scientific reports

OPEN

Counting generations in birth and death processes with competing Erlang and exponential waiting times

Giulia Belluccini[1], Martín López-García[1], Grant Lythe[1] & Carmen Molina-París[1,2]✉

Lymphocyte populations, stimulated in vitro or in vivo, grow as cells divide. Stochastic models are appropriate because some cells undergo multiple rounds of division, some die, and others of the same type in the same conditions do not divide at all. If individual cells behave independently, then each cell can be imagined as sampling from a probability density of times to division and death. The exponential density is the most mathematically and computationally convenient choice. It has the advantage of satisfying the memoryless property, consistent with a Markov process, but it overestimates the probability of short division times. With the aim of preserving the advantages of a Markovian framework while improving the representation of experimentally-observed division times, we consider a multi-stage model of cellular division and death. We use Erlang-distributed (or, more generally, phase-type distributed) times to division, and exponentially distributed times to death. We classify cells into generations, using the rule that the daughters of cells in generation n are in generation $n + 1$. In some circumstances, our representation is equivalent to established models of lymphocyte dynamics. We find the growth rate of the cell population by calculating the proportions of cells by stage and generation. The exponent describing the late-time cell population growth, and the criterion for extinction of the population, differs from what would be expected if N steps with rate λ were equivalent to a single step of rate λ/N. We link with a published experimental dataset, where cell counts were reported after T cells were transferred to lymphopenic mice, using Approximate Bayesian Computation. In the comparison, the death rate is assumed to be proportional to the generation and the Erlang time to division for generation 0 is allowed to differ from that of subsequent generations. The multi-stage representation is preferred to a simple exponential in posterior distributions, and the mean time to first division is estimated to be longer than the mean time to subsequent divisions.

Cells of the immune system patrol our bodies for months or years[1,2]. During an adaptive immune response, a subset of specific cells, initially a small fraction of the total population, expands as cells undergo multiple rounds of division over a few days[3]. Although most of these cells die as the infection is overcome, lasting immunity is ensured by the transformation, or "differentiation" of individual cells to a memory phenotype. The most convenient mathematical and computational models of the dynamics of cell populations, which can consider heterogeneity at the single-cell level, are Markov models. In these models, the variables describe the number of cells of each type as a function of time, and cellular events such as division, death or differentiation are defined by their associated rates; each event corresponds to a possible fate of an individual cell and cells are independent of each other. In this formulation, inter-event times are exponentially-distributed random variables, with probability density maximised at zero.

The rapid expansion of cohorts of lymphocytes is recreated in laboratories, either by stimulation in vitro or by transferring cells to lymphopenic mice. By labelling cells with carboxyfluorescein succinimidyl ester (CFSE) or cell trace violet at the beginning of an experiment, and then using flow cytometry at a later time, a cohort of cells can be classified into generations because each round of division dilutes the intracellular dye[4–9]. At the beginning of an experiment, when cells are labelled, it is considered that all cells are in generation 0. In time-lapse microscopy experiments, individual cells are tracked and correlations within family trees identified[10–17]. Over the timescales of such experiments, hours to days, it is not appropriate to treat cell division as an instantaneous

[1]School of Mathematics, University of Leeds, Leeds LS2 9JT, UK. [2]T-6, Theoretical Biology and Biophysics, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA. ✉email: molina-paris@lanl.gov

1

event. Rather, cells are "cycling" through gap, synthesis and mitosis phases (G$_1$/G$_2$, S and M), and daughter cells cannot immediately redivide[18]. To improve on the exponential distribution, Smith and Martin proposed a model in which the time between divisions is the sum of a fixed time spent in phase B, corresponding to S/G$_2$/M, and a variable time spent in phase A, corresponding to G$_1$[19]. In the "single stochastic division" model of Hogan et al., the rate of transition from A to B phase depends on the T cell clonotype and on the number of cells competing for the same resources[20]. If there is a common molecular mechanism controlling the time spent in all phases of the cell cycle, then phase B may, instead, occupy a fixed proportion of the total time[10]. Takahashi[21,22] divided the cell cycle into four phases, with the duration of each drawn from a Pearson type III distribution. Weber et al. postulated a delayed exponential waiting time for each of three phases, corresponding to G$_1$, S and G$_2$/M[23].

The cyton model is a general framework for modelling proliferating lymphocytes, based on the idea that each cell has a set of competing clocks, determining its fate[13]. A number of features are incorporated in the model: generation-dependent parameters, heritable factors, and correlations between cells of the same generation[11,16,24]. When the fate of an individual cell is determined by competing internal clocks, the probability density of observed times between divisions is not the same as that of the division clock because division only happens if another fate does not. The probability density of division times is said to be "censored"[25]. When all clocks have exponential probability densities, the probability density of observed division times is also exponential due to the memory-less property of exponential random variables; this conservation of shape does not hold for non-exponential distributions, including the log-normal and Erlang distributions.

Kendall[26] introduced the idea of cell division occurring at the end of a sequence of k phases, with an exponentially-distributed time spent in each phase, where each phase identifies a cell generation, to study methodological and computational issues related to CFSE labelling data. Gamma and Erlang distributions have been considered to model a cell's time to division[28,29]. Yates et al.[30] incorporated the idea of a sequence of exponentially-distributed phases, also referred as *stages*, before division in a multi-stage representation of a population of dividing cells. This yields an Erlang distribution of times to division, while retaining some of the mathematical and computational advantages of the exponential distribution[30]. The number of stages and their mean duration can be used as free parameters to compare with experimental data[31,32]. On the other hand, the internal stages are a mathematical construct that do not directly correspond to biological phases.

In this manuscript we adopt a multi-stage representation of the cell cycle and we include cell death as a competing fate. In the simplest case, obtained by assuming identical birth and death rates across stages, we derive an analytical expression for the expected number of cells in each stage as a function of time, and study the limiting behaviour of the system as $t \to +\infty$. We also consider cell generations, in order to make theoretical predictions comparable to CFSE experimental data[20]. The expected number of cells in each generation is found in the case of constant birth and death rates, and number of stages across generations. Although arising from different motivations, our multi-stage model with cell generations can be accommodated within the cyton framework, with the progressor fraction equal to one, division and death clocks following Erlang and exponential distributions, respectively. We show the applicability of this stochastic approach by calibrating our multi-stage model with cell generations and its exponential version with CFSE data from two populations of murine T cells[20]. Model calibration is performed making use of Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) approaches[33].

In the "Methods: multi-stage models of cell division and death with a Markovian framework" section, we introduce the multi-stage model including cell death and obtain analytical results for the multi-stage framework under some conditions on the model parameters. In the "Case study: lymphopenia-induced proliferation" section, we calibrate the multi-stage model with cell generations by making use of CFSE data from Hogan et al.[20]. Its performance is compared with a simple exponential model of cell division. A final discussion is provided in the "Discussion" section.

Methods: multi-stage models of cell division and death with a Markovian framework

We present a multi-stage (MS) model of the time between cell divisions. Cells pass through a sequence of N stages before dividing. The stages are not directly related to the biological phases of the cellular cycle. The time to progress from stage j to the next one, $j + 1$, is an exponentially-distributed random variable with mean $1/\lambda^{(j)}$. We will refer to these rates, $\lambda^{(j)}$, $j = 1, \ldots, N$, as *birth* rates. Times to death are also distributed exponentially, with per cell *death* rate μ. Thus, at each stage, each cell may either proceed to the next one, with probability $\lambda^{(j)}/(\lambda^{(j)} + \mu)$, or die, with probability $\mu/(\lambda^{(j)} + \mu)$. The time increment is a random variable, following the exponential distribution with mean $1/(\lambda^{(j)} + \mu)$.

Figure 1 illustrates the dynamics. Our multi-stage model is equivalent to considering two independent clocks for cell division and death, which compete to decide the cellular fate. The time-to-death clock follows an exponential distribution with rate μ, while the division time follows a continuous phase-type distribution with parameters $\boldsymbol{\tau}$ and $\mathbf{T}$[34]. A particular choice of phase-type distribution is the *Erlang*(λ, N), which is a concatenation of N identically distributed exponential steps, where all birth rates are equal: $\lambda^{(j)} = \lambda$, $j = 1, \ldots, N$. The case $\mu = 0$ has been considered by Yates et al.[30].

The number of cells in stage j at time t, is the random variable $S_j(t)$, $j = 1, \ldots, N$. Let $M_j(t) = \mathbb{E}[S_j(t)]$, be the expected value of $S_j(t)$. The following set of differential equations may be obtained by considering the events that can happen in a short time interval:

$$\frac{\mathrm{d}M_j(t)}{\mathrm{d}t} = \begin{cases} 2\lambda^{(N)}M_N(t) - (\lambda^{(1)} + \mu)M_1(t), & \text{if } j = 1, \\ \lambda^{(j-1)}M_{j-1}(t) - (\lambda^{(j)} + \mu)M_j(t), & \text{if } j = 2, \ldots, N. \end{cases} \tag{2.1}$$
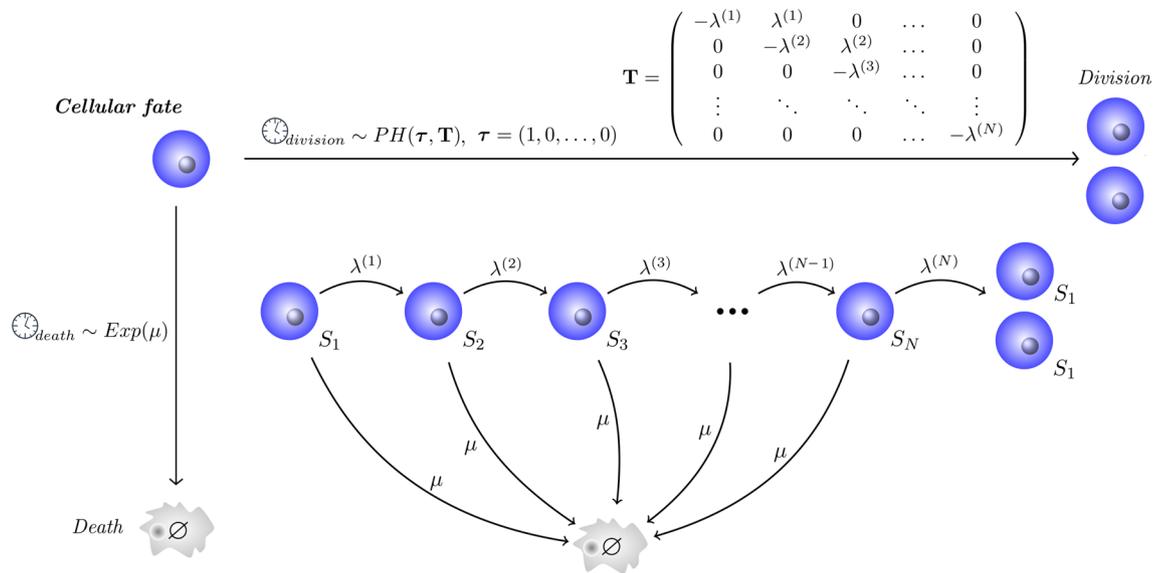
Figure 1. Multi-stage model of cell division and death (MS model). The cell cycle is divided into N different stages. A cell has to visit N stages in order to divide. At each stage j, $j = 1, \ldots, N$, the cell may proceed to the next stage, with birth rate $\lambda^{(j)}$, or die, with death rate μ.
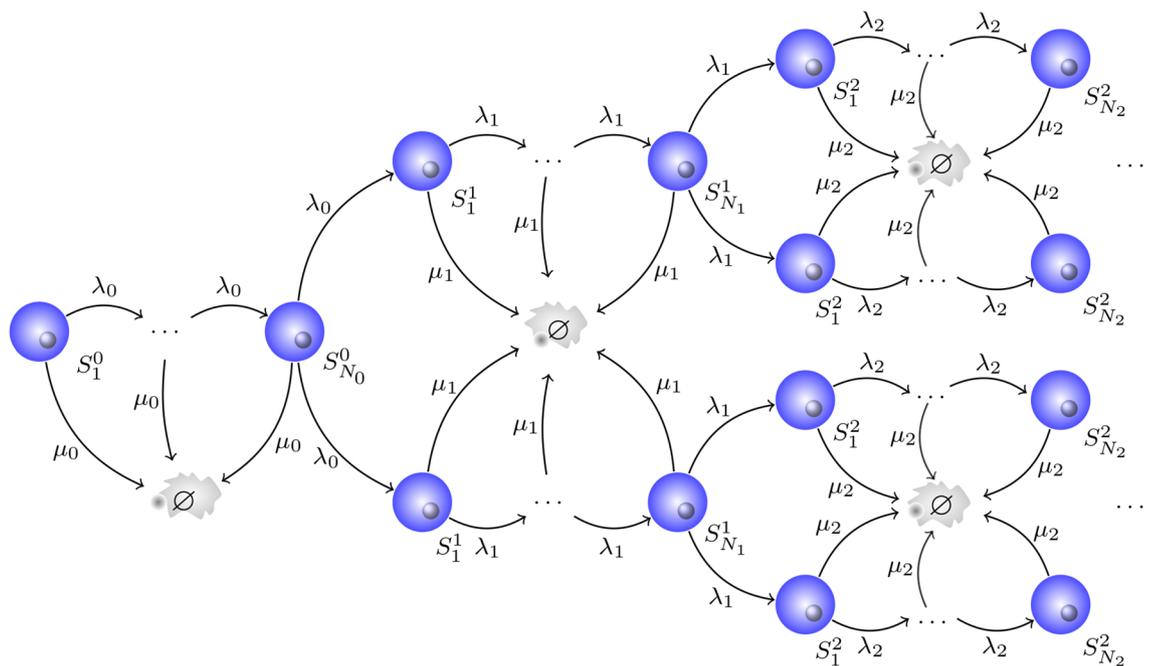
Figure 2. Multi-stage model with cell generations (MS-G model). Each cell in the first stage of generation 0 has to visit all the N_0 compartments (or stages) in order to divide. When cells arrive at the last stage of generation 0, N_0, they may divide with birth rate λ_0, or die with death rate μ_0. If a cell divides, its daughter cells join the first compartment of the next generation, and the process continues.

