Ken R. Duffy · Vijay G. Subramanian

On the impact of correlation between collaterally consanguineous cells on lymphocyte population dynamics

Abstract During an adaptive immune response, lymphocytes proliferate for five to twenty five cell divisions, then stop and die over a period of weeks. Based on extensive flow cytometry data, Hawkins et al. (PNAS, 2007, 104, 5032–5037) introduced a cell-level stochastic model of lymphocyte population dynamics, called the Cyton Model, that accurately captures mean lymphocyte population size as a function of time. In Subramanian et al. (J. Math. Biol., 2008, 56:6, 861–892), we performed a branching process analysis of the Cyton Model and deduced from parameterizations for *in vitro* and *in vivo* data that the immune response is predictable despite each cell's fate being highly variable.

One drawback of flow cytometry data is that individual cells cannot be tracked, so that it is not possible to investigate dependencies in the fate of cells within family trees. In the absence of this information, while the Cyton Model abandons one of the usual assumptions of branching processes (the independence of lifetime and progeny number), it adopts another of the standard branching processes hypotheses: that the fates of progeny are stochastically independent. However, new experimental observations of lymphocytes show that the fates of cells in the same family tree are not stochastically independent. Hawkins et al. (2008, submitted for publication) report on ciné lapse photography experiments where every founding cell's family tree is recorded for a system of proliferating lymphocytes responding to a mitogenic stimulus. Data from these experiments demonstrate that the death-or-division fates of collaterally consanguineous cells (those in the same generation within a founding cell's family tree) are strongly correlated, while there is little correlation between cells of distinct generations and between cells in distinct family trees.

As this finding contrasts with one of the assumptions of the Cyton Model, in this paper we introduce three variants of the Cyton Model with increasing levels of collaterally consanguineous correlation

Ken Duffy and Vijay Subramanian, Hamilton Institute, National University of Ireland, Maynooth, Ireland
E-mail: {ken.duffy, vijay.subramanian}@nuim.ie

structure to incorporate these new found dependencies. We investigate their impact on the predicted expected variability of cell population size. Mathematically we conclude that while the introduction of correlation structure leaves the mean population size unchanged from the Cyton Model, the variance of the population size distribution is typically larger. Biologically, through comparison of model predictions for Cyton Model parameterizations determined by *in vitro* and *in vivo* experiments, we deduce that if collaterally consanguineous correlation extends beyond cousins, then the immune response is less predictable than would be concluded from the original Cyton Model. That is, some of the variability seen in data that we previously attributed to experimental error could be due to intrinsic variability in the cell population size dynamics.

1 Introduction

Advances in flow cytometry and the introduction of techniques for following cell division by labeling with the fluorescent dye carboxyfluorescein succinimidyl ester (CFSE) have enabled the collection of experimental data on the kinetics of lymphocyte division progression and cell survival [1][2]. For example, the series of papers [3][4][5][6][7] report on extensive *in vitro* experiments on the behavior of purified naïve lymphocytes exposed to a mitogenic stimulus. Starting with a collection of founding lymphocytes, these techniques enable the determination of the number of cells alive as a function of time and identify the distribution of cell generation number. Data gathered in these experiments suggests that each cell's lifetime and progeny number are not stochastically independent [6]. Instead it suggests a mathematical mechanism to describe the dependence through the competition between two distinct processes within each cell that can be mathematically described by two random variables: a time to die, T_D, and a time to divide, T_B. Whichever process is completed first (i.e., whichever random variable is smaller) determines the fate of that cell. The experimental data indicates that T_B and T_D are a function of a cell's generation number and that, in particular, the distributions of the random variables for founding cells are parameterized significantly differently to those in later generations [6].

Lymphocytes responding to a mitogenic stimulus do not proliferate indefinitely, but undergo a finite number of divisions before ceasing to sub-divide, with this number varying within the population. It is demonstrated in [7] that this limit of proliferation in the *in vitro* systems is not due culture conditions (i.e., exhausting media or stimuli), but is an important and regulable feature of *in vitro* responses. Thus to describe lymphocytes responding to a mitogenic stimulus, there is a need for a mechanism that captures the fact that proliferation is transient and does not continue indefinitely. For lymphocytes, Hawkins *et al.* [6] coin the phrase *Division Destiny* for this mechanism, where the progeny of founder cells can only sub-divide at most a certain number of times. As division destiny varies within the population of founding cells, we will mathematically model it as a random variable.

Flow cytometry data motivated the introduction of the Cyton Model, a set of stochastic hypotheses for lymphocyte population dynamics, by Hawkins *et al.* [6], which encapsulates these properties. In [6], the Cyton Model is used to determine the evolution of the mean population size per generation and is fit to lymphocyte flow cytometry data. The Cyton Model has been further analyzed by Subramanian *et al.* [8] who determine the evolution of each generation's population size Probability Generating Function (PGF) by considering it as a non-standard branching process. In [8] it is reported that despite the fate of each cell being highly variable, for Cyton Model parameterizations from *in vitro* and *in vivo* experiments the overall population size variance is small. It is concluded that despite the potential for great variability in the experience of each individual cell, the immune response is predictable.

One drawback of flow cytometry methods is that individual cells cannot be tracked. Thus from flow cytometry data it is not possible to infer relationships between progeny of a single founder. In the

absence of this information, while the Cyton Model abandons the usual independence of lifetime and progeny number assumption, it adopts another of the standard branching processes hypotheses: that progeny are stochastically independent copies of the each other. However, new experimental observations show that siblings are <u>not</u> stochastically independent entities. This motivates the present article.

The reaction of B lymphocytes stimulated with the ligand for Toll Like Receptor 9 (TLR-9) has recently been proposed as a representative of cell population dynamics by Hawkins *et al.* [9]. This cell system has desirable experimental properties that enable the collection of detailed cell-by-cell data through optical observation. In particular, unlike most proliferating cell systems, the proliferating population does not form three dimensional structures, but remains in a two dimensional plane. Thus each individual cell and its progeny can be optically followed through time lapse photography, as they do not leave the field of vision. Although labor intensive, these optical experiments allow the collection of more detailed information than flow cytometry. They enable the construction of a family tree for each founder cell and the investigation of correlation structure between division, death and division destiny of relatives and between distinct family trees. From an identifiability perspective, Hawkins *et al.* [9] discovered an optical marker to distinguish cells that had experienced division destiny from those that had died through apoptosis before division occurred. Cells that have undergone division destiny fail to grow before dying, whereas those that divide or die by apoptosis continually expand until reaching their fate. Based on this marker, the data indicates that division destiny is typically experienced by all members of a given family tree in a single generation, with a lot of variation across distinct family trees. We will, therefore, mathematically represent division destiny as random variable, K, that independently assigns to each founding cell a maximum number of cell divisions that its family tree can experience.

The results reported in [9] show the presence of strong correlation between the time to divide of collaterally consanguineous cells (i.e. cells that are in the same generation of a single founder's family tree). Clearly this has serious implications for mathematical predictions as one of the Cyton Model assumptions, the independence of siblings, appears to be inaccurate.

In this article we consider three variants of the Cyton Model that incorporate increasing levels of collaterally consanguineous correlation structure within the paradigm of non-standard branching processes. The first variant introduces the least amount of additional dependence, where the only change is that siblings experience the same time to divide. The second variant introduces further correlation structure where all siblings, cousins and second cousins experience the same fate, but the fates of more distantly related cells are stochastically independent. The final variant, Variant 3, introduces the most extreme form of correlation for collaterally consanguineous cells: all cells in the same generation of a single family tree experience the same fate. When considered in conjunction with the Cyton Model, these three variants allow us to explore the full range of collaterally consanguineous correlation structures: from the Cyton Model with no correlations, through two variants with local dependencies, to a variant with highly non-local correlation structure.

Analyzing these variants of the Cyton Model enables us to draw conclusions regarding the impact of correlation structure on the predictability of immune response. We determine the time-evolution of the probability generating function of population size per generation and compare predictions with experimental data. So long as the probability generating function is finite in an open neighborhood of 1 (e.g. [10] pp 278), the population distribution can be recovered using numerical inversion techniques such as those described in [11][12][13][14][15], but in this article we focus on deductions regarding the first two moments of the distribution, the mean and the variance.

This paper is organized as follows. In Section 2 we recall the mathematical framework of the Cyton Model and introduce three new models based on alternate hypotheses that incorporate increasing

depth of correlation structure between collaterally consanguineous cells. These models are studied in detail in Sections 3, 4 and 5. For each variant an iterative scheme is introduced to determine the probability generation function of population size per generation, taking into account the possibility of different distributions per generation and division destiny, to enable deductions for parameterizations from data. Mathematically we find:

- The evolution of mean population size is the same in the original Cyton Model and its three variants with additional correlation structure.
- All higher moments of the population size distribution differ in each model. Thus there is, potentially, significantly higher intrinsic variability in population size than would be predicted based on the standard independence of siblings hypothesis.
- Dependencies typically lead to higher variance at all times. If the dependency is local with, for example, siblings (or siblings and cousins etcetera) sharing the same fate, but more distant relations are independent, then variance is larger but qualitatively similar with that from the equivalent system with independence. However, if the dependency is non-local where all collaterally consanguineous cells (i.e. all those in a given family tree that are of the same generation) share the same fate, we show by an analytic example that the variance can behave qualitatively differently, growing more quickly and remaining significantly higher at large times.

In Section 6 we compare deductions from the Cyton Model variants for parameterizations that have been proposed for *in vitro* [7] and *in vivo* [16] data. Biologically we find:

- If siblings having identical times to divide, but with all more distant relations being stochastically independent, this results in a negligible increase in variance over the Cyton Model and that, therefore, this would continue to suggest a highly predictable immune response.
- Assume siblings, cousins and second cousins have identical times to divide and to die, but with all more distant relations being stochastically independent. If proliferation occurs for less than (approximately) ten generation this results in a reasonable increase in variance that is still qualitatively similar to that of the Cyton Model. If proliferation occurs for a larger number of generations then these local dependencies are washed out and the model predictions are nearly identical to those of the Cyton Model. This would suggest that the immune response is still highly predictable, but that, in comparison with the Cyton Model, more of the variability observed in data may be attributable to intrinsic fluctuations of the cell dynamics if proliferation only occurs for a small number of generations.
- If all the progeny of a founding cell within a single generation experience the same division and death times, but cells in distinct generations are independent, the increase in variance is significant and the model makes qualitatively different predictions to the Cyton Model. Variance does not increase more dramatically than the other models during proliferation, but during cessation it decreases significantly more slowly. This suggests that some of the variability in data than we had previously attributed to experimental error based on analysis of the Cyton Model [8] could be due to intrinsic fluctuations, especially at large times.

In Section 7 we discuss related work and the deductions from the results in the present article.

2 The Cyton Model and three variants with correlated collaterally consanguineous cells

The Cyton Model was introduced by Hawkins *et al.* [6]. It provides a cell-level stochastic model for the regulation of lymphocyte proliferation and survival. It addresses the requirement for variability in

each cell's fate by postulating independent control of time to divide, T_B, and time to die, T_D, within each cell. Whichever processes concludes first determines the fate of that cell. The combination of the two time-based mechanisms in each cell is called the Cyton. In Subramanian et al. [8] a non-standard branching process is used to analyse the Cyton Model. There it is noted that while the independence of the random variables T_B and T_D is motivated by biological reasoning, for the mathematics it is not formally necessary. This is significant for model identifiability since we do not need to know the full distributions of T_B and T_D, but only certain conditional distributions that are directly observable if one conducts optical experiments.

