10	20460.4	20373	-17.4833	1
11	15997	20635.2	1187.6	0
12	17384.2	21583.3	989.366	0
13	9785.79	9893.49	45.0805	1.9×10^{-11}
14	9776.4	9932.86	65.5522	5.6×10^{-16}
15	11690.5	11423.5	-93.5471	1
16	13648.9	13749	30.0378	4.2×10^{-8}
17	13241	16056.1	870.828	0
18	14017.7	14104.8	25.4563	4.5×10^{-7}
19	9733.53	10234.3	210.737	0
20	10548.6	10566.6	6.97462	8.3×10^{-3}
21	9621.94	10349.5	309.72	0
22	10341.9	13651.4	1310.76	0
23	13075.5	13858.1	245.151	0
24	11323.2	14597.8	1184.55	0

a value of $m = 0.5$, which indicates perfectly symmetric division. For the remaining data sets, however, our parameter estimation algorithm returned a value of m less than 0.5. Furthermore, if we fix m at 0.5 for these data sets, the GLS cost tends to be significantly greater. So, for two thirds of the data sets examined, allowing for asymmetric division does appear to lead to better agreement between the model and the data.

We hypothesize that the *apparent* importance of incorporating asymmetry that is implied by the model comparison experiments might be due to a type of experimental confounding. Recall that the statistical model we use to relate experimental data to our mathematical model is not a constant coefficient of variance model (cf. (3.2)). In fact, data points (histogram bins) with higher cell counts tend to contribute more to the cost functional given by (3.3), even despite the moderating effect of the weights. Also, the cell counts in the various bins tend to be larger in the later days of the experiment after the cells have had the opportunity to multiply their numbers through cell division. Therefore, our parameter estimation scheme tends to favor parameter sets that give good model fits (i.e., low residuals) for observations at later days (e.g., Days 4 and 5) over those that give good model fits for observations at earlier days (e.g.,

Days 2 and 3). As supporting evidence for this hypothesis, consider Figures 2 and 3. These figures show data from Data Set 12 of Table 3 along with the asymmetric model fits obtained (1) when m is free to take on any admissible value and (2) when m is fixed at 0.5. (Note that the latter model fit is equivalent to a symmetric division model fit.) In Figure 2, we clearly see that the symmetric division ("m fixed") model does a better job of capturing all the peaks in the data for the earlier days than does the asymmetric division ("m free") model. In examining Figure 3, it might even be argued that the symmetric division model captures the peaks in the data for the later days just as well as the asymmetric division model, but clearly the residuals produced by the asymmetric division model are smaller. Thus, despite the better overall "peak capturing" performance of the symmetric division model, the asymmetric division model is selected by the model comparison tests because it produces smaller residuals for the later days of observations. It is conceivable that the introduction of the additional parameter m allows one to obtain better fits using the asymmetric model simply because this additional degree of freedom allows one to obtain *reasonably low residuals* for the *early days of observations* while simultaneously obtaining *very low residuals* for the *later days of observations*. Perhaps applying an analysis of residuals (as described by Banks and Tran [8] and Banks et al. [1]) to the asymmetric division model would lead to a different and better statistical model upon which future parameter estimations could be based.

It should also be pointed out that, in some cases, neither the asymmetric nor the symmetric division model seem to do an adequate job of fitting CFSE FI data of the type used in this study. Figure 4, which shows data from Data Set 22 of Table 3 along with the asymmetric ("m free") and symmetric ("m fixed") model fits, illustrates this point. Notice that the fits obtained with both models are particularly poor at Days 2 and 3. While such poor fits could occur because of imperfections in the parameter estimation algorithm (e.g., optimization routines identifying local rather than global minima) or simply because some of our data sets contain unidentified experimental errors, it is also feasible that our models have yet to account for some important parameters governing T cell proliferation.