When we extend the MS model to assign a generation to each cell, we refer to the model as the MS-G model. In this way, mean quantities can be compared with CFSE experimental data[8]. Histograms of CFSE intensity display a series of peaks, each corresponding to a *generation*, or a number of divisions over the course of the experiment[35].

In the MS-G model, generation $g \geq 0$ is split into N_g different stages. The notation N_g reflects the fact that the number of stages may depend on the generation g. A cell in generation g has to sequentially visit all N_g compartments to divide. On the other hand, cells might also die at any stage of the cycle. As depicted in Fig. 2, if a cell belongs to generation g and is in compartment j, $j = 1, \ldots, N_g - 1$, it may proceed to the following stage, with birth rate λ_g, or die with death rate μ_g. Again the notation reflects the potential for these rates to depend on the

generation. When a cell reaches the last stage, N_g, of generation g and divides, its two daughters will join the first compartment of generation $g + 1$. In summary, given a cell in generation g, its time to division follows an Erlang distribution with parameters (λ_g, N_g), whereas its time to death follows an exponential distribution with rate μ_g. These distributions correspond to two independent competing clocks to control cellular fate, similarly to those considered in Fig. 1.

The number of cells in stage j of generation g at time t is the random variable $S_j^g(t)$, $g \geq 0$, $j = 1, \ldots, N_g$. Let $M_j^g(t) = \mathbb{E}[S_j^g(t)]$ be the expected value of $S_j^g(t)$. The following set of differential equations may be obtained by considering the events that can happen in a short time interval:

$$\frac{dM_j^g(t)}{dt} = \begin{cases} -(\lambda_0 + \mu_0)M_1^0(t), & \text{if } g = 0, \ j = 1, \\ \lambda_g M_{j-1}^g(t) - (\lambda_g + \mu_g)M_j^g(t), & \text{if } g \geq 0, \ j = 2, \ldots, N_g, \\ 2\lambda_{g-1} M_{N_{g-1}}^{g-1}(t) - (\lambda_g + \mu_g)M_1^g(t), & \text{if } g \geq 1, \ j = 1. \end{cases} \tag{2.2}$$

We are interested in computing the mean number of cells over time for the MS and MS-G models. Specifically the MS-G model will provide the mean number of cells in each generation, and thus, can be used together with CFSE data to obtain division and death rates. When division times are Erlang distributed (MS model), or if one considers that those Erlang distributions are identical across generations (MS-G model), it is possible to carry out a comprehensive analytical study. This is shown in the "Analytical results" section.

When convenient analytical solutions cannot be obtained, (2.1)–(2.2) can be solved numerically in different ways. For example, for the MS-G model, and keeping in mind our interest in modelling CFSE data, we assume there exists a maximum generation G that can be measured by the dye. Thus, one might be interested in following cells within generations $g = 0, \ldots, G$. For these generations, Eq. (2.2) can be solved by making use of the matrix exponential. To this end, let $\boldsymbol{M}(t)$ be the column vector of the mean number of cells in each stage and generation as time evolves, i.e.,

$$\boldsymbol{M}(t) = \left(M_1^0(t), \ldots, M_{N_0}^0(t), M_1^1(t), \ldots, M_{N_1}^1(t), \ldots, M_1^G(t), \ldots, M_{N_G}^G(t)\right)^T$$
$$= (\mathbf{M}_0(t)^T, \mathbf{M}_1(t)^T, \ldots, \mathbf{M}_G(t)^T)^T,$$

which has length $\sum_{g=0}^G N_g$, and where the column sub-vectors $\mathbf{M}_g(t)$ contain the mean number of cells across stages in generations $g = 0, \ldots, G$. Let us also define the coefficient matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{00} & \mathbf{0}_{N_0 \times N_1} & \mathbf{0}_{N_0 \times N_2} & \cdots & \mathbf{0}_{N_0 \times N_{G-1}} & \mathbf{0}_{N_0 \times N_G} \\ \mathbf{A}_{10} & \mathbf{A}_{11} & \mathbf{0}_{N_1 \times N_2} & \cdots & \mathbf{0}_{N_1 \times N_{G-1}} & \mathbf{0}_{N_1 \times N_G} \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \mathbf{0}_{N_{G-1} \times N_0} & \mathbf{0}_{N_{G-1} \times N_1} & \mathbf{0}_{N_{G-1} \times N_2} & \cdots & \mathbf{A}_{G-1,G-1} & \mathbf{0}_{N_{G-1} \times N_G} \\ \mathbf{0}_{N_G \times N_0} & \mathbf{0}_{N_G \times N_1} & \mathbf{0}_{N_G \times N_2} & \cdots & \mathbf{A}_{G,G-1} & \mathbf{A}_{G,G} \end{pmatrix},$$

where

$$\mathbf{A}_{gg} = \begin{pmatrix} -(\lambda_g + \mu_g) & 0 & 0 & 0 & \cdots & 0 \\ \lambda_g & -(\lambda_g + \mu_g) & 0 & 0 & \cdots & 0 \\ 0 & \lambda_g & -(\lambda_g + \mu_g) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_g & -(\lambda_g + \mu_g) & 0 \\ 0 & \cdots & 0 & 0 & \lambda_g & -(\lambda_g + \mu_g) \end{pmatrix},$$

$$\mathbf{A}_{g,g-1} = \begin{pmatrix} 0 & \cdots & 0 & 2\lambda_{g-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix}.$$

$\mathbf{A}_{gg}$ is a square $N_g \times N_g$ matrix, whereas $\mathbf{A}_{g,g-1}$ is a $N_g \times N_{g-1}$ matrix. $\boldsymbol{A}$ is then a real square matrix of dimension $\sum_{g=0}^G N_g$, and $\mathbf{0}_{a \times b}$ represents a null matrix with dimension $a \times b$. Given the vector of the initial conditions $\boldsymbol{n_0}$, which has length $\sum_{g=0}^G N_g$, the system of Eq. (2.2) can be rewritten as the following Cauchy problem

$$\begin{cases} \dfrac{d\boldsymbol{M}(t)}{dt} = \boldsymbol{A} \cdot \boldsymbol{M}(t), \\ \boldsymbol{M}(0) = \boldsymbol{n_0}. \end{cases}$$

The solution of the system is given by $\boldsymbol{M}(t) = e^{\boldsymbol{A}t}\boldsymbol{n_0}$, where

$$e^{\boldsymbol{A}t} = \sum_{k=0}^{+\infty} \frac{(\boldsymbol{A}t)^k}{k!}$$

represents the matrix exponential. For efficient ways of computing this matrix, see Refs.[36–39]. Finally, we note that since CFSE data describe the number of cells in each generation, one can then compute the mean number of cells in each generation over time as

$$M^g(t) = \sum_{j=1}^{N_g} M_j^g(t), \qquad g \geq 0. \tag{2.3}$$

Using (2.2) and (2.3), we compute the time evolution of $M^g(t)$ as follows

$$\frac{\mathrm{d}M^0(t)}{\mathrm{d}t} = -\lambda_0 M_{N_0}^0(t) - \mu_0 M^0(t),$$
$$\frac{\mathrm{d}M^g(t)}{\mathrm{d}t} = 2\lambda_{g-1} M_{N_{g-1}}^{g-1}(t) - \lambda_g M_{N_g}^g(t) - \mu_g M^g(t), \qquad g \geq 1. \tag{2.4}$$

In a related approach by Zilman et al.[29], a cell's time to division is a gamma-distributed random variable, and time to death is exponentially distributed. Solutions are given in terms of integral equations. Here, with Erlang-distributed division times, we find a set of linear differential equations for the expected number of cells in each stage.

Analytical results

In this Section, we show how the Markovian framework of the proposed multi-stage models provides analytical tractability under some simplifying assumptions. Our aim is to compute the mean number of cells in each stage and generation over time, especially the limiting behaviour as $t \to +\infty$.

MS model with Erlang division time.   In this Section, we consider a simple case of the MS model, where identical birth rates are assumed across different stages; that is, $\lambda^{(j)} = \lambda$, $j = 1, \ldots, N$. The phase-type distribution for the time to division in Fig. 1 is $Erlang(\lambda, N)$ and the mean time to division is given by $\frac{N}{\lambda}$. Note that when $N = 1$ the MS model becomes a Markov linear birth-and-death process, with birth rate, λ, and death rate, μ. Equation (2.1) becomes

$$\frac{\mathrm{d}M_j(t)}{\mathrm{d}t} = \begin{cases} 2\lambda M_N(t) - (\lambda + \mu)M_1(t), & \text{if } j = 1, \\ \lambda M_{j-1}(t) - (\lambda + \mu)M_j(t), & \text{if } j = 2, \ldots, N. \end{cases} \tag{3.1}$$

As in Yates et al.[30], we introduce the new variables $m_j(t) = \mathrm{e}^{(\lambda+\mu)t} M_j(t)$, $j = 1, \ldots, N$, which satisfy the following ODEs:

$$\frac{\mathrm{d}m_j(t)}{\mathrm{d}t} = \begin{cases} 2\lambda m_N(t), & \text{if } j = 1, \\ \lambda m_{j-1}(t), & \text{if } j = 2, \ldots, N. \end{cases} \tag{3.2}$$

We find an Nth-order homogeneous differential equation for $m_N(t)$ that does not depend on μ:

$$\frac{\mathrm{d}^N m_N(t)}{\mathrm{d}t^N} = 2\lambda^N m_N(t), \tag{3.3}$$

together with a set of ODEs that relate $m_j(t)$ to the derivatives of $m_N(t)$ with respect to time

$$m_j(t) = \left(\frac{1}{\lambda}\right)^{N-j} \frac{\mathrm{d}^{N-j} m_N(t)}{\mathrm{d}t^{N-j}}, \ \ j = 1, \ldots, N-1. \tag{3.4}$$

The solution is[30]

$$m_j(t) = 2^{1-\frac{j}{N}} \sum_{k=0}^{N-1} c_k \, z^{-kj} \, \mathrm{e}^{2^{\frac{1}{N}} \lambda z^k t},$$

$j = 1, \ldots, N$ and $t \geq 0$, where $z = \mathrm{e}^{\frac{2\pi i}{N}}$ is the first Nth root of unity, and c_k (for $k = 0, \ldots, N-1$) are constants which depend on the initial conditions.