The reason that a non-standard branching process is needed to analyse the Cyton Model is as follows. In terms familiar to the literature in continuous time Branching processes [17][18][19][20], letting L denote a cell's lifetime random variable and ζ its number of progeny random variable, we can write each in terms of T_B and T_D:

$$L = \min(T_B, T_D) \qquad \text{and} \qquad \zeta = \begin{cases} 2 \text{ if } T_B < T_D \\ 0 \text{ if } T_B \geq T_D, \end{cases}$$

where T_B and T_D are assumed to be almost surely positive and if both processes complete simultaneously, $T_D = T_B$, we assume that the cell dies. Except in exceptional circumstances (e.g. independent and exponentially distributed T_B and T_D), L and ζ are not independent random variables even if T_B and T_D are. Thus the Cyton Model leads to a non-standard continuous time branching process rather than a Bellman-Harris process [18].

Explicitly, the Cyton Model hypotheses are:

- Each cell's Cyton is stochastically independent.
- Individual cells will, upon division, adopt new values for the time to die and time to divide that are drawn from appropriate distributions.
- The operation of the regulable Cyton controlling division and survival, seen leading up to the first division, is repeated through subsequent divisions.
- As in [8], there is a division destiny for each founding cell in the initial population that is drawn from a stochastic distribution.

Inspired by the evidence presented in Hawkins *et al.* [9] we consider three variants of the Cyton Model that incorporate increasing levels of dependencies between collaterally consanguineous cells[1].

- *Variant 1:* Founding cells have independent Cytons. Siblings have the same time to divide, but different times to die. All relations that are cousins or more distant (second cousins, etcetera) are independent.
- *Variant 2:* Founding cells have independent Cytons. Siblings, cousins and second cousins have the same time to divide and die. All relations that are more distant are independent.
- *Variant 3:* Founding cells have independent Cytons. For each individual generation in a founding cell's family there is only a single time to die and time to divide pair that is experienced by every cell in that generation. Thus in each generation either every cell, be they siblings, cousins, second cousins, etcetera, in a given family tree divides or dies and all cells do so at the same time.

In the three Variants, all cells that are progeny of a single founding cell experience the same division destiny. Also all three retain the hypothesis that times to die and and divide are independent across distinct generations.

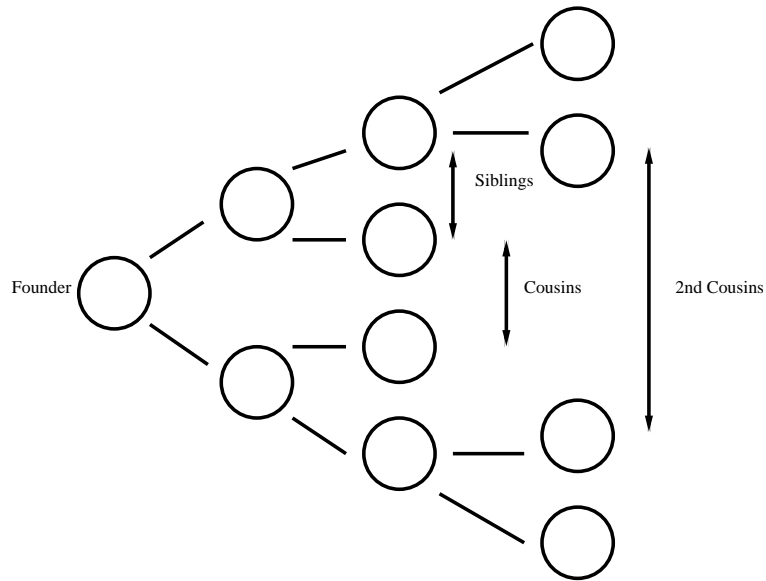[1] See Figure 1 for our family tree nomenclature.

Fig. 1 Founding cell's family tree indicating siblings, cousins and second cousins

Variant 1 enables us to explore the impact that correlation in the smallest part of a family tree, siblings, has on model predictions of variance. It is possible to construct analogous models to Variant 1, but where this dependence occurs up to the level of cousins or second cousins, etcetera, by using similar ideas to those in the present article.

In Variant 2, cells that are siblings, cousins and second cousins experience identical fates at identical times, but the Cytons of more distant relatives are independent. These cells share the same T_B and T_D random variables, where as the time to die and time to division random variables of more distant relations are stochastically independent.

Variant 3 represents the highest level of correlation structure that we consider. For each generation of cells from a single founder, all cells experience the same fate, death or division, and, moreover, this fate occurs at the same time. This contrasts with models as in Variant 2 where the correlation extends only a finite breadth in each generation. Variant 3 gives rise to significantly larger variance at large times, as members from a small number of family trees are likely to constitute the majority of the population and within these families, all cells die, divide and experience division destiny at the same time.

We begin by introducing a simple analytic example in an idealized setting. In Section 2.1, we consider the Cyton Model for a homogeneous population whose distributions do not depend on generation number and in the absence of division destiny. This example enables us to show that variance in Variant 3 behaves qualitatively differently to the other variants at large times. As we will compare the Cyton Model with the variants based on parameterizations from data, in Section 2.2 we briefly recall a mathematical framework [8] for determining the Probability Generating Function (PGF) of the Cyton Model. This treatment includes the full model where the distributions of T_B and T_D may change at each generation and where the population experience division destiny. For parameterizations from data, quantitative comments are made in Section 6.

2.1 Cyton Model without division destiny, homogeneous analytic example

Let $Z(t)$ denote the total number of cells alive at time t. We consider the Probability Generating Function (PGF) $F^{\text{cyt}}(s,t) := E(s^{Z(t)} | Z(0) = 1)$ of a single founding cell. We do this because if $Z(0) = d$, then $E(s^{Z(t)} | Z(0) = d) = E(s^{Z(t)} | Z(0) = 1)^d$ as founding cells are assumed to be stochastically independent.

For illustrative (rather than data matching) purposes we consider the fundamental (non-linear Volterra) integral equation for $F^{\text{cyt}}(s,t)$ for the Cyton Model [8] in the presence of a homogeneous population where T_B and T_D are the same for every generation and in the absence of division destiny:

$$F^{\text{cyt}}(s,t) = sP(T_B > t, T_D > t) + P(T_D \leq t, T_B \geq T_D) + \int_0^t F^{\text{cyt}}(s, t-r)^2 dP(T_D > r, T_B = r). \quad (1)$$

These three terms come from the usual considerations: neither death nor division has occurred and the founding cell is still alive; time to death has occurred and did so before time to division so that the founding cell died; time to division expired before time to die, the founder gave rise to two siblings. As the two siblings are independent and identically distributed with the founder, we get the $F^{\text{cyt}}(s, t-r)^2$ term: with $Z^1(t)$ denoting the progeny of one sister and $Z^2(t)$ denoting the progeny of the other, then using the independence property followed by the homogeneity assumption that is in force in this section:

$$E(s^{Z^1(t)+Z^2(t)} | Z^1(0) = Z^2(0) = 1) = E(s^{Z^1(t)} | Z^1(0) = 1)E(s^{Z^2(t)} | Z^2(0) = 1)$$
$$= E(s^{Z(t)} | Z(0) = 0)^2 = F^{\text{cyt}}(s,t)^2. \quad (2)$$

Recall that the mean population size $m(t)$ can be determined as

$$m(t) := \frac{\partial}{\partial s} F^{\text{cyt}}(s,t)|_{s=1} = E(Z(t)),$$

Defining

$$v^{\text{cyt}}(t) := \frac{\partial^2 F^{\text{cyt}}(s,t)}{\partial s^2}\Big|_{s=1} = E(Z(t)^2) - E(Z(t)),$$

the variance is

$$\text{var}^{\text{cyt}}(t) := v^{\text{cyt}}(t) + m(t)(1 - m(t)).$$

Taking partial derivatives of equation (1) with respect to s we get that

$$m(t) = P(T_D > t, T_B > t) + \int_0^t 2m(t-r) \, dP(T_B = r, T_D > r) \quad (3)$$

$$v^{\text{cyt}}(t) = \int_0^t 2 \left(m(t-r)^2 + v^{\text{cyt}}(t-r) \right) dP(T_B = r, T_D > r) \quad (4)$$

for the homogenous Cyton Model without division destiny.

Consider the setting where the time to division is exponentially distributed, $P(T_B^H > t) = e^{-\lambda_B t}$, and the time to death is independent of it and exponentially distributed $P(T_D^H > t) = e^{-\lambda_D t}$. In this case each cell's life time random variable L, the time to whichever of division or death occurs first, has distribution $P(L \leq t) = 1 - P(T_B^H > t, T_D^H > t) = 1 - \exp(-(\lambda_B + \lambda_D)t)$. The number of children ζ at the end of its lifetime satisfies $P(\zeta = 0) = \lambda_D/(\lambda_B + \lambda_D)$ and $P(\zeta = 2) = \lambda_B/(\lambda_B + \lambda_D)$.

As $P(\zeta = 0, L \leq t) = P(T_D^H < T_B^H, T_D^H \leq t) = \int_0^t P(T_B^H > r) dP(T_D^H = r) = \lambda_D/(\lambda_B + \lambda_D)(1 - \exp(-(\lambda_B + \lambda_D)t)) = P(\zeta = 0)P(L \leq t)$, the random variables ζ and L are independent. Thus this forms a classic Bellman-Harris process and the homogeneous part of the population falls within the standard branching process methodology where it is well known that

$$m(t) = e^{(\lambda_B - \lambda_D)t}$$

$$\text{and} \quad \text{var}^{\text{cyt}}(t) = \begin{cases} \left(e^{2(\lambda_B - \lambda_D)t} - e^{(\lambda_B - \lambda_D)t}\right) \left(\frac{\lambda_B + \lambda_D}{\lambda_B - \lambda_D}\right) & \text{if } \lambda_B \neq \lambda_D, \\ 2\lambda_B t & \text{if } \lambda_B = \lambda_D. \end{cases} \quad (5)$$

Note that exponentially distributed lifetimes are not biologically plausible for most cell systems. It has been known since the work of Prescott [21], Nachtwey and Cameron [22], and Smith and Martin [23] that there is a lower bound on the time between cell divisions that is thought to correspond to the minimum time cells take to progress through the S and G_2/M phases of the cell cycle. The purpose of this simple explicit example with exponential distributions is to enable us to demonstrate in closed form that the non-local correlation structure of collaterally consanguineous cells in model Variant 3 gives rise to qualitatively different behavior in the variance. When comparing model predictions with data in Section 6 we employ log-normal distributions and include a minimum time between divisions.

2.2 Cyton Model, generational PGF evolution

As with most branching processes, the fundamental equation (1) for the homogeneous Cyton Model without division destiny (and the equivalent ones for its variants) only admit analytic solutions for a limited class of T_B and T_D distributions. Experimental evidence from flow cytometry proposes skewed distributions, such as Weibull or log-normal, as providing an excellent fit to death and division time data [6]. These distributions are outside this class for which equation (1) has an analytic solution. Moreover, as the distributions change between the founding cells and their progeny, to match with data it is necessary to provide an iterative scheme to determine the probability generating function for the population size. This iterative scheme is unlikely to lead to easy analytic deductions, but enables simple numerical implementation for any class of distributions. We will present similar schemes for the three Variants.