5. Conclusion

We have presented a new cell proliferation model that incorporates aspects of the cyton model of Hawkins et al. [14, 15], the separable PDE models of Schittler
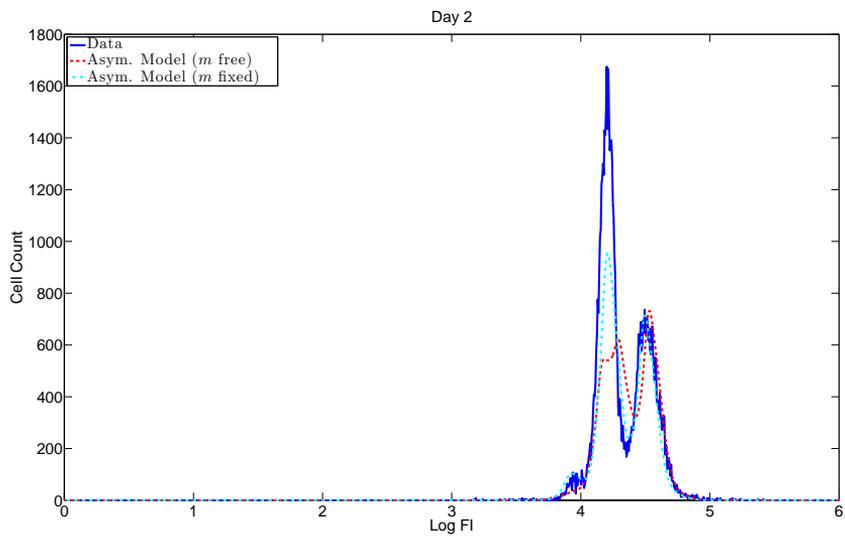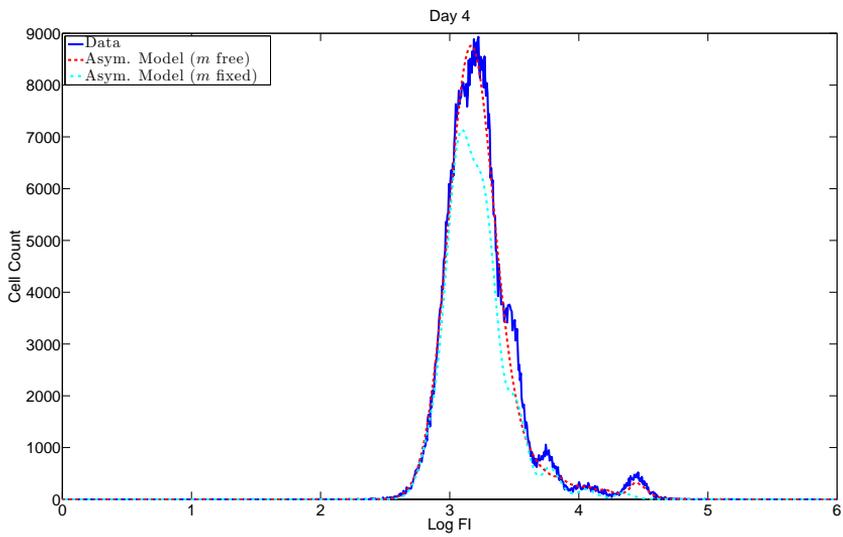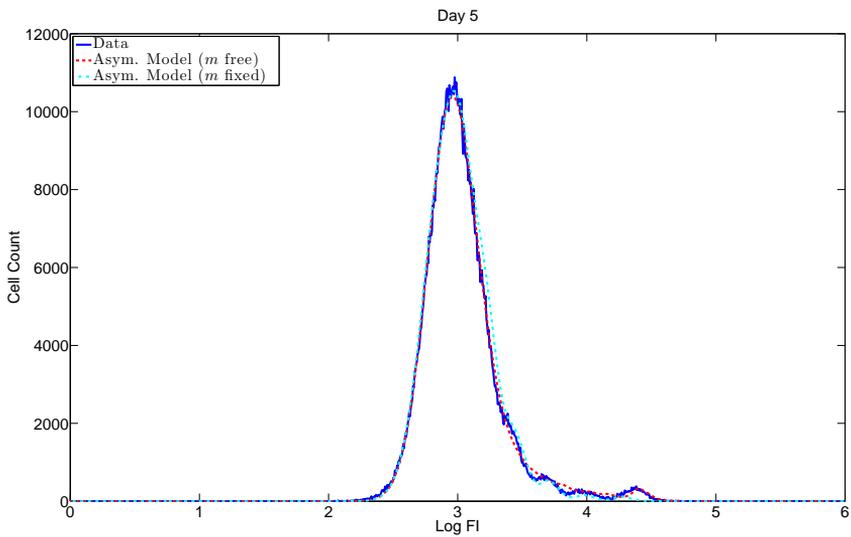
(a)
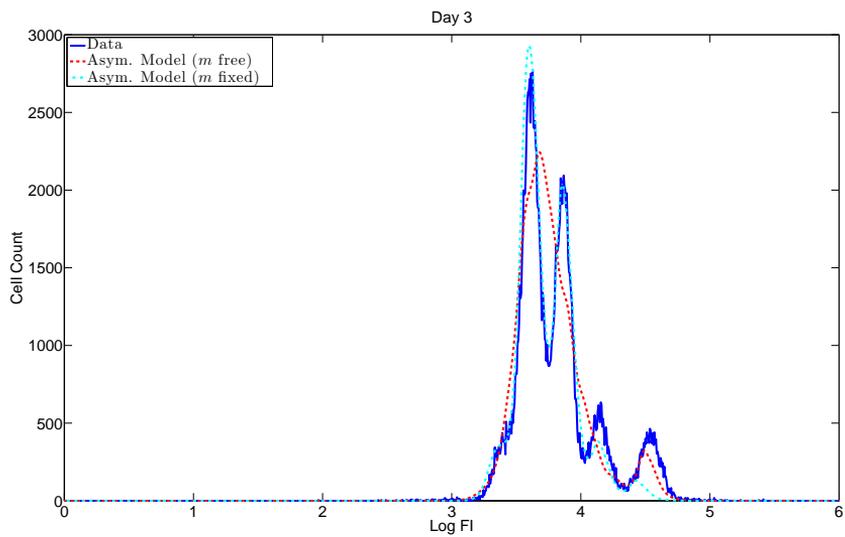
(b)