If, at time $t = 0$, there are C_0 cells in the first stage and zero cells in any other stage, we have

$$c_k = C_0 \frac{2^{\frac{1}{N}}}{2N} \, z^k, \ \ k = 0, \ldots, N-1.$$

Then the solutions of the system (3.1) in terms of the original variables are

$$M_j(t) = C_0 \frac{2^{\frac{1-j}{N}}}{N} \, \mathrm{e}^{-\mu t} \sum_{k=0}^{N-1} z^{(1-j)k} \mathrm{e}^{\left(2^{\frac{1}{N}} z^k - 1\right)\lambda t}, \qquad j = 1, \ldots, N. \tag{3.5}$$

Therefore, the expected total number of cells in the population at time t, $M(t)$, is given by

$$M(t) = \sum_{j=1}^{N} M_j(t) = C_0 \frac{2^{\frac{1}{N}}}{2N} \mathrm{e}^{-\mu t} \sum_{k=0}^{N-1} \frac{z^k}{2^{\frac{1}{N}} z^k - 1} \mathrm{e}^{\left(2^{\frac{1}{N}} z^k - 1\right)\lambda t}. \tag{3.6}$$
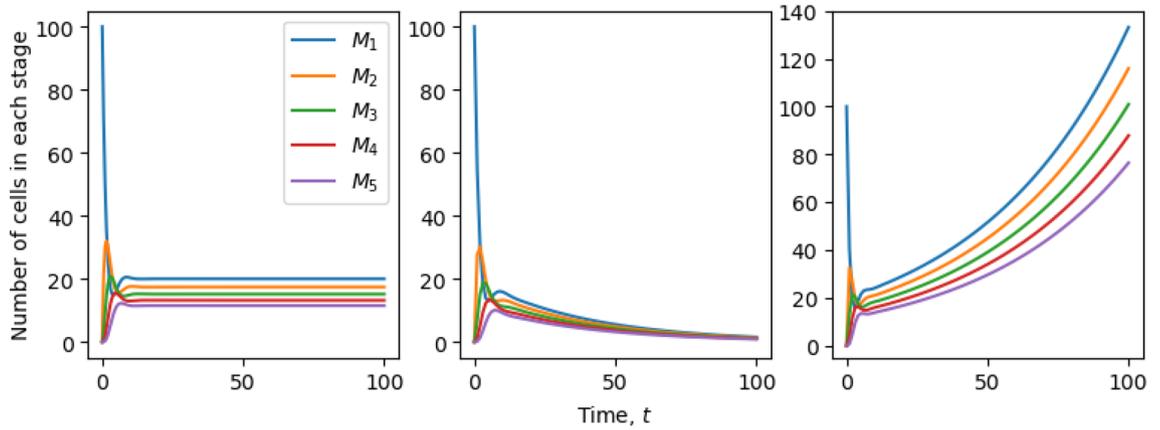
Figure 3. Limiting behaviour when $t \to +\infty$ of a population with an initial number of cells, $C_0 = 10^2$. Birth and death rates, λ and μ, have units of inverse time, t^{-1}. Left: Parameters: $N = 5$, $\lambda = 0.6$, $\mu = (2^{1/N} - 1)\lambda$. The population of cells in stage j levels out to $2^{\frac{1-j}{N}} C_0/N$ for sufficiently large times. Centre: Parameters: $N = 5$, $\lambda = 0.5$, $\mu = 0.1$. The population of cells at any stage becomes extinct at late times. Right: Parameters: $N = 5$, $\lambda = 0.8$, $\mu = 0.1$. The populations grow according to (3.11) and the relation between M_1 and M_5 given by Eq. (3.9) is satisfied. For example, at $t = 100$, $M_1(t) \simeq 2^{4/5} M_5(t)$.

Population extinction or growth. Let us consider the MS model with Erlang division time, and define $S(t) = \sum_{i=1}^{N} S_i(t)$, the total number of cells in the population at time $t \geq 0$, so that $M(t) = \mathbb{E}[S(t)]$. One can define

$$p_j^{(m)} = \lim_{t \to +\infty} \mathbb{P}(S(t) = 0 \mid S_j(0) = m, \ S_k(0) = 0, \ k \neq j), \quad j = 1, \ldots, N,$$

the probability of population extinction at late times, for m cells initially in stage j at time $t = 0$. Since cells behave independently, it is clear that $p_j^{(m)} = (p_j^{(1)})^m$, so we omit the superindex from now on and denote $p_j^{(1)} \equiv p_j$. A first-step argument leads to

$$p_i = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} p_{i+1}, \quad i = 1, \ldots, N-1,$$

$$p_N = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} p_1^2.$$

These equations can be solved recursively, leading to

$$p_1 = \begin{cases} \frac{1}{(1-r)^N} - 1, & \text{if } \mu < (2^{1/N} - 1)\lambda, \\ 1, & \text{if } \mu \geq (2^{1/N} - 1)\lambda, \end{cases} \tag{3.7}$$

where $r = \frac{\mu}{\mu + \lambda}$. When $N = 1$, $p_1 = \frac{\mu}{\lambda}$; when $N = 2$, $p_1 = \frac{\mu^2 + 2\mu\lambda}{\lambda^2}$.

The analytical solutions (3.5)–(3.6) provide another route to study the limiting behaviour as $t \to +\infty$. The largest term in the summation of (3.5) is the one corresponding to $k = 0$. The combination $(2^{1/N} - 1)\lambda - \mu$, is positive if $\mu < (2^{1/N} - 1)\lambda$, negative, when $\mu > (2^{1/N} - 1)\lambda$, or zero if $\mu = (2^{1/N} - 1)\lambda$. Thus, extinction of the cell population is certain if $\mu > (2^{1/N} - 1)\lambda$. Figure 3 (centre) shows an example of extinction when $N = 5$, $\lambda = 0.5$, $\mu = 0.1$ and the initial number of cells $C_0 = 10^2$. An example of behaviour when $\mu = (2^{1/N} - 1)\lambda$ is shown in the left panel of Fig. 3.

If $\mu < (2^{1/N} - 1)\lambda$ then, as $t \to +\infty$,

$$M_j(t) \simeq C_0 \frac{2^{\frac{1-j}{N}}}{N} e^{-\mu t} e^{(2^{\frac{1}{N}} - 1)\lambda t}, \qquad j = 1, \ldots, N, \tag{3.8}$$

and

$$\lim_{t \to +\infty} \frac{M_1(t)}{M_N(t)} = 2^{\frac{N-1}{N}}, \tag{3.9}$$

which is illustrated in Fig. 3 (right). The total population size is easily obtained using

$$\frac{dM(t)}{dt} = \lambda M_N(t) - \mu M(t). \tag{3.10}$$

As $t \to +\infty$,
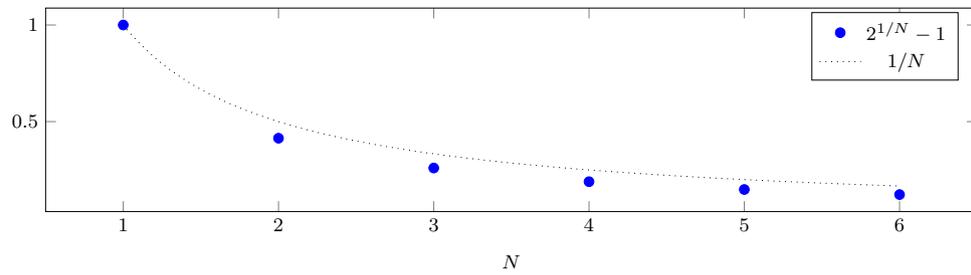
Figure 4. The exponent σ_N that determines the asymptotic growth rate of the population is shown against the number of stages. The dotted line would be expected if N steps with rate λ were equivalent to a single step of rate λ/N.

$$M(t) \simeq C_0 \, \frac{2^{\frac{1}{N}}}{2N\left(2^{\frac{1}{N}} - 1\right)} \, e^{(\sigma_N\lambda-\mu)t}, \qquad \text{where } \sigma_N = (2^{\frac{1}{N}} - 1)\lambda. \tag{3.11}$$

The exponent $\sigma_N = 2^{\frac{1}{N}} - 1$, describing the late-time cell population growth, plotted in Fig. 4, is lower than would be expected if N steps with rate λ were equivalent to a single step of rate λ/N. As $N \to +\infty$, we have $N\sigma_N \to \log 2$. In terms of (3.10), $\lambda\sigma_N < \lambda/N$ since $M_N(t) < 1/N$ as $t \to +\infty$. Because the cell population is unevenly distributed across stages, with a bias towards earlier stages in the long run, N steps with rate λ are not equivalent to a single step of rate λ/N.

Mean fraction of cells at each stage. We define the mean fraction of cells in each stage, $P_j(t)$, as the ratio between the mean number of cells in compartment j and the expected total number of cells in the population, i.e.,

$$P_j(t) = \frac{M_j(t)}{M(t)}, \ \ j = 1, \ldots, N. \tag{3.12}$$

We make use of (3.1) and (3.10) to write,

$$\frac{dP_j(t)}{dt} = \begin{cases} \lambda(2P_N(t) - P_1(t) - P_1(t)P_N(t)), & \text{if } j = 1, \\ \lambda(P_{j-1}(t) - P_j(t) - P_j(t)P_N(t)), & \text{if } j = 2, \ldots, N, \end{cases} \tag{3.13}$$

which have the following steady state solution

$$P_1^* = \frac{2P_N^*}{1 + P_N^*}, \quad P_j^* = \frac{P_{j-1}^*}{1 + P_N^*}, j = 2, \ldots, N. \tag{3.14}$$

One observes that $P_j^* < P_{j-1}^*$, $j = 1, \ldots, N-1$, which means (on average) the fraction of cells decreases stage by stage, independently of the initial distribution of cells. In fact, one can solve (3.14) to determine P_j^*, as follows

$$P_j^* = \left(\sqrt[N]{2}\right)^{N-j}\left(\sqrt[N]{2} - 1\right), j = 1, \ldots, N, \tag{3.15}$$

which does not depend on λ or μ. Thus, at late times the fraction of cells in each stage only depends on the number of stages considered; the parameter λ sets the timescale of the dynamical system, and all cells are equally susceptible to death, regardless of the stage they are in.

MS-G model with identical Erlang division times across generations.

The solutions of the system (2.2) can be written in a closed analytical form in particular cases. For example, one may consider a simplified scenario where the number of stages is equal to 1 for all the generations, i.e., $N_g = 1$ for all $g \geq 0$. Then, if we consider that at time $t = 0$, there are C_0 cells in generation 0, so that $\mathbf{n}_0 = (C_0, 0, \ldots, 0)$, this leads to the following solutions:

$$M^0(t) = C_0 e^{-(\lambda_0+\mu_0)t},$$

$$M^g(t) = 2^g \, C_0 \, \left(\prod_{l=0}^{g-1} \lambda_l\right) \sum_{i=0}^{g} e^{-(\lambda_i+\mu_i)t} \prod_{k=0,k\neq i}^{g} \frac{1}{\lambda_k + \mu_k - \lambda_i - \mu_i}, \quad g \geq 1. \tag{3.16}$$

In this case the MS-G model becomes a birth-and-death process tracking cell generations, and becomes identical to that considered in Refs.[27,40–42], where the inter-event times of cell death and division are modelled as exponential random variables, rather than Erlang distributions.