We say that founding cells are in generation 0, the cells that result from their division are in generation 1 and so forth. For each generation $k \in \mathbb{Z}_+$ let T_D^k denote the time to die and T_B^k denote the time to division random variables. To enable deductions regarding the joint distributions of distinct generations it will prove more convenient to work with Moment Generating Functions (MGFs), $M^{\text{cyt}}(\theta, t) := F^{\text{cyt}}(e^\theta, t)$, $\theta \in \mathbb{R}$, rather than the PGFs as in [8]. Let $Z_j(t)$ denote the number of cells of generation j that are alive at time t. We consider the MGF of random vector of cells in generations k through l at time t, $(Z_k(t), \ldots, Z_l(t))$, after starting with one cell of generation k at time 0. Defining $M^{\text{cyt}}(\boldsymbol{\theta}_k^l, t) := 1$ if $l < k$, for $k \leq l$ this MGF is defined by

$$M^{\text{cyt}}(\boldsymbol{\theta}_k^l, t) := E\left(e^{\sum_{j=k}^l \theta_j Z_j(t)} | Z_k(0) = 1, \; Z_i(0) = 0 \text{ if } i \neq k\right),$$

where $\boldsymbol{\theta}_k^l$ is the vector $(\theta_k, \theta_{k+1}, \ldots, \theta_l) \in \mathbb{R}^{l-k+1}$.

Considering all the possible fates of the initial cell from generation k that could occur until time t, with the logic as in equation (1), we get the following MGF relationship across generations:

$$M^{\text{cyt}}(\boldsymbol{\theta}_k^l, t) = e^{\theta_k} P(T_D^k > t, T_B^k > k) + P(T_D^k \leq t, T_B^k \geq T_D^k)$$
$$+ \int_0^t M^{\text{cyt}}(\boldsymbol{\theta}_{k+1}^l, t - r)^2 dP(T_B^k = r, \ T_D^k > r),$$

where we have used the fact that siblings, and their descendents, are independent.

To include division destiny, we recall that the progeny of a founder cell can undergo a maximum number of subdivisions. This number K is a random variable that is selected independently for each founding cell from a given distribution. If division destiny selects at generation k^*, then cell division cannot happen once generation k^* is reached, i.e., $P(T_B^{k^*} = +\infty) = 1$. Furthermore, we assume a different time to die random variable $\tilde{T}_D^{k^*}$ for the time to die distribution once division destiny occurs.

To include division destiny, we must derive

$$\tilde{M}^{\text{cyt}}(\boldsymbol{\theta}_k^{k^*}, t) := E\left(e^{\sum_{j=k}^{k^*} \theta_j Z_j(t)} | Z_k(0) = 1, \ Z_i(0) = 0 \text{ if } i \neq k\right)$$

for all $k \in \{0, 1, \ldots, k^*\}$ and for all k^* in the support of the distribution of K. These are determined in turn by each iteration given below:

1. Initialize by setting $k = k^*$ and $\tilde{M}^{\text{cyt}}(\theta_{k^*}, t) = 1 + (e^{\theta_{k^*}} - 1)P(\tilde{T}_D^{k^*} > t)$.
2. Set $k = k - 1$ and calculate

$$\tilde{M}^{\text{cyt}}(\boldsymbol{\theta}_k^{k^*}, t) = e^{\theta_k} P(T_D^k > t, T_B^k > k) + P(T_B^k \leq t, T_B^k \geq T_D^k)$$
$$+ \int_0^t \tilde{M}^{\text{cyt}}(\boldsymbol{\theta}_{k+1}^{k^*}, t - r)^2 dP(T_B^k = r, \ T_D^k > r).$$

3. If $k > 0$, then repeat the previous step and otherwise terminate.

The division destiny random variable K typically has finite support, with upper bound $K_{\max}$, so that the iterative scheme will terminate after a finite number of iterations[2]. We are now in a position to determine the MGF of the population starting with one founding cell at time 0 by performing the following:

$$\bar{M}^{\text{cyt}}(\boldsymbol{\theta}_0^{K_{\max}}, t) := \sum_{k^*=0}^{K_{\max}} P(K = k^*) \tilde{M}^{\text{cyt}}(\boldsymbol{\theta}_0^{k^*}, t). \tag{6}$$

With the distributions of $K, \{T_B^k, T_D^k\}$ given we can follow this scheme to determine $\bar{M}^{\text{cyt}}(\boldsymbol{\theta}_0^{K_{\max}}, t)$ in equation (6). As founding cells are independent, if there are d initial cells, the MGF of the total population is the dth power of $\bar{M}^{\text{cyt}}(\boldsymbol{\theta}_0^{K_{\max}}, t)$ in equation (6).

Once $\bar{M}^{\text{cyt}}(\boldsymbol{\theta}_0^{K_{\max}}, t)$ is known, the moments of the distribution of population size can be determined by numerically taking partial derivatives of $\bar{M}^{\text{cyt}}((\theta, \ldots, \theta), t)$ with respect to θ and evaluating at $(0, 0, \ldots, 0)$. As this may be numerically problematic, if one only wishes to determine a particular moment of the distribution, one can explicitly take derivatives of the iterative scheme to get a direct iterative scheme for the moment of interest. In Section 6 we will use the latter methodology to determine the predicted variance of the Cyton Model for distributions that have been proposed to match with data and compare it with the predictions based on its variants.

---

[2] Even if it is not the case that K has bounded support, there exists a minimum time between cell divisions [23]. Thus there exists some $b > 0$ such that $T_B^k > b$ almost surely for every generation k and for a given time t at most members of the first $[t/b]$ (the greatest integer smaller than t/b) generations can be present.

3 Cyton Model Variant 1, siblings with correlated division

Variant 1 introduces the least amount of additional correlation structure beyond the original Cyton Model, where siblings have the same time to divide, but different times to die. All cells that are more distantly related (e.g. cousins) are independent. It is possible to extend the analysis to models where this correlation is up to the level of cousins, second cousins, etcetera, but here we focus on siblings.

The Cyton Model Variant 1 hypotheses are:

- Each founding cell is stochastically independent.
- Cells in distinct generations are independent.
- Siblings share the same time to divide, but have differing times to die. Cousins (cells born of two siblings), second cousins, etcetera, have stochastically independent times to divide and die, which are all drawn from the appropriate distributions.
- There is a division destiny for each founding cell in the starting population that is drawn from a stochastic distribution.

3.1 Cyton Model Variant 1 without division destiny, homogenous analytic example

To identify $F^{\text{var1}}(s,t)$ for a single founding cell in a population with homogeneous T_B and T_D distributions and in the absence of division destiny, it is necessary to determine the evolution of two related PGFs:

$$F^{\text{var1}}(s,t) := E(s^{Z(t)}|Z(0) = 1) \qquad \text{and} \qquad F_{\text{sib}}^{\text{var1}}(s,t) := E(s^{Z(t)}|Z(0) = 2 \text{ siblings}).$$

For the former, we have the natural generalization to that in equation (1) for the Cyton Model:

$$F^{\text{var1}}(s,t) = sP(T_D > t, T_B > t) + P(T_D \leq t, T_B \geq T_D) + \int_0^t F_{\text{sib}}^{\text{var1}}(s,t-r)dP(T_D > r, T_B = r). \tag{7}$$

However, the evolution for $F_{\text{sib}}^{\text{var1}}(s,t)$ is somewhat more involved. With T_D^1 and T_D^2 being the times to death of two siblings who share a single time to division T_B, we have that

$$\begin{aligned}
F_{\text{sib}}^{\text{var1}}(s,t) = {} & s^2 P(\min(T_D^1, T_D^2) > t, T_B > t) \\
& + sP\left(\min(T_D^1, T_D^2) \leq t, T_B > t, \max(T_D^1, T_D^2) > t\right) \\
& + P\left(\max(T_D^1, T_D^2) \leq \min(T_B, t),\right) \\
& + \int_0^t F_{\text{sib}}^{\text{var1}}(s,t-r)dP\left(\min(T_D^1, T_D^2) \leq r, T_B = r, \max(T_D^1, T_D^2) > r\right) \\
& + \int_0^t F_{\text{sib}}^{\text{var1}}(s,t-r)^2 dP(T_B = r, \min(T_D^1, T_D^2) > r).
\end{aligned} \tag{8}$$

The terms can be described as follows: both siblings are still alive; one sister has died, the other has neither died nor divided; both siblings have died; one sister has died and the other divided; both siblings have divided. It is in the last integral where cousins appear. By inserting the $F_{\text{sib}}^{\text{var1}}(s,t-r)^2$ term, we are asserting the independence of cousins. It is here that an alternative PGF term would be inserted if we wished cousins to have dependencies.

Note that as

$$\frac{\partial}{\partial s}F_{\text{sib}}^{\text{var1}}(s,t)|_{s=1} = 2E(Z(t)) = 2m(t),$$

taking derivatives of equation (7) with respect to s, we see that $m(t)$ evolves as equation (3). That is, as in the original Cyton Model. The variance, however, is governed by a different equation to that for the Cyton Model, equation (4). With

$$v_{\text{sib}}^{\text{var1}}(t) := \frac{\partial^2}{\partial s^2}F_{\text{sib}}^{\text{var1}}(s,t)|_{s=1},$$

taking partial derivatives of equation (7) with respect to s, we have that

$$v^{\text{var1}}(t) := \frac{\partial^2}{\partial s^2}F^{\text{var1}}(s,t)|_{s=1} = \int_0^t v_{\text{sib}}^{\text{var1}}(t-r)dP(T_D > r, T_B = r).$$

Finally we obtain an integral equation for $v_{\text{sib}}(t)$ by taking partial derivatives of equation (8) with respect to s. Using the identity $P(A \cap B) + P(A \cup B) = P(A) + P(B)$, we get

$$v_{\text{sib}}^{\text{var1}}(t) = 2P(\min(T_D^1, T_D^2) > t, T_B > t) + 2\int_0^t v_{\text{sib}}^{\text{var1}}(t-r)dP(T_B = r, T_D^1 > r)$$

$$+ 8\int_0^t m(t-r)^2 dP(\min(T_D^1, T_D^2) > r, T_B = r).$$

From these expressions, it is unclear if var(t) is always higher in the Variant 1 model.

For the simple example with exponentially distributed and independent times to die and times to divide that was introduced in Section 2.1, var(t) can be calculated explicitly. If $\lambda_B \neq \lambda_D$, then

$$\text{var}^{\text{var1}}(t) = \frac{(5\lambda_B + 3\lambda_D)}{3(\lambda_B - \lambda_D)}\left(e^{2(\lambda_B - \lambda_D)t} - e^{(\lambda_B - \lambda_D)t}\right) + \frac{2\lambda_B}{3(2\lambda_B + \lambda_D)}\left(e^{-(\lambda_B + 2\lambda_D)t} - e^{(\lambda_B - \lambda_D)t}\right), \tag{9}$$

and if $\lambda_B = \lambda_D$, then

$$\text{var}^{\text{var1}}(t) = \frac{8\lambda_B t}{3} - \frac{2}{9}\left(1 - e^{-3\lambda_B t}\right). \tag{10}$$

It can be shown that regardless of λ_B and λ_D the variance in equations (9) and (10) for this variant is greater than that in equation (5) for the Cyton Model.

3.2 Cyton Model Variant 1; Generational PGF Evolution

As in Section 2.2, we introduce an iterative scheme for determining the PGF of the population size per generation for the Variant 1 model.