Figure 2: Plots of data from Data Set 12 of Table 3 and optimized model fits at (a) Day 2 and (b) Day 3.

(a)

(b)

Figure 3: Plots of data from Data Set 12 of Table 3 and optimized model fits at (a) Day 4 and (b) Day 5.
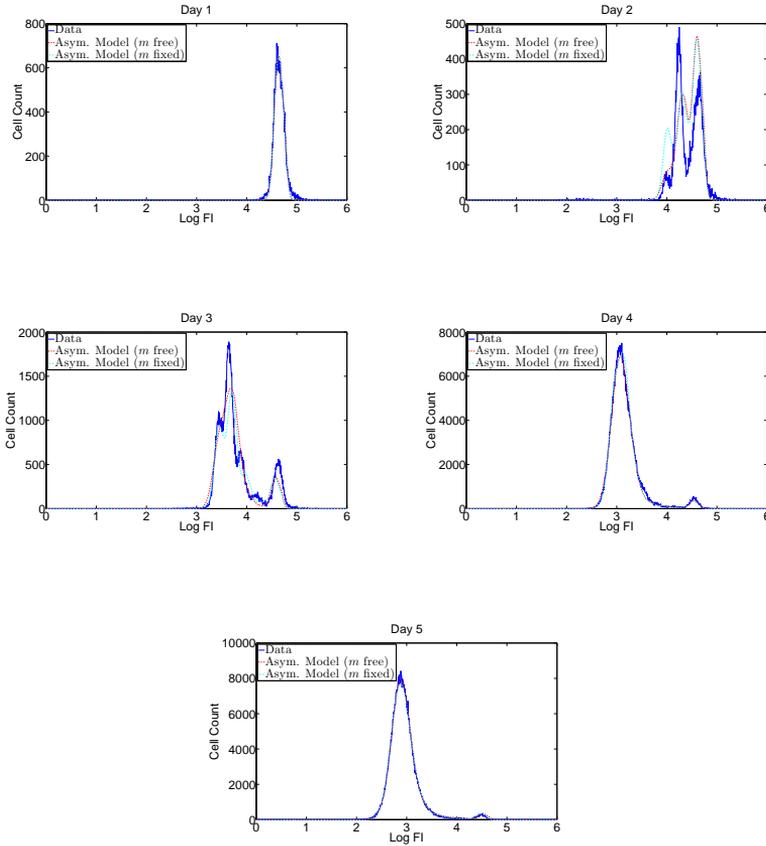
Figure 4: Plots of data from Data Set 22 of Table 3 and optimized model fits for Days 1 through 5.

et al. [20], and the asymmetric division parameter m proposed by Bocharov et al. [9] and Luzyanina et al. [17]. These latter authors [9, 17] consider asymmetry as well as a time delay in their analysis of the division process of mouse hepatitis virus specific CD8+ T cells. They use a Wiener filter with a smooth decreasing low-pass filter function to pretreat (reduce the noise) in the CFSE histogram data. In our models, the use of cytons accounts for variability in division times and time lags in the onset of cell divisions. Moreover, we use unfiltered data from PHA-stimulated human CD4+ and CD8+ T cell experiments.