In this Section we consider the case with identical number of stages, N, and rates, λ and μ, for each generation, so that division times are Erlang-distributed in each generation. Under these assumptions, it is possible to obtain an analytical expression for the mean number of cells in each generation. Then Eq. (2.2) becomes

$$\frac{\mathrm{d}M_j^g(t)}{\mathrm{d}t} = \begin{cases} -(\lambda+\mu)M_1^0(t), & \text{if } g=0, j=1, \\ \lambda M_{j-1}^g(t) - (\lambda+\mu)M_j^g(t), & \text{if } g \geq 0, j=2,\ldots,N, \\ 2\lambda M_N^{g-1}(t) - (\lambda+\mu)M_1^g(t), & \text{if } g \geq 1, j=1. \end{cases} \tag{3.17}$$

These equations can be rewritten in terms of the new variables $m_j^g(t) = e^{(\lambda+\mu)t}M_j^g(t)$, for $g \geq 0$, $j=1,\ldots,N$. This is equivalent to multiplying (3.17) by the integrating factor $e^{(\lambda+\mu)t}$. Thus, (3.17) becomes

$$\frac{\mathrm{d}m_j^g(t)}{\mathrm{d}t} = \begin{cases} 0, & \text{if } g=0, j=1, \\ \lambda m_{j-1}^g(t), & \text{if } g \geq 0, j=2,\ldots,N, \\ 2\lambda m_N^{g-1}(t), & \text{if } g \geq 1, j=1. \end{cases} \tag{3.18}$$

To determine the solutions of (3.18), we focus here on the case $M_1^0(0) = m_1^0(0) = C_0$ and all the other compartments are empty at time $t=0$. This implies that $m_1^0(t) = C_0$ for $t \geq 0$, and by solving (3.18) recursively one gets

$$m_j^0(t) = C_0 \lambda^{j-1} \frac{t^{j-1}}{(j-1)!}, \ j=1,\ldots,N.$$

This expression allows one then to determine the mean number of cells in each stage of generation 1,

$$m_j^1(t) = 2C_0 \lambda^{N+j-1} \frac{t^{N+j-1}}{(N+j-1)!}, \ j=1,\ldots,N. \tag{3.19}$$

By recursion of (3.18) the mean number of cells in each compartment j of generation g is given by

$$m_j^g(t) = 2^g C_0 \lambda^{gN+j-1} \frac{t^{gN+j-1}}{(gN+j-1)!} \text{ for } g \geq 0, j=1,\ldots,N.$$

Going back to the original variables, $M_j^g(t)$, the solutions of (3.17) are

$$M_j^g(t) = 2^g C_0 \lambda^{Ng+j-1} \frac{t^{Ng+j-1}}{(Ng+j-1)!} e^{-(\lambda+\mu)t}, \ g \geq 0, j=1,\ldots,N. \tag{3.20}$$

From the previous equations, one can show that

$$\lim_{t \to +\infty} M_j^g(t) = 0, \text{ for all } g \geq 0, j=1,\ldots,N,$$

since cells in each generation and compartment either proceed to the next stage within their generation, divide (proceeding to the next generation), or die.

Once the mean number of cells in each compartment for a given generation is at hand, the expected number of cells in each generation can be determined according to (2.3). We can write

$$M^g(t) = \sum_{j=1}^N M_j^g(t) = 2^g C_0 (\lambda t)^{Ng} e^{-(\lambda+\mu)t} \sum_{j=1}^N \frac{(\lambda t)^{j-1}}{(Ng+j-1)!}, \ g \geq 0. \tag{3.21}$$

This equation is consistent with the results of the exponential model ($N=1$)[27]. On the other hand, if one is interested in the mean number of cells in each compartment, $M_j(t)$ for $j=1,\ldots,N$, regardless of the generation they belong to, this can be computed as follows

$$\begin{aligned} M_j(t) &= \sum_{g=0}^{+\infty} M_j^g(t) = \sum_{g=0}^{+\infty} 2^g C_0 \lambda^{gN+j-1} \frac{e^{-(\lambda+\mu)t} t^{gN+j-1}}{(gN+j-1)!} \\ &= C_0 e^{-(\lambda+\mu)t} 2^{\frac{1-j}{N}} \sum_{g=0}^{+\infty} \frac{\left(2^{\frac{1}{N}}\lambda t\right)^{gN+j-1}}{(gN+j-1)!}, \end{aligned} \tag{3.22}$$

for $j=1,\ldots,N$ and $t \geq 0$. In practice, one could truncate the series above to get an approximation of the mean number of cells in each stage. However, we note that one can use instead the solution provided by (3.5), since the dynamics of the MS-G model is equivalent to the dynamics of the MS model, when the parameters N, λ and μ are generation-independent. It can be numerically checked that this indeed provides equivalent results. In fact, when $N=1$ or $N=2$, one can analytically show the equivalence. In the former case ($N=1$), it is enough to recall the power series of the exponential function. In the latter case ($N=2$), we derive from (3.5)

8

$$M_1(t) = \frac{C_0}{2} e^{-(\lambda+\mu)t} \left(e^{\sqrt{2}\lambda t} + e^{-\sqrt{2}\lambda t} \right),$$

$$M_2(t) = \frac{C_0}{2\sqrt{2}} e^{-(\lambda+\mu)t} \left(e^{\sqrt{2}\lambda t} - e^{-\sqrt{2}\lambda t} \right),$$

where we used the fact that $z = e^{\pi i} = -1$. On the other hand, from (3.22) we obtain

$$M_1(t) = C_0 e^{-(\lambda+\mu)t} \cosh\left(\sqrt{2}\lambda t\right) = C_0 e^{-(\lambda+\mu)t} \frac{e^{\sqrt{2}\lambda t} + e^{-\sqrt{2}\lambda t}}{2},$$

$$M_2(t) = C_0 e^{-(\lambda+\mu)t} \sinh\left(\sqrt{2}\lambda t\right) = \frac{C_0}{\sqrt{2}} e^{-(\lambda+\mu)t} \frac{e^{\sqrt{2}\lambda t} - e^{-\sqrt{2}\lambda t}}{2}.$$

This shows that the two models lead to the same expected number of cells in each stage.

Comparison between the MS-G model and the cyton model. The *cyton model* is a stochastic model proposed to describe the population dynamics of B and T lymphocytes[13]. Division and death times are regulated by two independent clocks, and the competition between both clocks determines the fate of the cell. When a cell divides, these clocks, which depend on the number of divisions the cell has undergone, are reset for each daughter cell. However, when analysing an in vitro experiment with this type of cells, there is evidence that not all cells either divide or die. For instance, a portion of them may not respond to the stimulation[43], or may respond without division[44]. This is the reason why a progressor fraction is defined in the cyton model. This progressor fraction represents for a given generation, the fraction of cells that are capable of undergoing further division. Each clock is described by a probability density function, and the parameters that define these probabilities are the free parameters in the model. Right skewed distributions, such as log-normal or gamma, are usually adopted to characterise the two independent clocks that regulate cell division and death. In summary, the cyton model is based on the following assumptions:

- death and division are random events, characterised by a probability density function for the time to divide or die, respectively,
- these processes are independent, and compete to determine the fate of the cell,
- the clocks responsible for these processes are reset when a cell divides,
- only a fraction of the cells in each generation are capable to undergo further divisions, and
- the machineries that regulate cellular fate depend on the cell's generation.

In order to translate these assumptions into mathematical terms, let γ_g be the progressor fraction characterising cells having undergone g divisions, and let $\phi_g(\cdot)$ and $\psi_g(\cdot)$ represent the probability density functions for the time to division and death, respectively, for cells in generation g. The number of cells dividing for the first time, or dying, per unit time at time $t \geq 0$ can be calculated, respectively, as[13]:

$$n_0^{div}(t) = \gamma_0 \, C_0 \left(1 - \int_0^t \psi_0(s)\mathrm{d}s \right) \phi_0(t), \tag{3.23}$$

$$n_0^{die}(t) = C_0 \left(1 - \gamma_0 \int_0^t \phi_0(s)\mathrm{d}s \right) \psi_0(t), \tag{3.24}$$

where C_0 is the initial number of cells in the population. Consequently, the time evolution of the expected number of cells in generation 0, $\widetilde{M}^0(t)$, obeys the differential equation

$$\frac{\mathrm{d}\widetilde{M}^0(t)}{\mathrm{d}t} = -\left[n_0^{div}(t) + n_0^{die}(t) \right]. \tag{3.25}$$

The number of cells in generation g dividing, or dying, per unit time at time t can be computed, respectively, as

$$n_g^{div}(t) = 2\gamma_g \int_0^t n_{g-1}^{div}(s) \left[1 - \int_0^{t-s} \psi_g(l)\mathrm{d}l \right] \phi_g(t-s)\mathrm{d}s, \tag{3.26}$$

$$n_g^{die}(t) = 2 \int_0^t n_{g-1}^{div}(s) \left[1 - \gamma_g \int_0^{t-s} \phi_g(l)\mathrm{d}l \right] \psi_g(t-s)\mathrm{d}s. \tag{3.27}$$

Hence, the dynamics of the average number of cells in each generation, $\widetilde{M}^g(t)$, is governed by the differential equations

$$\frac{\mathrm{d}\widetilde{M}^g(t)}{\mathrm{d}t} = 2n_{g-1}^{div}(t) - n_g^{div}(t) - n_g^{die}(t), \, g \geq 1. \tag{3.28}$$

In the next sections we show how the cyton model is equivalent to our model for particular choices of the probability density functions of the division and death clocks, $\phi_g(\cdot)$ and $\psi_g(\cdot)$, and the progressor faction γ_g.

Exponential time to division and death. We consider here the MS-G model with number of stages across generations equal to one, i.e., $N_g = 1$ for all $g \geq 0$. This means that cells in generation g divide after an exponentially distributed time with rate λ_g, and die with rate μ_g. We note that this is different to a standard Markov birth-and-death process, since rates are generation-dependent. Equation (2.4) become

$$\frac{dM^0(t)}{dt} = -(\lambda_0 + \mu_0)M^0(t),$$
$$\frac{dM^g(t)}{dt} = 2\lambda_{g-1}M^{g-1}(t) - (\lambda_g + \mu_g)M^g(t),\ g \geq 1. \tag{3.29}$$

In this case, our model is equivalent to the cyton model with exponential times for division and death, and progressor fraction $\gamma_g = 1$, $g \geq 0$. One can show this equivalence by proving that $n_g^{div}(t) = \lambda_g M^g(t)$ and $n_g^{die}(t) = \mu_g M^g(t)$, by induction on g. In the cyton model, the assumption of exponential time to division and death implies that $\phi_g(t) = \lambda_g e^{-\lambda_g t}$ and $\psi_g(t) = \mu_g e^{-\mu_g t}$, $g \geq 0$. Therefore, according to (3.23) and (3.24), the number of cells at time t dividing for the first time or dying to exit generation 0 per unit time is given by

$$n_0^{div}(t) = C_0 \lambda_0 e^{-(\lambda_0 + \mu_0)t}, \quad n_0^{die}(t) = C_0 \mu_0 e^{-(\lambda_0 + \mu_0)t}.$$

We know from (3.16) that $M^0(t) = C_0 e^{-(\lambda_0 + \mu_0)t}$. Therefore, we can write $n_0^{div}(t) = \lambda_0 M^0(t)$ and $n_0^{die}(t) = \mu_0 M^0(t)$, which proves the case $g = 0$. We assume $n_g^{div}(t) = \lambda_g M^g(t)$ and $n_g^{die}(t) = \mu_g M^g(t)$ hold for generation g and we need to show they also hold for generation $g + 1$. We make use of (3.16) and (3.26) to write

$$n_{g+1}^{div}(t) = 2 \int_0^t \lambda_g 2^g C_0 \prod_{l=0}^{g-1} \lambda_l \sum_{i=0}^g e^{-(\lambda_i+\mu_i)s} \prod_{k=0,k\neq i}^g \frac{1}{\lambda_k + \mu_k - \lambda_i - \mu_i} \lambda_{g+1} e^{-(\lambda_{g+1}+\mu_{g+1})(t-s)} ds$$

$$= \lambda_{g+1} 2^{g+1} C_0 \prod_{l=0}^g \lambda_l \sum_{i=0}^g e^{-(\lambda_{g+1}+\mu_{g+1})t} \int_0^t \prod_{k=0,k\neq i}^g \frac{e^{(\lambda_{g+1}+\mu_{g+1}-\lambda_g-\mu_g)s}}{\lambda_k + \mu_k - \lambda_i - \mu_i} ds$$

$$= \lambda_{g+1} M^{g+1}(t).$$

For the number of cells in generation $g + 1$ dying, Eq. (3.27), together with (3.16) lead to

$$n_{g+1}^{die}(t) = 2 \int_0^t \lambda_g 2^g C_0 \prod_{l=0}^{g-1} \lambda_l \sum_{i=0}^g e^{-(\lambda_i+\mu_i)s} \prod_{k=0,k\neq i}^g \frac{1}{\lambda_k + \mu_k - \lambda_i - \mu_i} \mu_{g+1} e^{-(\lambda_{g+1}+\mu_{g+1})(t-s)} ds$$

$$= \mu_{g+1} 2^{g+1} C_0 \prod_{l=0}^g \lambda_l \sum_{i=0}^g e^{-(\lambda_{g+1}+\mu_{g+1})t} \int_0^t \prod_{k=0,k\neq i}^g \frac{e^{(\lambda_{g+1}+\mu_{g+1}-\lambda_g-\mu_g)s}}{\lambda_k + \mu_k - \lambda_i - \mu_i} ds$$

$$= \mu_{g+1} M^{g+1}(t),$$

which concludes the proof. With the identities $n_g^{div}(t) = \lambda_g M^g(t)$ and $n_g^{die}(t) = \mu_g M^g(t)$ in (3.25) and (3.28), one can show that $M^g(t)$ and $\widetilde{M}^g(t)$ obey the same differential equations for all $g \geq 0$. Thus, the two models are equivalent.