Again it is more convenient to work with MGFs than PGFs, with the MGF $M^{\text{var1}}(\theta, t)$ taking the role of the PGF $F^{\text{var1}}(e^\theta, t)$ and $H^{\text{var1}}(\theta, t)$ taking on the role of $F_{\text{sib}}^{\text{var1}}(e^\theta, t)$. With $Z_k(t)$ denoting the number of cells of generation k that are alive at time t. We must evaluate the moment generating

function of random vector of cells in generations k through l at time t, $(Z_k(t), \ldots, Z_l(t))$, starting with one cell of generation k at time 0. Define $M^{\mathrm{var1}}(\boldsymbol{\theta}_k^l, t) := 1$ if $l < k$ and if $l \geq k$ define

$$M^{\mathrm{var1}}(\boldsymbol{\theta}_k^l, t) := E\left(e^{\sum_{j=k}^l \theta_j Z_j(t)} | Z_k(0) = 1, \ Z_i(0) = 0 \text{ if } i \neq k\right)$$

Unlike in the Cyton Model model, we also need to define the MGF of the random vector of cells in generations k through l at time t starting with two siblings of generation k at time 0. Define $H^{\mathrm{var1}}(\boldsymbol{\theta}_k^l, t) := 1$ if $l < k$ and if $l \geq k \geq 1$ define

$$H^{\mathrm{var1}}(\boldsymbol{\theta}_k^l, t) := E\left(e^{\sum_{j=k}^l \theta_j Z_j(t)} | Z_k(0) = 2 \text{ siblings}, \ Z_i(0) = 0 \text{ if } i \neq k\right) \ \forall \, k \in \mathbb{N}, l \in \mathbb{Z}_+.$$

We insist on $k \geq 1$ as cells from the zeroth/founder generation are assumed to not be siblings, but are stochastically independent.

Following similar logic to the integral equations (7) and (8), by considering the possible fates and likelihoods of those fates for the cells under consideration, the inter-generational relationships are:

$$M^{\mathrm{var1}}(\boldsymbol{\theta}_k^l, t) = e^{\theta_k} P(T_D^k > t, T_B^k > t) + P(T_D^k \leq t, T_B^k \geq T_D^k)$$
$$+ \int_0^t H^{\mathrm{var1}}(\boldsymbol{\theta}_{k+1}^l, t-r) dP(T_B^k = r, T_D^k > r)$$
$$H^{\mathrm{var1}}(\boldsymbol{\theta}_k^l, t) = e^{2\theta_k} P(\min(T_D^{k,1}, T_D^{k,2}) > t, T_B^k > t)$$
$$+ e^{\theta_k} P(\min(T_D^{k,1}, T_D^{k,2}) \leq t, T_B^k > t, \max(T_D^{k,1}, T_D^{k,2}) > t)$$
$$+ P(\max(T_D^{k,1}, T_D^{k,2}) \leq \min(T_B^k, t))$$
$$+ \int_0^t H^{\mathrm{var1}}(\boldsymbol{\theta}_{k+1}^l, t-r) dP(\min(T_D^{k,1}, T_D^{k,2}) \leq r, T_B^k = r, \max(T_D^{k,1}, T_D^{k,2}) > r)$$
$$+ \int_0^t H^{\mathrm{var1}}(\boldsymbol{\theta}_{k+1}^l, t-r)^2 dP(T_B^k = r, \min(T_D^{k,1}, T_D^{k,2}) > r)$$

Division destiny is again represented by random the variable K. If $K = k^*$, then division cannot happen for any cell in this family tree once generation k^* is reached. Furthermore, again, we assume a different time to die random variable $\tilde{T}_D^{k^*}$ for the time to die distribution once division destiny sets in. In this setting we must determine

$$\tilde{M}^{\mathrm{var1}}(\boldsymbol{\theta}_k^{k^*}, t) := E\left(e^{\sum_{j=k}^{k^*} \theta_j Z_j(t)} | Z_k(0) = 1, \ Z_i(0) = 0 \text{ if } i \neq k\right) \text{ for } k = 0, 1, \ldots, k^* \text{ and}$$

$$\tilde{H}^{\mathrm{var1}}(\boldsymbol{\theta}_k^{k^*}, t) := E\left(e^{\sum_{j=k}^{k^*} \theta_j Z_j(t)} | Z_k(0) = 2 \text{ siblings}, \ Z_i(0) = 0 \text{ if } i \neq k\right) \text{ for } k = 1, \ldots, k^*$$

and for all k^* in the support of K. These are derived by iterating the following scheme:

1. Initialize by setting $k = k^*$, $\tilde{M}^{\mathrm{var1}}(\theta_{k^*}, t) = 1 + (e^{\theta_{k^*}} - 1) P(\tilde{T}_D^{k^*} > t)$, and if $k^* > 0$ also

$$\tilde{H}^{\mathrm{var1}}(\boldsymbol{\theta}_{k^*}, t) = e^{2\theta_{k^*}} P(\min(\tilde{T}_D^{k^*,1}, \tilde{T}_D^{k^*,2}) > t) + P(\max(\tilde{T}_D^{k^*,1}, \tilde{T}_D^{k^*,2}) \leq t)$$
$$+ e^{\theta_{k^*}} P(\min(\tilde{T}_D^{k^*,1}, \tilde{T}_D^{k^*,2}) \leq t, \max(\tilde{T}_D^{k^*,1}, \tilde{T}_D^{k^*,2}) > t).$$

2. Set $k = k - 1$ and calculate

$$
\begin{aligned}
\tilde{M}^{\mathrm{var1}}(\boldsymbol{\theta}_k^{k^*}, t) = & e^{\theta_k} P(T_D^k > t, T_B^k > k) + P(T_D^k \leq t, T_B^k \geq T_D^k) \\
& + \int_0^t \tilde{H}^{\mathrm{var1}}(\boldsymbol{\theta}_{k+1}^{k^*}, t - r) dP(T_B^k = r, \, T_D^k > r).
\end{aligned}
$$

If $k > 0$, also calculate the following

$$
\begin{aligned}
\tilde{H}^{\mathrm{var1}}(\boldsymbol{\theta}_k^{k^*}, t) = & \, e^{2\theta_k} P(\min(T_D^{k,1}, T_D^{k,2}) > t, T_B^k > t) \\
& + e^{\theta_k} P(\min(T_D^{k,1}, T_D^{k,2}) \leq t, T_B^k > t, \max(T_D^{k,1}, T_D^{k,2}) > t) \\
& + P(\max(T_D^{k,1}, T_D^{k,2}) \leq \min(T_B^k, t)) \\
& + \int_0^t \tilde{H}^{\mathrm{var1}}(\boldsymbol{\theta}_{k+1}^l, t - r) dP(\min(T_D^{k,1}, T_D^{k,2}) \leq r, T_B^k = r, \max(T_D^{k,1}, T_D^{k,2}) > r) \\
& + \int_0^t \tilde{H}^{\mathrm{var1}}(\boldsymbol{\theta}_{k+1}^l, t - r)^2 dP(T_B^k = r, \min(T_D^{k,1}, T_D^{k,2}) > r).
\end{aligned}
$$

3. If $k > 0$, then repeat the previous step and otherwise terminate.

Assuming that K has finite support with upper limit $K_{\max}$, we can derive the moment generating function starting with one founder cell at time 0 by evaluating:

$$
\bar{M}^{\mathrm{var1}}(\boldsymbol{\theta}_0^{K_{\max}}, t) := \sum_{k^*=0}^{K_{\max}} P(K = k^*) \tilde{M}^{\mathrm{var1}}(\boldsymbol{\theta}_0^{k^*}, t). \tag{11}
$$

As in Section 2.2 for the Cyton Model, with the distributions of $K, \{T_B^k, T_D^k\}$ given we can follow this scheme to determine $\bar{M}^{\mathrm{var1}}(\boldsymbol{\theta}_0^{K_{\max}}, t)$ in equation (11). In Section 6 we will use this scheme to determine the predicted variance of this model for distributions that have been proposed to match with data.

4 Cyton Model Variant 2, second cousins with correlated division and death

Variant 2 introduces a stronger correlation structure than Variant 1. All siblings, cousins, and second cousins experience the same fate at the same time. All more distant relations (third cousins etcetera) are independent. It is possible to consider this dependency up to any finite depth in the family tree. We report on second cousins as this is the least amount of dependency that gives rise to predictions that are recognizably distinct from the Cyton Model for parameterizations from data.

The Cyton Model Variant 2 hypotheses are:

− Each founder cell is stochastically independent.
− For each pair of siblings, quad of cousins and octet of second cousins there is a single, shared, time to divide and time to die pair, drawn from the appropriate distributions. More distant relations are stochastically independent.
− There is a division destiny for each founding cell in the starting population that is drawn from a stochastic distribution.

If founding cells were not independent, but constituted independent groups of eight second cousins, the impact of these revised assumptions would be that for a given number of founding cells the variance predicted by this Variant would be eight times that in the Cyton Model. This occurs as each octet of second cousins has a family tree based on the Cyton Model, but with each element representing eight second cousins, so that the population is eight times that of the Cyton Model. Thus the variance is sixty four times that of the Cyton Model. However, we should compare an octet of founding second cousins with an octet of founding Cyton Model cells, so that the increase in variance due to this additional dependency is eight times that of Cyton Model. Despite this back of the envelope calculation, for the flow cytometry parameterizations, we find that the independence of founding cells (the assumption that they are not related) dramatically reduces this difference in predicted variance to far less than a factor of twice that predicted by the Cyton Model. However, we do show that the variance of this model (or indeed, any model with independence beyond siblings or cousins etcetera) is necessarily higher than that in the Cyton Model.

4.1 Cyton Model Variant 2 without division destiny, homogeneous analytic example

This model Variant, with dependencies between second cousins, can be determined through first solving for the PGF of the Cyton Model in equation (1) and then taking three integrals:

$$F^{\mathrm{var2}}(s,t) = sP(T_B > t, T_D > t) + P(T_D \leq t, T_B \geq T_D) + \int_0^t F_{\mathrm{sib}}^{\mathrm{var2}}(s, t-r)dP(T_D > r, T_B = r)$$

$$F_{\mathrm{sib}}^{\mathrm{var2}}(s,t) = sP(T_B > t, T_D > t) + P(T_D \leq t, T_B \geq T_D) + \int_0^t F_{\mathrm{cous}}^{\mathrm{var2}}(s, t-r)dP(T_D > r, T_B = r)$$

$$F_{\mathrm{cous}}^{\mathrm{var2}}(s,t) = sP(T_B > t, T_D > t) + P(T_D \leq t, T_B \geq T_D) + \int_0^t F_{\mathrm{2nd\text{-}cous}}^{\mathrm{var2}}(s, t-r)dP(T_D > r, T_B = r)$$

$$F_{\mathrm{2nd\text{-}cous}}^{\mathrm{var2}}(s,t) = F^{\mathrm{cyt}}(s^8, t),$$

where F^{cyt} is the PGF for the Cyton Model given in equation (1). This final equality occurs as second cousins follow the rules of the Cyton Model, but with each Cyton corresponding to an octet of second cousins.

In order to see that the variance of this model is always larger than that of the Cyton Model, it suffices to note that the only difference between the Cyton Model and this Variant occurs in the integrand in the final integral. Compare the Variant 2 term

$$\frac{\partial^2}{\partial s^2} F^{\mathrm{cyt}}(s^8, t)|_{s=1} = 56m(t) + 64v^{\mathrm{cyt}}(t) = 56(m(t) + v^{\mathrm{cyt}}(t)) + 8v^{\mathrm{cyt}}(t)$$

with the corresponding term in the Cyton Model where there are eight independent second cousins

$$\frac{\partial^2}{\partial s^2} F^{\mathrm{cyt}}(s, t)^8|_{s=1} = 56m(t)^2 + 8v^{\mathrm{cyt}}(t).$$

As $\mathrm{var}^{\mathrm{cyt}}(t) = v^{\mathrm{cyt}}(t) + m(t) - m(t)^2 \geq 0$, the variance for this variant is always greater than that of the Cyton Model. A similar argument holds for any depth of correlation (siblings, cousins, etcetera) of this sort. Thus this finite-depth correlation structure never leads to a decrease in variance.