The model comparisons performed here do not provide conclusive evidence that including a provision for asymmetric division in our models consistently leads to improved fits for the CFSE-based flow cytometry data sets we considered. In some cases, including the parameter m allowed for model fits with

significantly lower costs (as evaluated using (3.3)), but in the majority of cases the estimated value of m was not substantially different from 0.5 (which corresponds to perfectly symmetric cell divisions). Upon visual inspection of some of the model fits obtained in Section 4, it was observed that the selection of the asymmetric model over the symmetric model might be due to the form of the statistical model underlying the cost functional (3.3). Thus, it is unclear as to whether the incorporation of asymmetric division into previously studied cyton-based division- and label-structure cell proliferation models is worth the associated increase in computational cost.

Acknowledgments

References

[1] H.T. Banks, S. Hu, W.C. Thompson, *Modeling and Inverse Problems in the Presence of Uncertainty*, CRC Press, Boca Raton, 2014.

[2] H.T. Banks, D.F. Kapraun, K.G. Link, W.C. Thompson, C. Peligero, J. Argilaguet, and A. Meyerhans, Experimental and biological variability in CFSE-based flow cytometry data, CRSC-TR13-10, North Carolina State University, September 2013.

[3] H.T. Banks, D.F. Kapraun, K.G. Link, W.C. Thompson, C. Peligero, J. Argilaguet, and A. Meyerhans, Analysis of variability in estimates of cell proliferation parameters for cyton-based models using CFSE-based flow cytometry data, CRSC-TR13-14, North Carolina State University, November 2013; *J. Inverse Ill-Posed Probl.*, to appear.

[4] H.T. Banks, D.F. Kapraun, W.C. Thompson, C. Peligero, J. Argilaguet, and A. Meyerhans, A novel statistical analysis and interpretation of flow cytometry data, CRSC-TR12-23, North Carolina State University, March 2013; *J. Biol. Dyn.* **7** (2013), 96–132.

[5] H.T. Banks, Z.R. Kenz, and W.C. Thompson, An Extension of RSS-based Model Comparison Tests for Weighted Least Squares, CRSC-TR12-18, North Carolina State University, August 2012; *Intl. J. Pure and Appl. Math.* **79** (2012).

[6] H.T. Banks, K.L. Sutton, W.C. Thompson, G. Bocharov, M. Doumic, T. Schenkel, J. Argilaguet, S. Giest, C. Peligero, and A. 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. Immunol. Methods* **373** (2011), 143–160.

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

[8] H.T. Banks and H.T. Tran, *Mathematical and Experimental Modeling of Physical and Biological Processes*, CRC Press, Boca Raton, 2009.

[9] G. Bocharov, T. Luzyanina, J. Cupovic, and B. Ludewig, Asymmetry of cell division in CFSE-based lymphocyte proliferation analysis, *Front. Immunol.* **4** (2013), published online.

[10] G. Casella and R.L. Berger, *Statistical Inference* (2nd Edition), Duxbury, 2002.

[11] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, London, 2000.

[12] A.V. Gett and P.D. Hodgkin, A cellular calculus for signal integration by T cells, *Nat. Immunol.* **1** (2000), 239–244.

[13] J. Hasenauer, D. Schittler, and F. Allgöwer, Analysis and simulation of division- and label-structured population models: a new tool to analyze proliferation assays, *Bull. Math. Biol.* **74** (2012), 2692–2732.

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

[15] E.D. Hawkins, M.L. Turner, M.R. Dowling, C. van Gend, and 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.

[16] D.F. Kapraun, *Cell Proliferation Models, CFSE-Based Flow Cytometry Data, and Quantification of Uncertainty*, Ph.D. Dissertation, Department of Mathematics, North Carolina State University, Raleigh, NC (24 July 2014).

[17] T. Luzyanina, J. Cupovic, B. Ludewig, and G. Bocharov, Mathematical models for CFSE labelled lymphocyte dynamics: asymmetry and time-lag in division, *J. Math. Biol.* **69** (2014), 1547–83.

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

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

[20] D. Schittler, J. Hasenauer, and 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 2011, Zurich, Switzerland, p. 165–168.

[21] G.A. Seber and C.J. Wild, *Nonlinear Regression*, Wiley, Hoboken, NJ, 2003.

[22] R. Sennerstam, Partition of protein (mass) to sister cell pairs at mitosis: a re-evaluation, *J. Cell. Sci.* **90** (1988), 301–306.

[23] W.C. Thompson, *Partial Differential Equation Modeling of Flow Cytometry Data from CFSE-based Proliferation Assays*, Ph.D. Dissertation, Department of Mathematics, North Carolina State University, Raleigh, NC (8 December 2011).

[24] S.A. Weston and C.R. Parish, New fluorescent dyes for lymphocyte migration studies: Analysis by flow cytometry and fluorescence microscopy, *J. Immunological Methods* **133** (1990), 87–97.