Erlang time to division and exponential time to death. We now consider the more interesting case where the number of stages in each generation is greater than one, and the cell cycle can be described as a multi-stage process. We focus here on the case where identical number of stages N and birth and death rates, λ and μ, respectively, are considered across generations. Similarly to the previous case, we prove that $n_g^{div}(t) = \lambda M_N^g(t)$ and $n_g^{die}(t) = \mu M^g(t)$ by induction on g. Since a cell's time to division is Erlang distributed and a cell's time to death is exponentially distributed, $\psi_g(t) = \mu e^{-\mu t}$ for all $g \geq 0$ and

$$\phi_g(t) = \frac{\lambda^N t^{N-1} e^{-\lambda t}}{(N-1)!}, g \geq 0,$$

where the progressor fraction is again set to 1 for each generation. Note that in this case the parameters in $\phi_g(\cdot)$ and $\psi_g(\cdot)$ are independent of the generation g, since the number of stages and the birth and death rates are identical for all generations. From (3.23) and (3.24), the number of cells dividing for the first time or dying to exit generation 0 per unit time at time t is

$$n_0^{div}(t) = \frac{C_0 \lambda^N t^{N-1}}{(N-1)!} e^{-(\lambda+\mu)t}, \quad n_0^{die}(t) = C_0 \mu e^{-(\lambda+\mu)t} \sum_{j=0}^{N-1} \frac{(\lambda t)^j}{j!}.$$

The dynamics of the expected number of cells in generation 0 is given by (3.25), as in the previous case. Using (3.20) and (3.21), we observe that

$$M^0(t) = C_0 e^{-(\lambda+\mu)t} \sum_{j=0}^{N-1} \frac{(\lambda t)^j}{j!}, \quad M_N^0(t) = \frac{\lambda^N t^{N-1}}{(N-1)!} e^{-(\lambda+\mu)t}.$$

Therefore, $n_0^{div}(t) = \lambda M_N^0(t)$ and $n_0^{die}(t) = \mu M^0(t)$, which concludes the case $g = 0$. We make use of these identities in (3.25) to obtain

$$\frac{d\widetilde{M}^0(t)}{dt} = -\lambda M_N^0(t) - \mu M^0(t),$$

which is the differential equation derived in (2.4) for $M^0(t)$. Now, let us suppose that the identities $n_g^{div}(t) = \lambda M_N^g(t)$ and $n_g^{die}(t) = \mu M^g(t)$ hold for generation g and we prove them for generation $g + 1$. Using (3.26) and the induction hypothesis, we have

$$
\begin{aligned}
n_{g+1}^{div}(t) &= 2 \int_0^t \lambda 2^g C_0 \frac{(\lambda s)^{Ng+N-1}}{(Ng+N-1)!} e^{-(\lambda+\mu)s} e^{-\mu(t-s)} \frac{\lambda^N (t-s)^{N-1} e^{-\lambda(t-s)}}{(N-1)!} ds \\
&= 2^{g+1} \frac{\lambda^{N(g+2)}}{(N(g+1)-1)!} C_0 e^{-(\lambda+\mu)t} \frac{1}{(N-1)!} \int_0^t s^{N(g+1)-1}(t-s)^{N-1} ds \\
&= 2^{g+1} \frac{\lambda^{N(g+2)}}{(N(g+1)-1)!} C_0 e^{-(\lambda+\mu)t} \sum_{j=0}^{N-1} \frac{(-1)^j t^{N-1-j}}{j!(N-1-j)!} \int_0^t s^{N(g+1)-1+j} ds \\
&= \lambda 2^{g+1} \frac{(\lambda t)^{N(g+1)+N-1}}{(N(g+1)+N-1)!} C_0 e^{-(\lambda+\mu)t} = \lambda M_N^{g+1}(t),
\end{aligned}
$$

where we have used (3.20) for the last step. The same arguments can be used to look at the number of cells in generation $g + 1$ dying per unit of time, (3.27). Together with the induction hypothesis, we can write

$$
\begin{aligned}
n_{g+1}^{die}(t) &= 2 \int_0^t \lambda 2^g C_0 \frac{(\lambda s)^{Ng+N-1}}{(Ng+N-1)!} e^{-(\lambda+\mu)s} e^{-\lambda(t-s)} \sum_{j=0}^{N-1} \frac{\lambda^j (t-s)^j}{j!} \mu e^{-\mu(t-s)} ds \\
&= 2^{g+1} \lambda^{Ng+N} C_0 \frac{e^{-(\lambda+\mu)t} \mu}{(Ng+N-1)!} \sum_{j=0}^{N-1} \frac{\lambda^j}{j!} \int_0^t s^{Ng+N-1}(t-s)^j ds \\
&= 2^{g+1} \lambda^{Ng+N} C_0 e^{-(\lambda+\mu)t} \mu \sum_{j=0}^{N-1} \lambda^j \sum_{k=0}^j \frac{t^j}{k!(j-k)!} \frac{t^{k+N+Ng}}{k+N+Ng} \frac{(-1)^k}{(Ng+N-1)!} \\
&= \mu 2^{g+1} C_0 e^{-(\lambda+\mu)t} \sum_{j=0}^{N-1} \frac{(\lambda t)^{N(g+1)+j}}{(N(g+1)+j)!} = \mu M^{g+1}(t),
\end{aligned}
$$

where the last identity was obtained making use of (3.21). Hence, (3.28) becomes

$$\frac{d\widetilde{M}^g(t)}{dt} = 2\lambda M_N^{g-1}(t) - \lambda M_N^g(t) - \mu M^g(t), \; g \geq 1,$$

which is identical to (2.4) for $M^g(t)$, $g \geq 1$. This concludes the proof of the equivalence between the cyton model and the multi-stage model with generations when a cell's time to divide is Erlang distributed with parameters λ and N, and a cell's time to die is exponential with rate μ. In summary, the analysis presented in this section for the multi-stage model with Erlang division time and exponential death time leads to exact closed solutions for the cyton model with the previous choice of clocks.

Case study: lymphopenia-induced proliferation

In this Section we illustrate the applicability of the MS-G model to CFSE data, making use of an experimental study of lymphopenia-induced proliferation[20]. In particular, we compare the performance of the MS-G model to that of a simple exponential (or single stage) model with generations, which is equivalent to making $N_g = 1$ for all g in the MS-G model.

Differences in T cell proliferation have been observed to vary between different T cell clonotypes (i.e., the set of T cells with the same T cell receptor). Hogan et al.[20] transferred CFSE-labelled OT-I or F5 T cells intravenously to lymphopenic mice. A certain number of days (3, 4, 5, 6, 7, 10, 12 and 18 days) after the transfer, spleens and lymph nodes were recovered from the mice and analysed by flow cytometry to quantify the expression levels of CD8, CD5, CD44, and CFSE dilution[20]. For each time point, the number of mice analysed was between 3 and 7. We note that two independent transfer experiments, carried out under identical conditions, were performed: one for OT-I cells and a second one for F5. In Fig. 5 both data sets are shown: for each time point the number of cells is plotted for each mouse and generation (identified via the CFSE dilution measurement). On the left (right), OT-I (F5) cells are represented by the green (blue) histograms. In order to infer model parameters, we will consider all cells which have divided five or more times as a single class, denoted 5+. This is similar to the approach considered in Refs.[4,29,35]. The rationale behind this choice is to reduce errors in the quantification of labelled cells with low CFSE fluorescence, as is the case for five or more divisions.

Figure 5 clearly shows that OT-I T cells proliferate faster than F5 cells, so that by day 7 there are OT-I cells in generation 10, whereas for F5 cells the maximum generation observed at day 7 is 6. This greater proliferative capacity of OT-I cells eventually leads, after one week, to competition for resources (e.g., IL-7 cytokine) and the
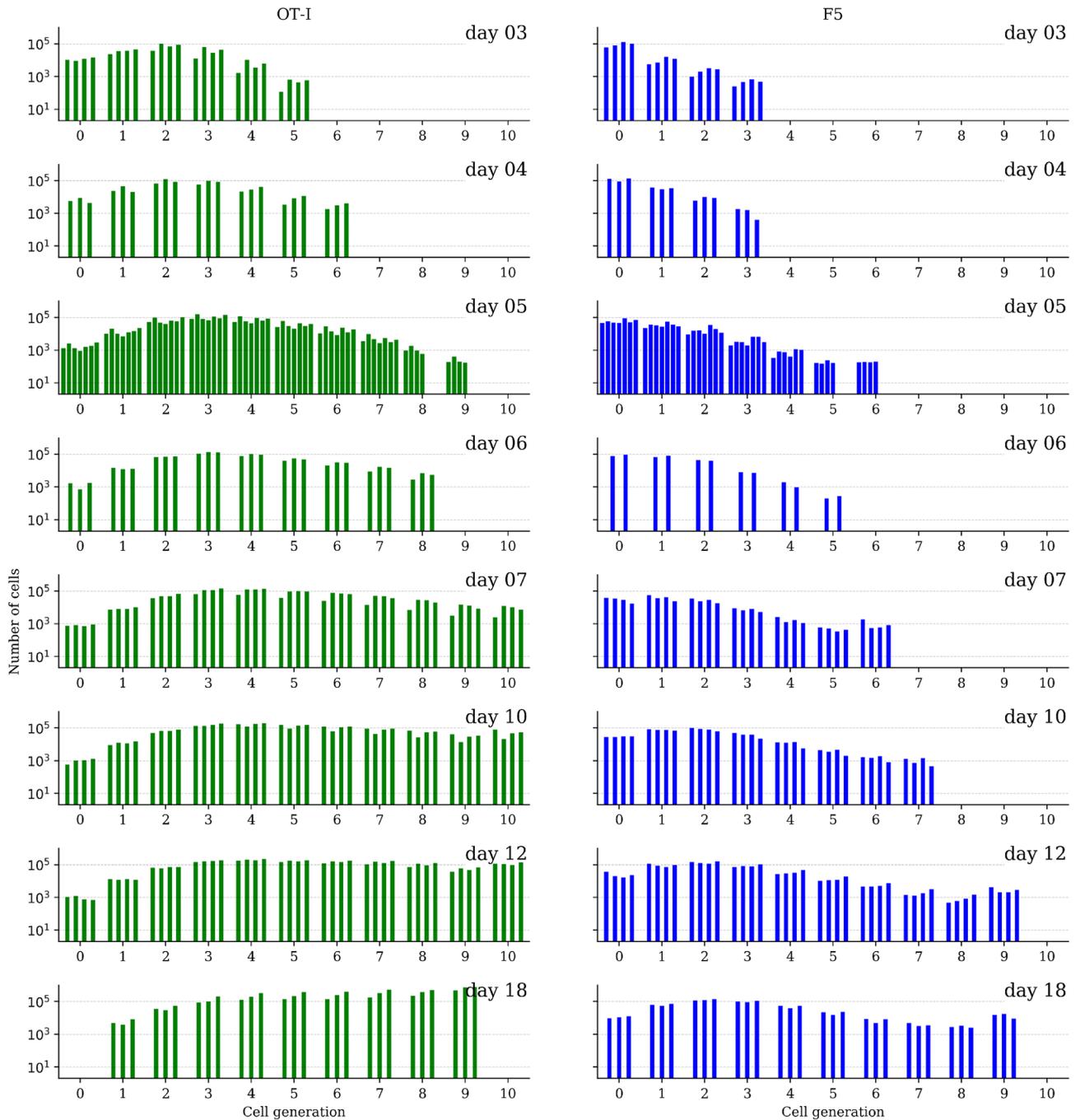
Figure 5. Data set of murine T lymphocytes from Hogan et al.[20]. Left: OT-I T cells. Right: F5 T cells. For each time point, the number of cells is plotted for each mouse and generation.