For completeness, we give expressions for the variance in the simple example introduced in Section 2.1. If $\lambda_B \neq \lambda_D$ and $3\lambda_B \neq \lambda_D$ then

$$\text{var}^{\text{var2}}(t) = \frac{\lambda_B + \lambda_D}{3\lambda_B - \lambda_D}\left(8\lambda_B^2 t^2 + \frac{4(13\lambda_B - 3\lambda_D)}{3\lambda_B - \lambda_D}\lambda_B t + \frac{115\lambda_B^2 - 54\lambda_B\lambda_D + 7\lambda_D^2}{(3\lambda_B - \lambda_D)^2}\right)e^{-(\lambda_B + \lambda_D)t}$$
$$+ \frac{(\lambda_B + \lambda_D)(101\lambda_B^3 - 47\lambda_B^2\lambda_D + 11\lambda_B\lambda_D^2 - \lambda_D^3)}{(\lambda_B - \lambda_D)(3\lambda_B - \lambda_D)^3}e^{2(\lambda_B - \lambda_D)t} - \frac{8(\lambda_B + \lambda_D)}{\lambda_B - \lambda_D}e^{(\lambda_B - \lambda_D)t}.$$

If $\lambda_B = \lambda_D$ or $3\lambda_B = \lambda_D$, $\text{var}^{\text{var2}}(t)$ can be determined by taking limits in the equation given above.

4.2 Cyton Model Variant 2, generational PGF evolution

For Variant 2, the generational approach is similar to that for Variant 1, so we do not expound on it. It suffices to say that one first determines the MGF for the Cyton Model as described in Section 2.2, but starting with the distributions of second cousins, to identify the moment generating function for second cousins. The joint moment generating function for the entire population can then be determined by taking the obvious integrals.

5 Cyton Model Variant 3, correlations within generations

Variant 3 introduces the strongest correlation structure that we consider. Within each family tree, all cells in a given generation experience the same fate at the same time, but the experience of cells in distinct generations are independent. Thus we replace the Cyton Model hypotheses with the following:

- Each founder cell is stochastically independent.
- For all the progeny of a given founder cell in a single generation, a single, shared, new time to divide and time to die are drawn from the appropriate distributions.
- The operation of the regulable cyton controlling division and survival, seen leading up to the first division, is repeated through subsequent divisions.
- There is a division destiny for each founding cell in the starting population that is drawn from a stochastic distribution.

5.1 Cyton Model Variant 3 without division destiny, homogeneous analytic example

We first consider the consequences of these revised hypotheses on the fundamental equation that determine the evolution of the PGF for a homogeneous population in the absence of division destiny, where (T_B, T_D) are chosen independently for each generation and from the same distribution.

Let $Z(t)$ denote the total number of cells alive at time t. We consider the equation that determines the PGF $F^{\text{var3}}(s,t) := E(s^{Z(t)}|Z(0) = 1)$:

$$F^{\text{var3}}(s,t) = sP(T_B > t, T_D > t) + P(T_D \leq t, T_B \geq T_D) + \int_0^t F^{\text{var3}}(s^2, t - r)dP(T_D > r, T_B = r).$$
(12)

These terms can be understood as in Section 2.1, where the difference between the Cyton Model and this variant is in the final integral. The integrand is now

$$E(s^{Z^1(t)+Z^2(t)} \mid Z^1(0) = Z^2(0) = 1) = E(s^{2Z^1(t)} \mid Z^1(0) = 1) = E(s^{2Z(t)} \mid Z(0) = 0) = F^{\mathrm{var3}}(s^2, t),$$

having used the fact that siblings share the same time to divide and time to die. This can be contrasted with the expression in equation (2). The equation (12) has the effect that this propagates and all descendents of a given founder in a single generation have the same time to die and time to divide.

For the original Cyton Model, equations (3) and (4) describe the evolution of the mean population size, $m(t)$, and variance $\mathrm{var}^{\mathrm{var3}}(t)$. For the new hypotheses, we get the same evolution of the mean (equation (3)), but a different equation for the variance since the following equation holds:

$$v^{\mathrm{var3}}(t) = \int_0^t 2\left(m(t-r) + 2v^{\mathrm{var3}}(t-r)\right) dP(T_B = r, T_D > r). \tag{13}$$

It is the case that $v^{\mathrm{var3}}(t)$ is necessarily greater than or equal to $v^{\mathrm{cyt}}(t)$ in equation (4). This can be seen by considering $v^{\mathrm{var3}}(t) - v^{\mathrm{cyt}}(t)$, which is 0 when $t = 0$, and noting that $m(t-r) - m(t-r)^2 + v^{\mathrm{var3}}(t-r) = \mathrm{var}^{\mathrm{var3}}(t-r) \geq 0$:

$$v^{\mathrm{var3}}(t) - v^{\mathrm{cyt}}(t) = \int_0^t \left(\mathrm{var}^{\mathrm{var3}}(t-r) + 2(v^{\mathrm{var3}}(t) - v^{\mathrm{cyt}}(t))\right) dP(T_B = r, T_D > r) \geq 0.$$

If the PGF $F^{\mathrm{var3}}(s, t)$ is finite in an open s neighborhood of 1, then it completely specifies the distribution of $Z(t)$, e.g. [10]. Analytically, particular care must be taken with equation (12) to ensure this is the case, as we now illustrate based on the example in Section 2.1. For the particular case where T_B is exponentially distributed with rate λ_B and T_D is independent and exponentially distributed with parameter λ_D, the distribution of $Z(t)$ for this variant can be calculated directly. Let $\{T_D^i\}$ denote the sequence of death times and $\{T_B^i\}$ the sequence of division times so that $L_i := \min\{T_D^i, T_B^i\}$ is the lifetime of the ith generation. For $n \geq 1$, the probability that $Z(t) = 2^n$ is the probability that n divisions have occurred:

$$P(Z(t) = 2^n) = P(L_0 + \ldots + L_{n-1} \leq t, L_0 + \ldots + L_n > t, T_B^i < T_D^i \text{ for all } i \in \{0, \ldots, n-1\}) \tag{14}$$

$$= P(L_0 + \ldots + L_{n-1} \leq t, L_0 + \ldots + L_n > t) P(T_B^i < T_D^i \text{ for all } i \in \{0, \ldots, n-1\})$$

$$= \frac{(\lambda_B + \lambda_D)^n t^n}{n!} e^{-(\lambda_B + \lambda_d)t} \left(\frac{\lambda_B}{\lambda_B + \lambda_D}\right)^n$$

$$= \frac{(\lambda_B t)^n}{n!} e^{-(\lambda_B + \lambda_d)t}$$

and $P(Z(t) = 0) = 1 - \sum_{n \geq 0} P(Z(t) = 2^n) = 1 - \exp(-\lambda_D t)$. This distribution has every moment finite as

$$E(Z(t)^k) = \sum_{n \geq 1} (2^n)^k P(Z(t) = 2^n) = e^{((2^k - 1)\lambda_B - \lambda_D)t}, \tag{15}$$

but $F^{\mathrm{var3}}(s, t) = \sum_{n=0}^{\infty} s^n P(Z(t) = 2^n)$ is infinite for all $s > 1$. This difficulty arises as the probability that a large number of generations pass in a brief period of time does not decrease sufficiently fast to for the expectation to be finite. In practical cell systems this concern does not arise due to the lower bound on the time between cell divisions [23]. For this example $\mathrm{var}^{\mathrm{var3}}(t)$ can be calculated explicitly using equation (15):

$$\mathrm{var}^{\mathrm{var3}}(t) = e^{(3\lambda_B - \lambda_D)t} - e^{2(\lambda_B - \lambda_D)t}. \tag{16}$$
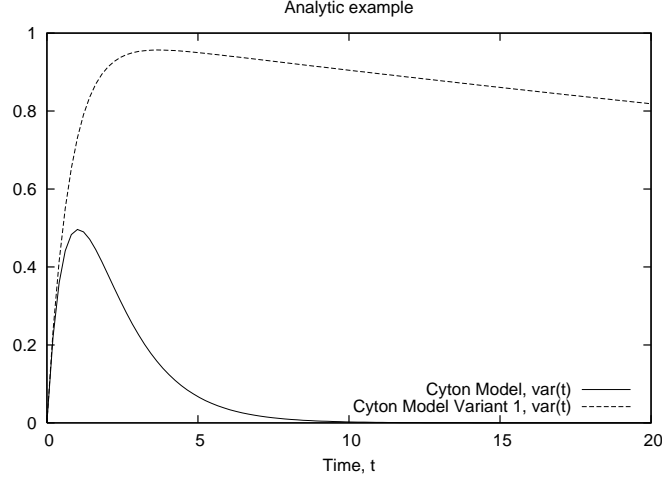
Fig. 2 Variance comparison between Cyton Model and Variant 3. Independent exponentially distributed time to die, $\lambda_D = 1$, and time to divide $\lambda_B = 0.33$. Demonstrates qualitative difference in variances for large t

Comparing the Cyton Model variance in equation (5) with the variance from the Variant 3 model in equation (16), the variance of the Variant 3 model decreases significantly more slowly at large t. This is illustrated in Figure 2 when $\lambda_D = 1$ and $\lambda_B = 0.33$.

5.2 Cyton Model Variant 3, generational PGF evolution

As in Sections 2.2, 3.2 and 4.2, we now present an iterative scheme that enables us to numerically evaluate the PGF for the population size. Using equation (14) we can explicitly write down a formula for the MGF, but in practice it is easiest to evaluate it using the following recursive approach.

For each $k \in \mathbb{Z}_+$ let T_D^k denote the time to die and T_B^k denote the time to division random variables. It will again prove more convenient to work with Moment Generating Functions (MGFs), $M^{\text{var}3}(\theta, t) := F^{\text{var}3}(e^\theta, t)$. Let $Z_j(t)$ denote the number of cells of generation j that are alive at time t. We consider the MGF of random vector of cells in generations k through l at time t, $(Z_k(t), \ldots, Z_l(t))$, after starting with one cell of generation k at time 0. Defining $M^{\text{var}3}(\boldsymbol{\theta}_k^l, t) := 1$ if $k > l$, for $k \leq l$ this MGF is defined by

$$M^{\text{var}3}(\boldsymbol{\theta}_k^l, t) := E\left(e^{\sum_{j=k}^{l} \theta_j Z_j(t)} | Z_k(0) = 1, \ Z_i(0) = 0 \text{ if } i \neq k\right),$$

where $\boldsymbol{\theta}_k^l$ is the vector $(\theta_k, \theta_{k+1}, \ldots, \theta_l) \in \mathbb{R}^{l-k+1}$.

Considering all the possible fates of the initial cell from generation k that could occur until time t we get the following MGF relationship across generations:

$$M^{\text{var}3}(\boldsymbol{\theta}_k^l, t) = e^{\theta_k} P(T_D^k > t, T_B^k > k) + P(T_D^k \leq t, T_B^k \geq T_D^k)$$
$$+ \int_0^t M^{\text{var}3}(2\boldsymbol{\theta}_{k+1}^l, t - r) dP(T_B^k = r, \ T_D^k > r)$$

where $2\boldsymbol{\theta}_{k+1}^{l} = (2\theta_{k+1}, 2\theta_{k+2}, \ldots, 2\theta_{l})$ and we have used the fact that all cells in a given generation have the same time to die and divide.