OT-I population approaching its carrying capacity[20]. Since our model does not account for competition, it can only appropriately describe the dynamics of OT-I cells during the first week of the experiment. Thus, for OT-I cells we will only make use of the data set up to that time (one week). Yet for the F5 population we will use the entire data set. In Hogan et al.[20] this competition was incorporated with a density-dependent birth rate, $\lambda(P)$, as follows

$$\lambda(P) = \bar{\lambda}\, e^{-\delta P}, \tag{4.1}$$

where $\bar{\lambda}$ is the rate of growth under unlimited resources, δ the size of reduction caused by the expansion of competing cells, and P is the size of the population[20]. Figure 6 shows the density-dependent birth rate, $\lambda(P)$, as a function of the population size P. It suggests that the competition for resources is greater in the case of OT-I T cells. In the experiments the number of OT-I cells after one week (about 5×10^5) is larger than the population
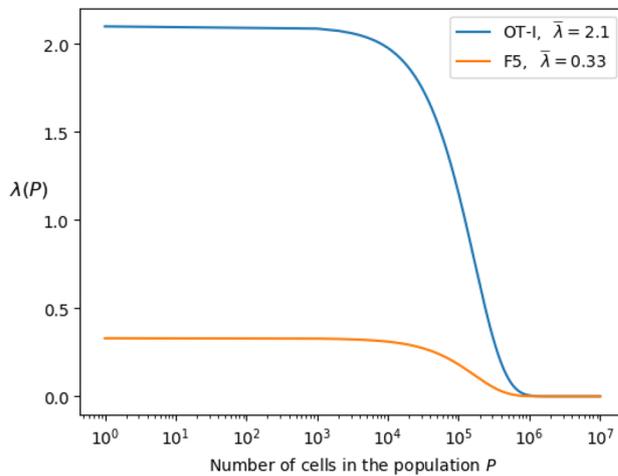
Figure 6. Density-dependent birth rate, $\lambda(P)$, as a function of the population size, P. The parameter $\bar{\lambda}$, with units of $cell \cdot day^{-1}$, represents the rate of growth under no competition and δ quantifies the level of reduction caused by the expansion of competing cells. Values for $\bar{\lambda}$ (shown in the inset) and $\delta = 6.0 \times 10^{-6}$ are taken from[20], Table 1].

Model parameters	Description	Prior distribution
C_0	Initial number of cells	$C_0 = 10^x, x \sim U(4,6)$
N_0, N	Number of stages	$U_{\text{discrete}}(1, 50)$
λ_0, λ	Birth rate	$\lambda_0 = 10^y, \lambda = 10^z, y, z \sim U(-3, 1)$
α	Death rate slope	$\alpha = 10^w, w \sim U(-5, -1)$

Table 1. Prior distributions for model parameters. Units for λ_0, λ and α are inverse hours (h^{-1}).

of F5 T cells at day 18 (about 4×10^5). Therefore, the population of F5 T cells never reaches its carrying capacity and the role of competition for resources can be neglected.

We estimate model parameters with the ABC-SMC algorithm[33]. Thus, the posterior distribution of the parameters is obtained by T sequential applications of the ABC algorithm, where the posterior obtained in each iteration is used as prior for the next one. This algorithm requires the definition of prior distributions for the first iteration, a distance function, a tolerance threshold for each iteration, and a perturbation kernel[33]. We assume all parameters are initially distributed according to a uniform prior distribution, as described in Table 1. When a prior distribution spans several orders of magnitude, the uniform distribution is taken over the exponent to efficiently explore parameter space. Given $x_D^g(t)$, the experimentally determined mean number of cells in generation g at time t, for $g \in \{0, 1, 2, 3, 4, 5+\}$, and its corresponding model prediction, $x_M^g(t) = M^g(t)$ for a particular choice of parameters $\boldsymbol{\theta} = (C_0, N_0, N, \lambda_0, \lambda, \alpha)$, the distance function is defined as

$$d(\text{model,data} \mid \boldsymbol{\theta}) = \sqrt{\sum_{g=0}^{G} \sum_{t \in \mathcal{T}} \left[\frac{x_M^g(t) - x_D^g(t)}{\sigma_D^g(t)} \right]^2}, \tag{4.2}$$

where $\mathcal{T}$ is the set of time points and depends on the clonotype of interest, $\sigma_D^g(t)$ represents the standard deviation of the experimental data at time t and generation g, and G is the merged (and maximum) generation, $G = 5+$. In practice, we define the first tolerance threshold, ε_1, in the ABC-SMC algorithm as the median value of the distances obtained from 10^4 preliminary realisations, with the parameters sampled from the prior distributions in Table 1. The subsequent tolerance thresholds, ε_j, $j = 2, \ldots, T$ can be then defined as the median of the distance values obtained from the previous iterations of the algorithm. Finally, we use a uniform perturbation kernel to perturb the parameters during the sequence of iterations[33], and implement the algorithm for $T = 16$ in the case of the multi-stage model and $T = 7$ for the single stage one.

Before performing the Bayesian inference, we make some assumptions based on the experimental set up. Several studies have shown that the time to a first division is larger the time to subsequent divisions, since cells require time to become activated before they divide[13,15,16]. Thus, we assume that all generations but 0 are comprised of the same number of stages N, whereas generation 0 is characterised by N_0 stages. Similarly, cells in generation 0 proceed to divide with birth rate λ_0, whilst all the other generations have a birth rate λ. Therefore, in contrast to the inference in[29], the number of stages N_0 and N are free parameters in the model. On the other
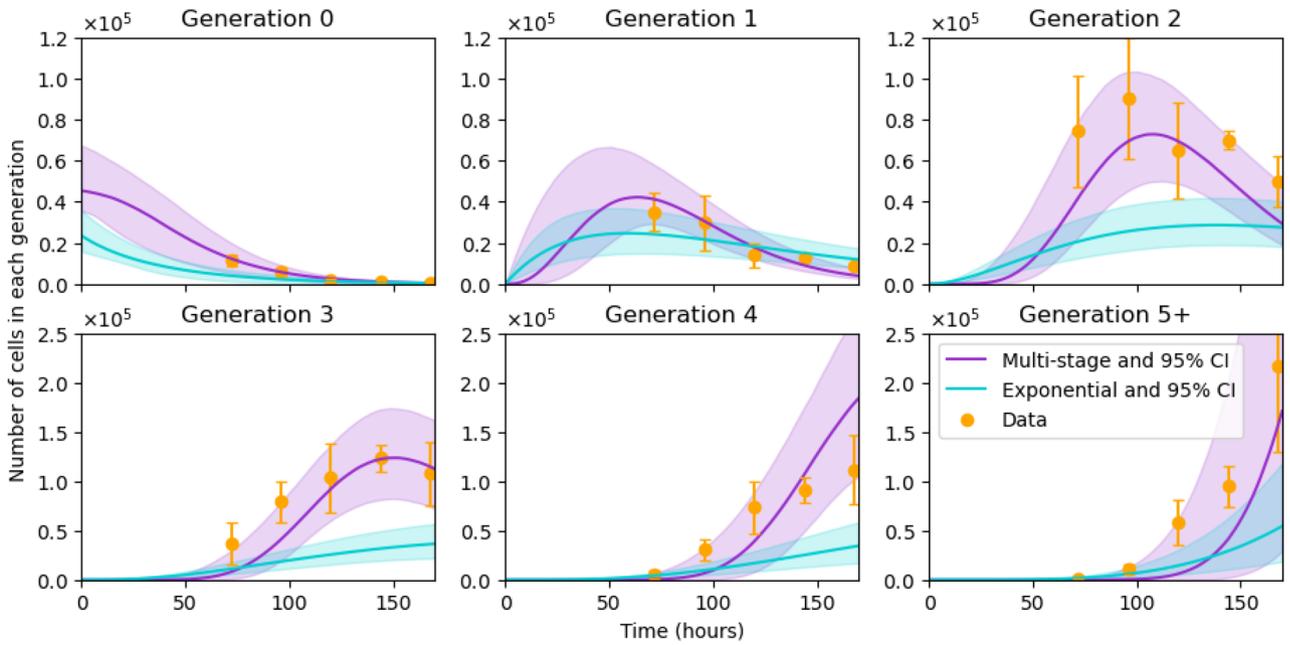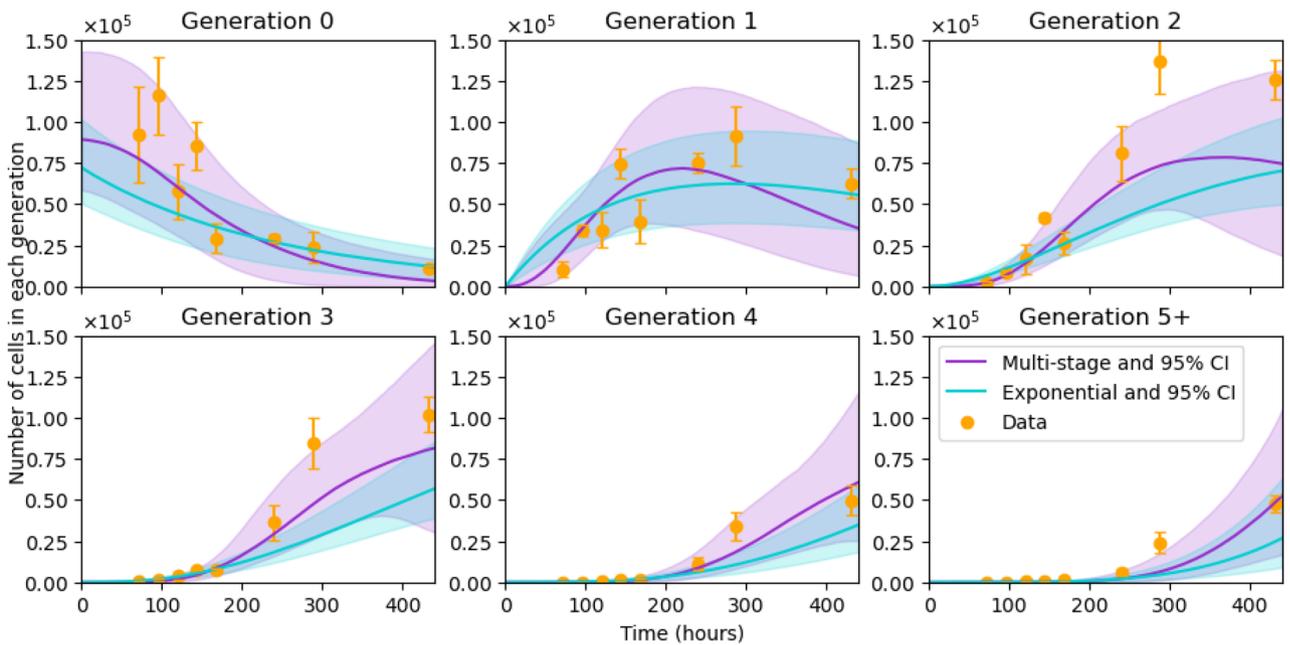
Figure 7. Exponential (solid turquoise line) and multi-stage (solid magenta line) model predictions compared to the data sets (orange dots) for OT-I (**A**) and F5 (**B**) T cells. Bars on data points represent their standard deviation. The expected number of cells in each generation is plotted as a function of time. These predictions represent the median value of 10^4 simulations with the accepted parameter values from the posterior distributions. Shaded areas represent 95% confidence intervals.

hand, we propose that the per cell death rate in a given generation is linear on the number of cell divisions that the cell has undergone[35,45]. We write

$$\mu_g = \alpha\, g,\ g \geq 0, \tag{4.3}$$

where α is a parameter to estimate. These linear death rates encode the fact that cells are more likely to die when they have already undergone several divisions[35,45]. Finally, the initial number of cells, C_0, is considered a parameter to be estimated, since the actual number of transferred cells which make it to the lymph nodes or spleen cannot be measured.

Mathematical model	Cell type	Value of AIC_C
Multi-stage	OT-I T cells	50.4
Exponential	OT-I T cells	283
Multi-stage	F5 T cells	206
Exponential	F5 T cells	317

Table 2. AIC_C values for the exponential and multi-stage models calibrated with CFSE data of murine T lymphocytes.
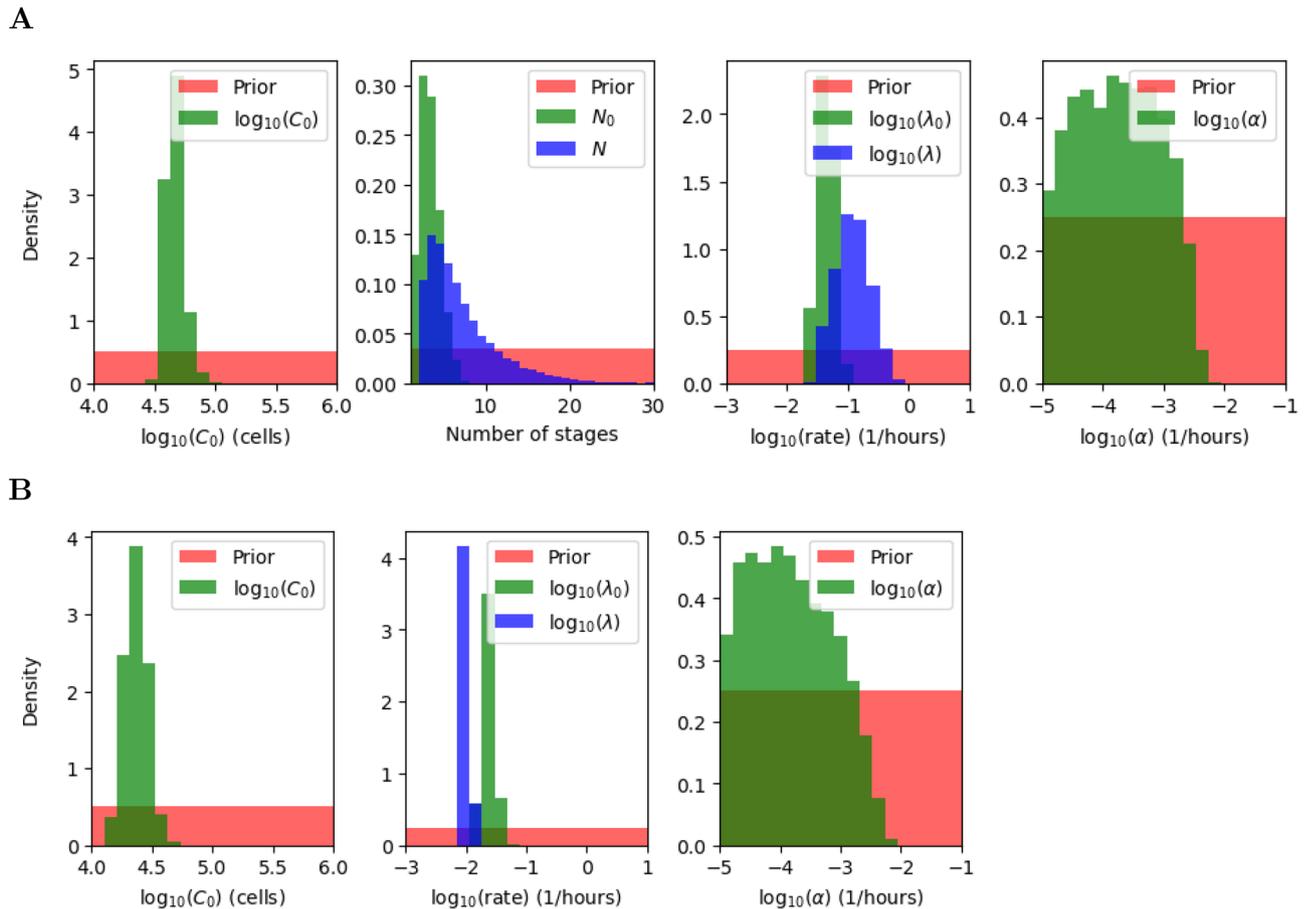
Figure 8. Posterior distributions (green and blue) for the parameters in the multi-stage (**A**) and exponential (**B**) model for OT-I T cells. In the exponential model, the number of stages for all generations is equal to 1, i.e., $N_0 = N = 1$. Prior distributions are shown in red.

The predictions obtained for each model, and for each clonotype (OT-I or F5), are shown in Fig. 7. We run the model with the parameters being sampled from the estimated posterior distributions and compute the median of all the simulations, which corresponds to the solid magenta (multi-stage model) and turquoise (exponential model) lines in Fig. 7. The bands around median predictions represent 95% confidence intervals. Data points are plotted with the standard deviation from the multiple experimental replicates. As shown in Fig. 7, the calibrated MS-G model successfully captures the dynamics of the proliferating T cell populations (OT-I and F5), whereas the single stage model significantly underestimates the expected number of cells beyond generation 1, particularly in the case of OT-I T cells. We made use of the corrected version of the Akaike Information Criterion (AIC_C)[46,47] to quantify the better fit of the MS-G model accounting for the extra parameters N_0 and N in the multi-stage model. The values of AIC_C for each model and clonotype are listed in Table 2. Despite the two extra parameters, the values of AIC_C corresponding to the multi-stage model are significantly lower for both clonotypes. Overall, the MS-G model is able to explain the data from the OT-I transfer experiment better, since this data set is less noisy than the F5 set.

The marginal posterior distributions for each parameter are shown in green and blue in Figs. 8 and 9, for the multi-stage and exponential models, respectively, and the (uniform) prior distributions are plotted in red. Summary statistics of these posterior distributions are shown in Tables 3, 4, 5 and 6. Cell death is governed by

Parameter	Minimum	Maximum	Mean	Median	Standard deviation
C_0	3.00×10^4	1.03×10^5	4.67×10^4	4.54×10^4	8.03×10^3
N_0	1	7	2.83	3	1.23
N	2	34	6.59	5	4.30
λ_0	1.98×10^{-2}	1.08×10^{-1}	4.64×10^{-2}	4.56×10^{-2}	1.45×10^{-2}
λ	2.80×10^{-2}	8.08×10^{-1}	1.48×10^{-1}	1.20×10^{-1}	1.01×10^{-1}
α	1.00×10^{-5}	5.97×10^{-3}	5.06×10^{-4}	1.76×10^{-4}	7.47×10^{-4}

Table 3. Summary statistics of OT-I clonotype posterior distributions for the multi-stage model.
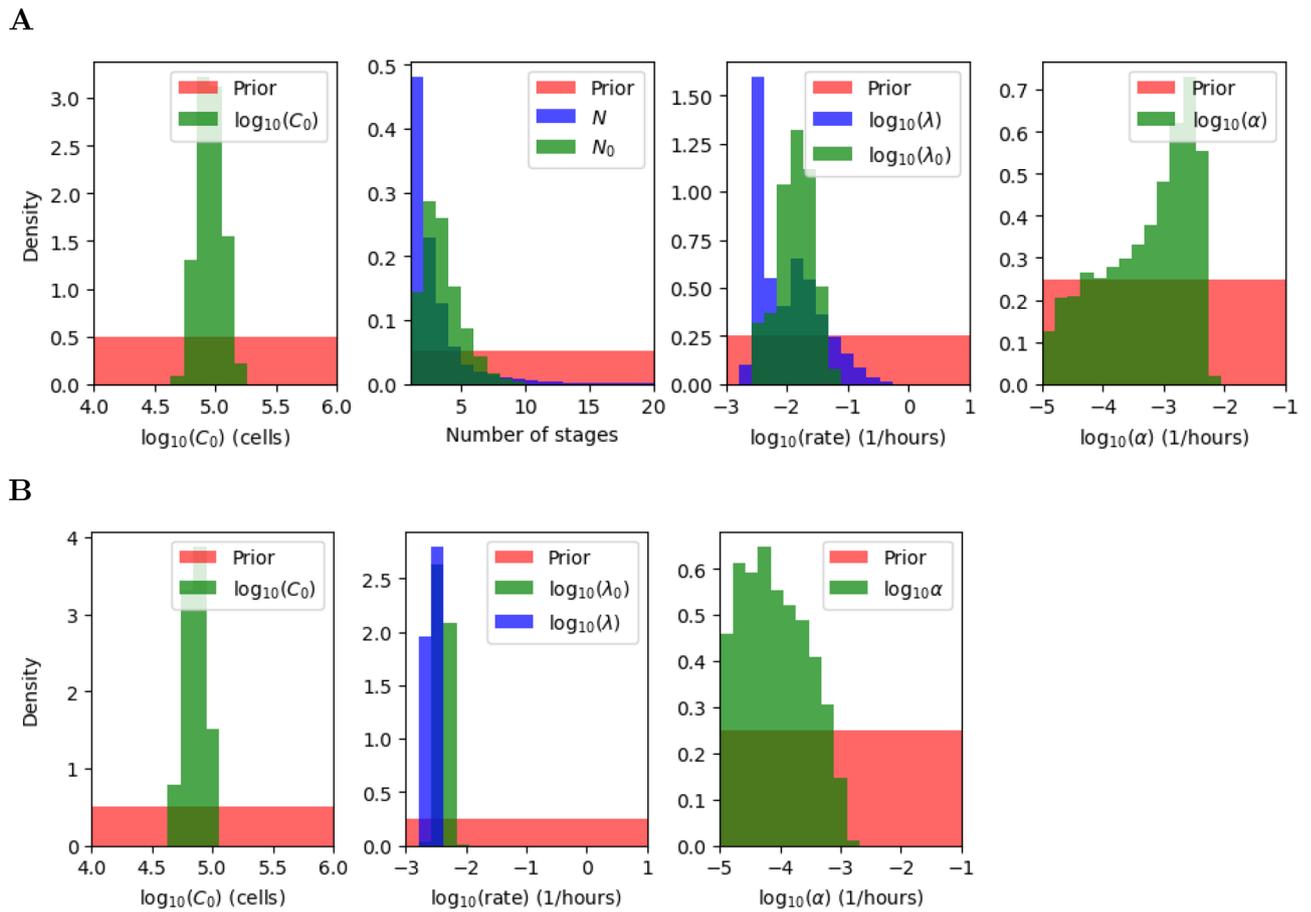
Figure 9. Posterior distributions (green and blue) for the parameters in the multi-stage (**A**) and exponential (**B**) model for F5 T cells. In the exponential model, the number of stages for all generations is equal to 1, i.e., $N_0 = N = 1$. Prior distributions are shown in red.