To include division destiny, we must first derive

$$\tilde{M}^{\text{var3}}(\boldsymbol{\theta}_{k}^{k^*}, t) := E\left(e^{\sum_{j=k}^{k^*} \theta_j Z_j(t)} | Z_k(0) = 1, \ Z_i(0) = 0 \text{ if } i \neq k\}\right) \text{ for } k = 0, 1, \ldots, k^*.$$

These must be evaluated for all k^* in the support of K. These are determined in turn by each iteration given below:

1. Initialize by setting $k = k^*$ and $\tilde{M}^{\text{var3}}(\theta_{k^*}, t) = 1 + (e^{\theta_{k^*}} - 1)P(\tilde{T}_D^{k^*} > t)$.
2. Set $k = k - 1$ and calculate

$$\tilde{M}^{\text{var3}}(\boldsymbol{\theta}_{k}^{k^*}, t) = e^{\theta_k} P(T_D^k > t, T_B^k > k) + P(T_D^k \leq t, T_B^k \geq T_D^k)$$
$$+ \int_0^t M^{\text{var3}}(2\boldsymbol{\theta}_{k+1}^{l}, t - r) dP(T_B^k = r, \ T_D^k > r).$$

3. If $k > 0$, then repeat the previous step and otherwise terminate.

Once again, letting $K_{\max} < \infty$ denote the maximum value in the support of K, we are now in a position to determine the MGF of the population starting with one founding cell at time 0 by performing the following:

$$\bar{M}^{\text{var3}}(\boldsymbol{\theta}_{0}^{K_{\max}}, t) := \sum_{k^*=0}^{K_{\max}} P(K = k^*) \tilde{M}(\boldsymbol{\theta}_{0}^{k^*}, t). \tag{17}$$

As in Section 2.2 for the Cyton Model, with the distributions of K and $\{T_B^k, T_D^k\}$, for each $k \geq 0$, given we can follow this this scheme to determine $\bar{M}^{\text{var3}}(\boldsymbol{\theta}_0^{K_{\max}}, t)$ in equation (17). If K is unbounded and there is not a minimum time between cell divisions, then care needs to be taken to ensure that $\bar{M}$ is finite as explained in the example in Section 5.1.

6 Comparison with data

We compare the predictions of the Cyton Model and each of its variants based on parameterizations determined from two sets of experiments. One set is from *in vitro* experimental data reported in Turner, Hawkins and Hodgkin [7] that follows B lymphocytes stimulated by 3 μM CpG DNA. The other set of data is from an *in vivo* experiment reported on in Homann, Teyton and Oldstone [16] where CD8+ T lymphocytes specific to a single peptide/MHC epitope were followed after infection with lymphocytic choriomeningitis virus.

We use the Cyton Model parameterizations for these experiments based on those given in [6]. It has been reported [5][6] that the empirical clock distributions appear to be well matched by members of the family of log-normal distributions. For example, $P(T_D > t) \approx P(N(\mu, \sigma^2) > \log t)$, where $N(\mu, \sigma^2)$ is a Normally distributed random variable with mean μ and variance σ^2. Assuming the T_B and T_D random variables are independent, with the (μ, σ^2) information for each random variable it is possible to make predictions from each model variant that can be compared with the data.

In the *in vitro* experiments, approximately fifteen thousand CFSE-labeled purified B-cells were exposed to the mitogenic stimulus. Roughly 10% of the founding cells do not respond to the stimulus. The

overall cell population size were recorded from the introduction of the stimulus to beyond division destiny by flow cytometric analysis [6].

In the *in vivo* experiment CD8+ T lymphocytes specific to a single peptide/MHC epitope were followed after infection with lymphocytic choriomeningitis virus. The overall cell population size was recorded from the introduction of the virus to beyond division destiny and on to the period where remaining cell numbers are retained at a homeostatic level. The true starting cell number is unknown, although it is estimated to be approximately one hundred [24][6]. This starting value of approximately one hundred is a minimal estimate for a virus response, as it represents a single epitope and is representative of similar data that could be collected on at least four other epitopes, bringing the population response to an initial number over five hundred [16][24].

The *in vitro* division destiny has the probability mass function given in Figure 3, with $K_{\max} = 8$. The division destiny distribution is such that a maximum of 8 generations can occur. The complete list of clock parameters are:

- Time to division of founding cells: mean 34.86 hours (including 1 hour minimum); standard deviation 4.9 hours. Time to death of founding cells: mean 151.27 hours; standard deviation 19.75 hours.
- Time to division of cells in each subsequent generation: mean 7.01 hours (including 1 hour minimum); standard deviation 1.27 hours. Time to death of cells in each subsequent generation: mean 83.93 hours; standard deviation 113.55 hours.

The *in vivo* division destiny has the probability mass function given in Figure 3, with $K_{\max} = 27$. The division destiny distribution is such that (barring death by apoptosis) a minimum of 12 and maximum of 27 generations can occur. The complete list of clock parameters are:

- Time to division of founding cells: mean 40.44 hours (including 1 hour minimum); standard deviation 7.55 hours. Time to death of founding cells: mean 400.50 hours; standard deviation 20.04 hours.
- Time to division of cells in each subsequent division: mean 8.73 hours (including 1 hour minimum); standard deviation 0.70 hour. Time to death cells in each subsequent division: mean 67.56 hours; standard deviation 59.94 hours.

Figure 4 plots the predicted variance for the cell population starting with a single founding cell for both the *in vitro* and *in vivo* experiments. For these parameterizations, the difference between the variance in the original Cyton Model and the Variant 1 model is miniscule, so the latter is not shown. The Variant 2 model, where all relations not more distant than second cousins share a time to divide and die, results in greater variance for the *in vitro* data, but is near identical to the Cyton Model predictions for the *in vivo* data. This suggests that local dependencies have little impact on variability in population size once a sufficient number of divisions have occurred, as there is enough independence in the population to limit the effect of the local dependencies. If the founding cells were all second cousins and the Variant 2 model was used, the variance would be eight times that of the Cyton Model. Clearly the reduction that occurs as founders are assumed independent has a significant impact. The Variant 3 model, where all collaterally consanguineous cells experience the same fate, gives a dramatic increase in variance, particularly at longer time-scales. Qualitatively, note that its variance decays more slowly at large times than those of the other models.

For 13,500 founding *in vitro* and 90 founding *in vivo* cells, the upper two plots in Figure 5 compare the predicted variability in the overall population based on the Cyton Model and Variant 3, which has the greatest impact on variance. Variants 1 and 2, which have local dependencies, are not plotted as they are not easily distinguished from the Cyton Model on this scale. The Variant 3 model, however,
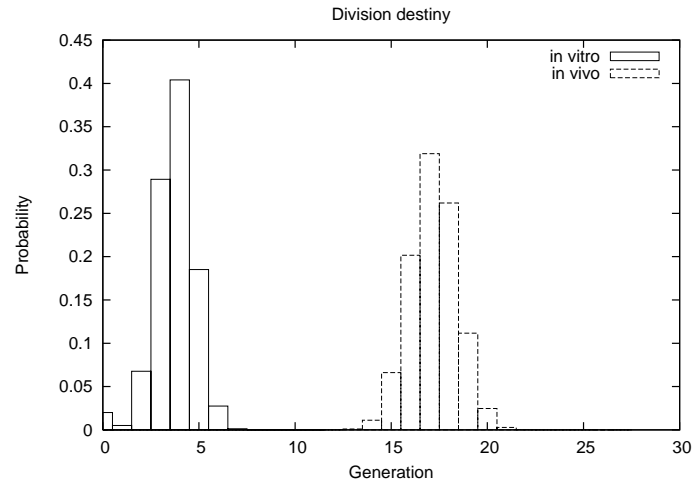
Fig. 3 *In vitro* and *in vivo* division destiny probability mass functions
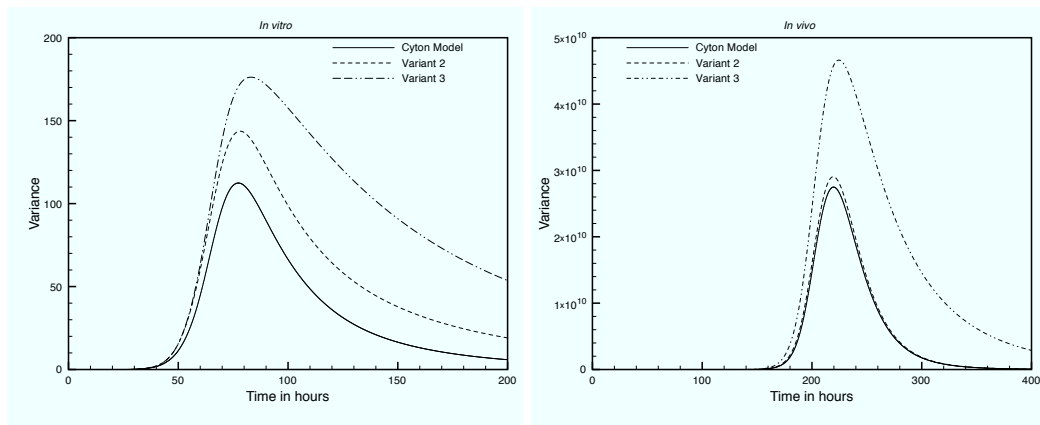

Fig. 4 Variance comparison for a single founding cell based on *in vitro* and *in vivo* data parameterizations. Cyton Model, Variant 2 and Variant 3 shown. Variant 1 indistinguishable from Cyton Model so is not shown.

produces wider error bars for both experiments. This can be understood as, unlike the other variants, cell division does not lead to additional statistical averaging beyond that from the independence of the founding cells. This can most easily be seen for the *in vivo* data as there is a small number of founding cells whose family trees undergo a large number of cell divisions. When the population size increases, there is no additional statistical averaging beyond that available in the original population in order to reduce variance.

To better see the variability predicted by the models, the lower two plots in Figure 5 show variability around the mean population size. Their content is the similar to that in the upper two plots, but with the model's mean prediction subtracted from all lines and the data. Only the Cyton Model, Variant 2 and Variant 3 are shown, as Variant 1 is indistinguishable from the Cyton Model. For both the *in vitro* and *in vivo* data, the wider error bars of the Variant 3 model includes more data points. This
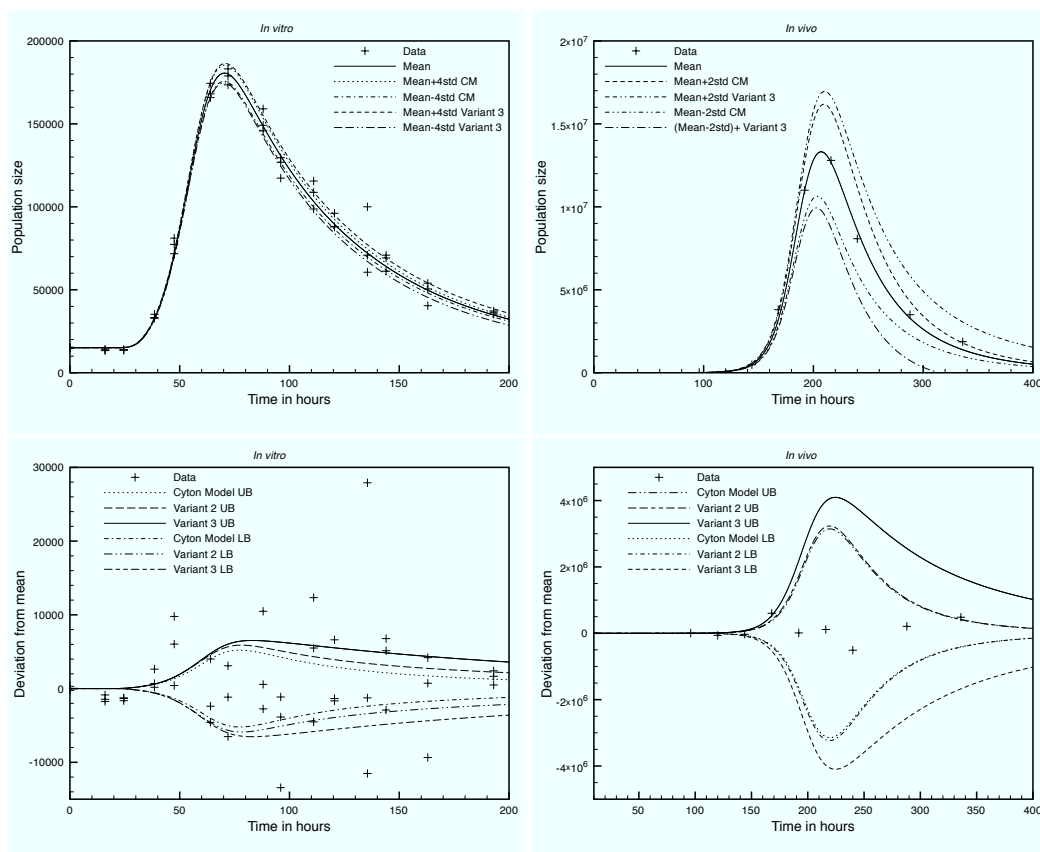
Fig. 5 Top two graphs show predicted variability comparison for *in vitro* and *in vivo* data. Cyton Model and Variant 3 shown. Variant 3 shows qualitatively different behavior at large times. Lower two graphs show predicted variability around mean for *in vitro* and *in vivo* data. Cyton Model, Variant 2 and Variant 3 shown.

suggests that some of the variability seen in data that had previously been attributed to experimental error [8] could, in the presence of dependencies within family trees, be due to intrinsic variability in the cell-system population dynamics.

7 Discussion and conclusions

Since their introduction, branching processes have been used extensively in the mathematical analysis of cell population dynamics. For certain cell systems, one of the key assumptions of branching processes, the stochastic independence of siblings, has been known not to hold since Powell's work [25] on bacterial cells in the mid 1950s (see also Staudte et al. [26] for a modern reinvestigation of Powell's data). Powell discovered independence across generations, but correlation in the lifetimes of siblings. Kubitschek [27] later found that these positive correlations extend to the lifetimes of cousins and second cousins also. This work motivated, for example, the introduction by Crump and Mode [28] and Olofsson [29] of generalized Bellman-Harris processes with local dependencies, but with the

retention of the standard assumption of independent lifetime and progeny number distributions. They do not consider transient dynamics, but prove limit theorems in the absence of replicative senescence or division destiny.

Data from both flow cytometry [6] and optical experiments [9] studying the response of lymphocytes to a mitogenic signal demonstrate that two of the standard assumptions of continuous time branching processes are inappropriate for this cell system: the independence of lifetime and progeny number random variables, and the stochastic independence of siblings. In modeling cell system dynamics, the independence of lifetime and progeny number is typically assumed for one of two reasons: (I) an absence of mechanistic reasons to motivate why there should be dependence; or (II) to ensure that the resulting model is identifiable from experimentally observable quantities. For lymphocytes, the data that led to the Cyton Model [6] gives biological motivation for a mechanistic structure that gives rise to dependency asked for in (I). For (II), given a solid explanation for dependence between lifetime and progeny number random variables, assuming they are independent in order to keep a model identifiable seems to have questionable merits. Thankfully, based on data from optical experiments of the sort described in [9], the resulting models are identifiable.

General mathematical frameworks for not-necessarily binary branching processes that treat populations where lifetime and offspring random variables that need not be independent were introduced by Crump and Mode [30] and Jagers [31]. The dependencies there are not described as arising from competing clocks as in the Cyton Model, but are motivated by human reproduction where offspring can occur before death of the the parent. Good introductions to these frameworks and their applications can be found in Jagers [31] and Haccou, Jagers and Vatutin [32].

Biologically reasonable mechanisms have been proposed that naturally give rise to dependencies between the lifetimes of siblings, but those that we are aware of would not explain strong dependencies between collaterally consanguineous cells in conjunction with weak mother-daughter dependence. For example, one well explored approach is setting the probability of division to be a function of cell-size and, at division, sharing cell-size amongst progeny, e.g. Clifford and Sudbury [33,34], Heijmans [35], Alt and Tyson [36] Tyson [37] Huggins and Staudte [38]. In these models, size is a loose label that can be considered short-hand for mass, quantity of a particular enzyme, etcetera. The dependencies that arise from models of this sort are typically decaying geometrically between generations and geometrically between distinct branches of each family tree.

There are cell systems other than lymphocytes whose behavior has been investigated using branching processes where predictions have been compared to data. One good example is oligodendricyte population dynamics that motivated the models introduced by Boucher et al. [39][40]. There, as in the general Yakovlev-Yanev framework [41] that also includes immigration, the usual branching hypotheses are adopted. For oligodendricytes, using time-lapse photography to follow sibling cells these hypotheses were checked by Hyrien et al. [42]. They found significant correlation between sibling cells' lifetimes. They investigated the impact of this correlation by comparing the predictions of the Boucher et al. models with *in silico* simulation of the system with dependencies. They found this extra correlation structure had little impact on model predictions, as we did for lymphocytes in our Variant 1 model. They comment that they are not aware of mathematical models that explore the impact of these dependencies, which is what we do in this article for lymphocytes.

As the standard branching process hypotheses do not hold for lymphocytes, it is important to study the mathematical and modeling implications of the new hypotheses, as we have done through the introduction of three variants of the Cyton Model in this paper. The mathematical conclusion of studying variants of the Cyton Model that incorporate dependencies between siblings is that the evolution of the mean population size is identical in the variants of the Cyton Model and in the
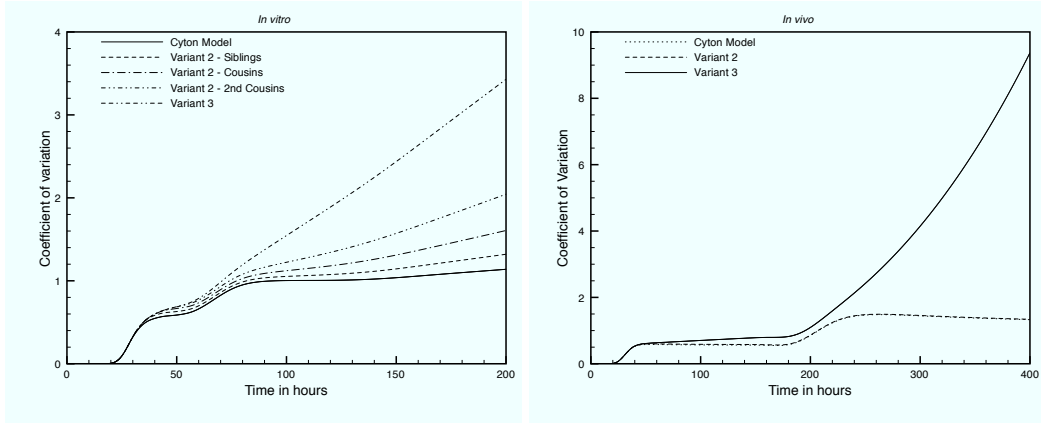
Fig. 6 Coefficient of variation comparison for *in vitro* and *in vivo* data. For the *in vivo* parameterization, the Variant 2 model line overlays the Cyton Model. For both parameterizations, the Variant 1 model line overlays the Cyton Model and so is not shown

original Cyton Model. However, the variability in population sizes in each model differs and can be substantially larger in the presence of correlation between collaterally consanguineous cells.

The qualitative differences seen between the model with extreme correlations, Variant 3, and the others (original Cyton Model, Variant 1 and Variant 2) can most readily be seen by considering the Coefficient of Variation (CV) of the population size distribution. The CV is a normalized measure of the dispersion of a distribution. With d independent founding cells, the CV is $\sqrt{\mathrm{var}(t)}/(m(t)\sqrt{d})$. For the *in vitro* and *in vivo* parameterizations and $d = 1$, these are reported in Figure 6. For the *in vitro* parameterization, two additional lines are drawn based on Variant 2, but with independence beyond siblings and independence beyond cousins. This enables one to see the impact of local dependencies when proliferation progresses for a relatively small number of divisions. The initial rise in CV occurs at the onset of proliferation (approximately 25 hours *in vitro* and 20 hours *in vivo*), with all models being qualitatively similar at the point. The second increase occurs at the onset of division destiny (approximately 60 hours *in vitro* and 180 hours *in vivo*), but the Variant 3 model behaves qualitatively differently to the Cyton Model. Its CV continues to increase at a rapid rate leading to predictions of significantly greater variability at long time-scales. For the *in vitro* parameterization, this rise is also captured by the Variant 2 model, with variance increasing as the depth of dependence is increased. However, for the *in vivo* parameterization, while the Variant 3 model still exhibits the same effect, the Variant 2 model (based on second cousins) is indistinguishable from the Cyton Model. This occurs due to the statistical averaging that arises once a large number of generations are present during proliferation.

These differences can be further explored by considering the predicted variances per generation. In order to determine the moments of the kth generation one first determines $\bar{M}(\boldsymbol{\theta}_0^{K_{\max}}, t)$ (equations (6), (11) and (17)), then takes partial derivatives of it with respect to θ_k and evaluates the resultant function at $\theta_0^{K_{\max}} = (0, 0, \ldots, 0)$. Alternatively, a recursive scheme for any particular moment can be determined by performing the above procedure on the model's iterative scheme.

Figure 7 reports the variance of the population size in for selected generations as a function of time for the Cyton Model, Variant 2 and Variant 3 for the *in vivo* parameterization in Section 6. Note that vertical scale changes on each graph, so that although they look qualitatively similar, they are
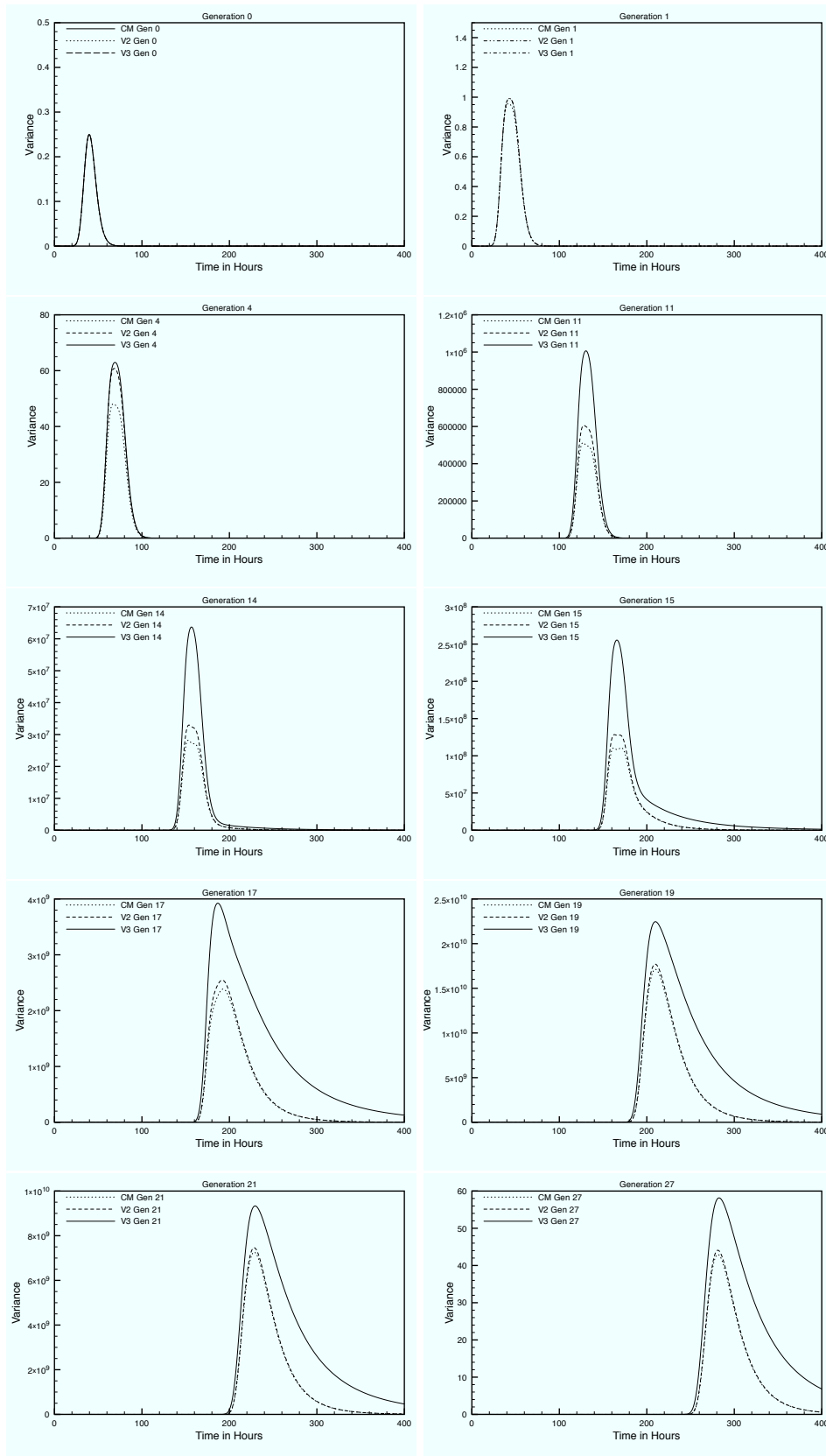
Fig. 7 Variance of population size as a function of time for selected generations between 0th and 27th for the *in vivo* parameterization

quantitatively different. One can see that for the 0th generation, all the predictions are near identical. For the following three generations, the Variant 2 and Variant 3 models predict similar higher variance than the Cyton Model, as illustrated by generations 1 and 4. However, beyond generation 4 the Variant 2 and Variant 3 models separate, as the Variant 2 model's correlation structure only extends to second cousins. By the 11th generation, the Variant 3 model variance is significantly higher than that of the Variant 2. This is also illustrated by the plot for generation 14. By the 19th generation the Variant 2 model variance is nearly identical to that of the Cyton Model, but the Variant 3 model predicts significantly higher variance throughout. Note also that the variance predictions are tightly centered at individual times until the onset of division destiny, whereupon a tail appears.

In summary, by considering the model predictions for model parameterizations based on *in vitro* [7] and *in vivo* [16] experiments, we have investigated the impact of correlation between collaterally consanguineous cells. We deduce that in order for more of the variability seen in experimental data to be explained by the inherent nature of the cell dynamics, rather than experimental error, it would be necessary that the dependencies within each cell's family tree be non-local and extend beyond siblings and cousins. Moreover, the impact of local collaterally consanguineous dependencies is qualitatively similar to that of independence, whereas the impact of non-local collaterally consanguineous dependencies is qualitatively different at longer time-scales.

References

1. A. Bruce Lyons and Christopher R. Parish. Determination of lymphocyte division by flow cytometry. *Journal of Immunological Methods*, 1(1):131–137, 1996.
2. Christopher R. Parish. Fluorescent dyes for lymphocyte migration and proliferation studies. *Immunology & Cell Biology*, 77(6):499–508, 1999.
3. Amanda V. Gett and Philip D. Hodgkin. A cellular calculus for signal integration by T cells. *Nature Immunology*, 1(4):239–244, 2000.
4. Stuart G. Tangye, Danielle T. Avery, Elissa K. Deenick, and Philip D. Hodgkin. Intrinsic differences in the proliferation of naive and memory human B cells as a mechanism for enhanced secondary immune responses. *Journal of Immunology*, 170(2):686–694, 2003.
5. Elissa K. Deenick, Amanda V. Gett, and Philip D. Hodgkin. Stochastic model of T cell proliferation: A calculus revealing IL-2 regulation of precursor frequencies, cell cycle time, and survival. *Journal of Immunology*, 170(10):4963–4972, 2003.
6. Edwin D. Hawkins, Marian L. Turner, Mark R. Dowling, C. van Gend, and Philip D. Hodgkin. A model of immune regulation as a consequence of randomized lymphocyte division and death times. *Proceedings of the National Academy of Sciences of the United States of America*, 104:5032–5037, 2007.
7. Marian L. Turner, Edwin D. Hawkins, and Philip D. Hodgkin. Quantitative regulation of B cell division destiny by signal strength. *Journal of Immunology*, 181(1):374–382, 2008.
8. Vijay G. Subramanian, Ken R. Duffy, Marian L. Turner, and Philip D. Hodgkin. Determining the expected variability of immune responses using the Cyton Model. *Journal of Mathematical Biology*, 56(6):861–892, 2008.
9. E. D. Hawkins, J. F. Markham, L. P. McGuinness, and P. D. Hodgkin. Divide or die: A single cell pedigree analysis of alternative stochastic fates. Submitted for publication, 2008.
10. P. Billingsley. *Probability and Measure.* John Willey & Sons, 1995.
11. Joseph Abate and Ward Whitt. The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems Theory Appl.*, 10(1-2):5–87, 1992.
12. Joseph Abate and Ward Whitt. Numerical inversion of probability generating functions. *Oper. Res. Lett.*, 12(4):245–251, 1992.
13. Gagan L. Choudhury, David M. Lucantoni, and Ward Whitt. Multidimensional transform inversion with applications to the transient $M/G/1$ queue. *Ann. Appl. Probab.*, 4(3):719–740, 1994.

14. E. P. C. Kao. *An Introduction to Stochastic Processes*. Duxbury Press, NY, 1997.
15. Joseph Abate, Gagan L. Choudhury, and Ward Whitt. *Computational Probability*, chapter An Introduction to Numerical Transform Inversion and its Application to Probability Models, W. Grassman ed., pages 257–323. Kluwer, Boston, 1999.
16. D. Homann, L. Teyton, and M.B.A. Oldstone. Differential regulation of antiviral T-cell immunity results in stable CD8+ but declining CD4+ T-cell memory. *Nature Medicine*, 7:913–919, 2001.
17. Andrej Yu. Yakovlev and Nikolaj M. Yanev. *Transient processes in cell proliferation kinetics*, volume 82 of *Lecture Notes in Biomathematics*. Springer-Verlag, Berlin, 1989.
18. Theodore E. Harris. *The theory of branching processes*. Dover Phoenix Editions. Dover Publications Inc., Mineola, NY, 2002.
19. K. B. Athreya and P. E. Ney. *Branching processes*. Dover Publications Inc., Mineola, NY, 2004.
20. Marek Kimmel and David E. Axelrod. *Branching processes in biology*, volume 19 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, 2002.
21. David M. Prescott. Regulation of cell reproduction. *Cancer Research*, 28(9):1815–1820, 1968.
22. D. S. Nachtwey and I. L. Cameron. *Methods in Cell Physiology*, volume III, pages 213–257. Academic Press, New York, 1968.
23. J. A. Smith and L. Martin. Do cells cycle? *Proceedings of the National Academy of Sciences of the United States of America*, 70(4):1263–1267, 1973.
24. R.J. De Boer, D. Homann, and A.S. Perelson. Different dynamics of CD4+ and CD8+ T cell responses during and after acute lymphocytic choriomeningitis virus infection. *Journal of Immunology*, 171(8):3928–3935, 2003.
25. E.O. Powell. Some features of the generation times of individual bacteria. *Biometrika*, 42:16–44, 1955.
26. R. G. Staudte, J. Zhang, R. M. Huggins, and R. Cowan. A reexamination of the cell-lineage data of E. O. Powell. *Biometrics*, 52(4):1214–1222, 1996.
27. H.E. Kubitschek. Cell generation times: ancestral and internal controls. In *Proc. Fifth Berkely Symp. Math. Stat. and Prob.*, volume 4, pages 549–572, 1967.
28. Kenny S. Crump and Charles J. Mode. An age-dependent branching process with correlations among sister cells. *J. Appl. Probability*, 6(1):205–210, 1969.
29. Peter Olofsson. Branching processes with local dependencies. *Ann. Appl. Probab.*, 6(1):238–268, 1996.
30. Kenny S. Crump and Charles J. Mode. A general age-dependent branching process. I, II. *J. Math. Anal. Appl. 24 (1968), 494-508; ibid.*, 25:8–17, 1968.
31. Peter Jagers. A general stochastic model for population development. *Skand. Aktuarietidskr.*, pages 84–103, 1969.
32. Patsy Haccou, Peter Jagers, and Vladimir A. Vatutin. *Branching Processes: Variation, Growth and Extinction of Populations*. Cambridge University Press, 2005.
33. Peter Clifford and Aidan Sudbury. The linear cell-size-dependent branching process. *J. Appl. Probability*, 9:687–696, 1972.
34. Aidan Sudbury and Peter Clifford. Some results for general cell-size-dependent branching processes. *J. Appl. Probability*, 10:289–298, 1973.
35. H. J. A. M. Heijmans. On the stable size distribution of populations reproducing by fission into two unequal parts. *Math. Biosci.*, 72(1):19–50, 1984.
36. Wolfgang Alt and John J. Tyson. A stochastic model of cell division (with application to fission yeast). *Math. Biosci.*, 84(2):159–187, 1987.
37. John J. Tyson. Effects of asymmetric division on a stochastic model of the cell division cycle. *Math. Biosci.*, 96(2):165–184, 1989.
38. R. M. Huggins and R. G. Staudte. Variance components models for dependent cell populations. *J. Amer. Statist. Assoc.*, 89:19–29, 1994.
39. Kenneth Boucher, Alexander Zorin, Andrej Y. Yakovlev, Margot Mayer-Proschel, and Mark Noble. A stochastic model of generation of oligodendrocytes in cell culture. *Math. Biosci.*, 159(1):47–78, 1999.
40. Kenneth Boucher, Alexander Zorin, Andrej Y. Yakovlev, Margot Mayer-Proschel, and Mark Noble. An alternative stochastic model of generation of oligodendrocytes in cell culture. *J. Math. Biol.*, 43(1):22–36, 2001.
41. Andrei Yakovlev and Nikolai Yanev. Branching stochastic processes with immigration in analysis of renewing cell populations. *Math. Biosci.*, 203(1):37–63, 2006.
42. Ollivier Hyrien, Ibro Ambeskovic, Margot Mayer-Proschel, Mark Noble, and Andrei Yakovlev. Stochastic modeling of oligodendrocyte generation in cell culture: model validation with time-lapse data. *Theoretical Biology and Medical Modelling*, 3(2):37–63, 2006.