Parameter	Minimum	Maximum	Mean	Median	Standard deviation
C_0	4.74×10^4	1.85×10^5	9.26×10^4	8.94×10^4	2.22×10^4
N_0	1	10	3.01	3	1.53
N	1	35	2.42	2	2.57
λ_0	2.68×10^{-3}	7.20×10^{-2}	1.70×10^{-2}	1.47×10^{-2}	1.07×10^{-2}
λ	2.06×10^{-3}	5.88×10^{-1}	2.20×10^{-2}	9.54×10^{-3}	3.90×10^{-2}
α	1.00×10^{-5}	6.21×10^{-3}	1.35×10^{-3}	8.19×10^{-4}	1.40×10^{-3}

Table 4. Summary statistics of F5 clonotype posterior distributions for the multi-stage model.
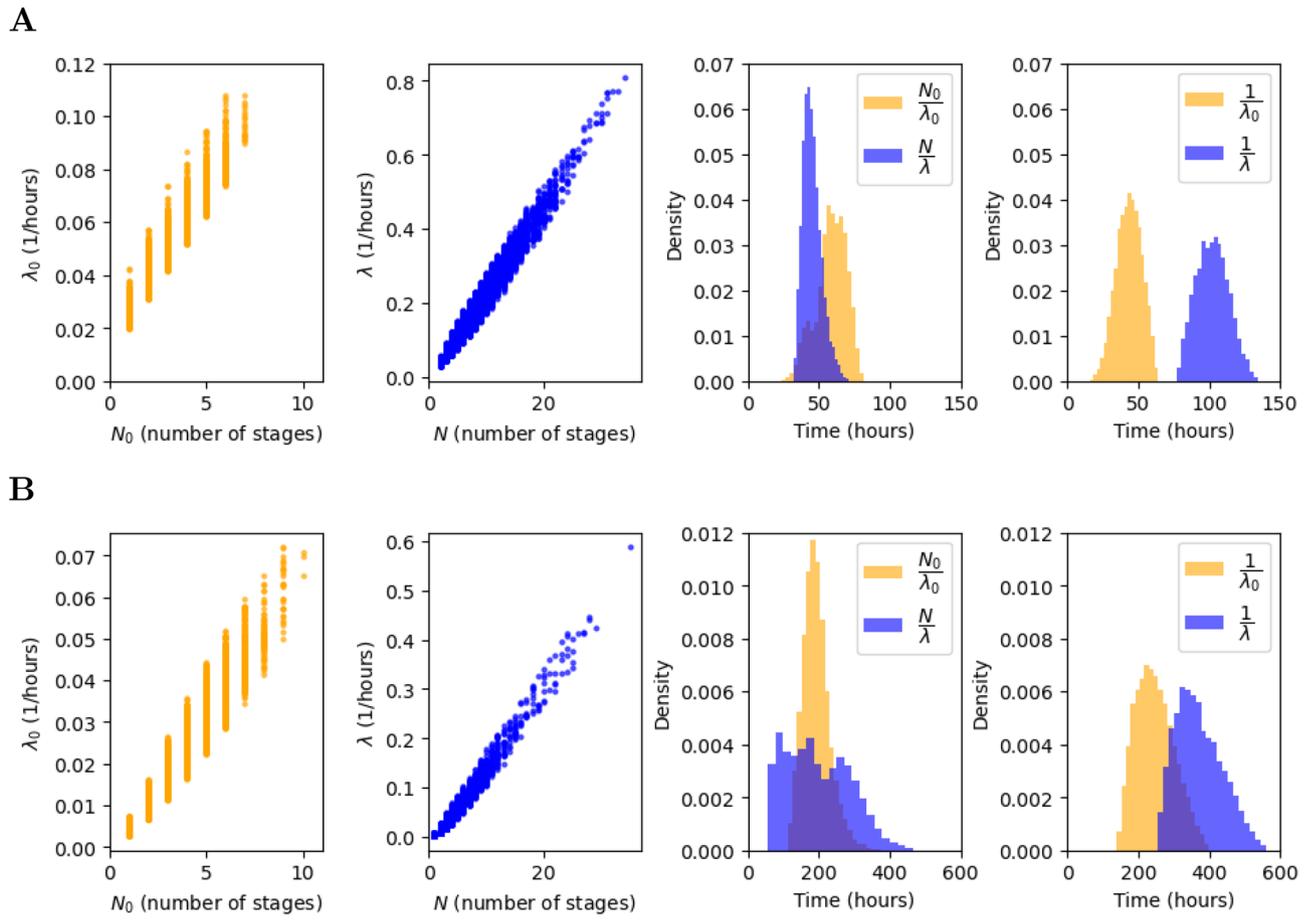
Figure 10. Joint posterior distributions (left two plots) of the number of stages N_0, N and the birth rates λ_0, λ. Marginal posterior distributions (right two plots) for the mean time to first and subsequent divisions estimated from the multi-stage model (third column) and the exponential model (fourth column). Panel **A** for OT-I T cells and **B** for F5 T cells.

Parameter	Minimum	Maximum	Mean	Median	Standard deviation
C_0	1.43×10^4	5.46×10^5	2.40×10^4	2.34×10^4	5.28×10^3
λ_0	1.56×10^{-2}	6.19×10^{-2}	2.40×10^{-2}	2.28×10^{-2}	5.69×10^{-3}
λ	7.45×10^{-3}	1.29×10^{-2}	9.88×10^{-3}	9.78×10^{-3}	1.10×10^{-3}
α	1.00×10^{-5}	7.23×10^{-3}	4.76×10^{-4}	1.32×10^{-4}	8.17×10^{-4}

Table 5. Summary statistics for the posterior distributions of the exponential model for the OT-I clonotype.

Parameter	Minimum	Maximum	Mean	Median	Standard deviation
C_0	4.54×10^3	1.10×10^5	7.33×10^4	7.22×10^4	1.39×10^4
λ_0	2.50×10^{-3}	7.18×10^{-3}	4.24×10^{-3}	4.13×10^{-3}	9.28×10^{-4}
λ	1.79×10^{-3}	3.92×10^{-3}	2.77×10^{-3}	2.76×10^{-3}	4.66×10^{-4}
α	1.00×10^{-5}	1.52×10^{-3}	1.72×10^{-4}	7.30×10^{-5}	2.23×10^{-4}

Table 6. Summary statistics for the posterior distributions of the exponential model for the F5 clonotype.

the parameter α, and is estimated to be low for both models and clonotypes, suggesting that cell death does not have a significant impact on the dynamics during lymphopenia, which is in fact dominated by cell division. This result is in agreement with Hogan et al.[20], where the death rate is assumed to be zero. The initial number of cells can be estimated with relative success, and does not seem to depend heavily on the model considered. On the other hand, cell division is governed by parameters $(N_0, \lambda_0, N, \hat{\lambda})$, with $N_0 = N = 1$ in the exponential model.

We note that in both models, $\frac{N_0}{\lambda_0}$ and $\frac{N}{\lambda}$ represent the mean time to the first and subsequent divisions, respectively. Although all division-related parameters can be estimated from the data, for both models and clonotypes, a correlation between the division rate and the number of stages is seen in the scatter plots of Fig. 10. Instead of plotting the marginal posterior distributions for these parameters, one can consider the posterior distribution for the mean times $\frac{N_0}{\lambda_0}$ and $\frac{N}{\lambda}$ (see Fig. 10). The fact that $N = 1$ is never chosen as an accepted parameter value in the posterior distribution for the multi-stage model and the OT-I clonotype already suggests that a multi-stage representation of cell division is preferred for this clonotype. On the other hand for the F5 clonotype the marginal distribution for N shows a non-zero frequency for the value 1, but larger values of N are also represented in its posterior distribution. The mean time to both first and subsequent divisions, $\frac{N_0}{\lambda_0}$ and $\frac{N}{\lambda}$, are significantly longer for the F5 clonotype than the OT-I. In fact, our results estimate that F5 T cells divide slowly compared to OT-I cells, requiring on average 192 h to carry out a first division (59 h taken by OT-I T cells), as shown in Fig. 10 for the multi-stage model. The time to subsequent divisions is represented by the blue histograms. Interestingly, our estimation of the mean time to first division of OT-I cells, on average 59 h, is close to the value obtained by Hogan et al.[20] (52 h when considering the best fit parameter estimates). In the case of F5 cells, we predict an average of 192 h to undergo their first division, whereas Hogan et al. obtained a value of 137 h. We note that the value 137 h is within the range covered by our predicted posterior distribution.

Our results indicate that OT-I T lymphocytes require on average 59 h for their first division, and a bit less, 46 h, for subsequent divisions (see upper left plot of Fig. 10). Based on our Bayesian approach, we conclude that a multi-stage model with a constant division rate after the first division event, is a suitable description of lymphopenia-induced proliferation[5,29,40]. The MS-G model estimates that F5 cells take on average slightly less than 200 h to divide, both for the first or subsequent division rounds, as shown in the lower left plot of Fig. 10. This difference can be explained by the different characteristics of OT-I and F5 T cells, and was previously observed[20]. The posterior distributions of the expected time to subsequent divisions in the MS-G model, $\frac{N}{\lambda}$, and in the exponential one $\frac{1}{\lambda}$, shown as blue histograms in Fig. 10, indicate that the exponential model predicts a longer division time than the multi-stage model for both clonotypes. This can be justified by the implementation of the ABC-SMC algorithm. Indeed, when parameterising the exponential model, the algorithm tries to keep the distance between the model predictions and the experimental observations low. This leads to the choice of parameter sets which limit cell proliferation, as shorter division times in the exponential model would lead to an increase in cell numbers not observed in the data set, and thus larger distance values. This is why the estimated birth rates in the exponential model are lower than the ones in the multi-stage representation. As a result, the exponential model predicts a greater average division time than the multi-stage model for both clonotypes. Finally, our results indicate that for both clonotypes the exponential model (see Fig. 10) found a shorter time to first division than to subsequent ones, contradicting previous findings[13,15,16], which support longer first division times. This is related to the fact that, overall, the exponential model is not able to capture the observed cell dynamics for neither of the clonotypes, as can be seen in Fig. 7.

Discussion

We analyse a multi-stage model of cell proliferation and death, tracking cell generations, in a framework that retains the benefits of a Markov process. With particular choices of rates, the models are equivalent to others in the literature[13,27,29,40–42]. In the case study of "Case study: lymphopenia-induced proliferation" section, the MS-G model performs better than the exponential model of time to division. The model implemented here provides a flexible framework for estimating the birth and death rates that describe the dynamics of lymphocyte populations[48,49]. The representation retains the advantages of a Markovian approach, including analytical tractability in some cases, and computational efficiency of numerical simulations with the Gillespie algorithm[50,51]. The expected number of cells in each generation satisfy a set of linear differential equations. Further comparison of this and of published models[13,27,29,40–42] with different experimental datasets is the aim of future work.

It has been observed[13,15,16] that immune cells typically need longer to divide for the first time, whereas later divisions require shorter times[5]. It is possible to assume that divided and undivided cells have different probability densities of time to cell division in exponential and Smith-Martin models[7,35,45]. With the multi-stage model introduced here, the separation need not be explicit because it is incorporated in the generation-dependent parameters. A longer mean time to first division, $\frac{N_0}{\lambda_0}$, than mean time to subsequent divisions, $\frac{N}{\lambda}$, is a natural part of the framework. Extension of the mathematical analysis in "Analytical results" section to the case $\lambda_0 \neq \lambda$, $N_0 \neq N$ and possibly generation-dependent death rate μ_g, $g \geq 0$, would be desirable.

Our calculations rely on the assumption that cells are independent of each other. In particular, no fate correlation is assumed between daughter cells and their progenitors, or between siblings. However, data sets from time-lapse microscopy of B and T cell families[10–12,14,16,17,25] show that division and death times for siblings are correlated, and "division destiny" is a familial characteristic[24]. A further potential extension of the MS-G model is the introduction of a population carrying capacity. In the model as described in "Analytical results" section, the mean number of cells over time either increases without bound, dies out or reaches a steady-state, depending on the relation between division-related parameters (birth rate and number of stages in the cell cycle), and the death rate. Competition for resources may be modelled using density-dependent birth and/or death rates[20,52], or by rates that depend on the time-dependent availability of resources[53].

Data availability

Python codes to carry out parameter calibration in the "Case study: lymphopenia-induced proliferation" section are available at https://doi.org/10.5281/zenodo.6337679.

References

1. den Braber, I. *et al.* Maintenance of peripheral naive T cells is sustained by thymus output in mice but not humans. *Immunity* **36**(2), 288–297 (2012).
2. Westera, L. *et al.* Closing the gap between T-cell life span estimates from stable isotope-labeling studies in mice and humans. *Blood* **122**(13), 2205–221 (2013).
3. Antia, R., Ganusov, V. V. & Ahmed, R. The role of models in understanding CD8+ T-cell memory. *Nat. Rev. Immunol.* **5**(2), 101–111 (2005).
4. De Boer, R. J., Ganusov, V. V., Milutinović, D., Hodgkin, P. D. & Perelson, A. S. Estimating lymphocyte division and death rates from CFSE data. *Bull. Math. Biol.* **68**(5), 1011–1031 (2006).
5. Gett, A. V. & Hodgkin, P. D. A cellular calculus for signal integration by T cells. *Nat. Immunol.* **1**(3), 239–244 (2000).
6. Hasbold, J. *et al.* Quantitative analysis of lymphocyte differentiation and proliferation in vitro using carboxyfluorescein diacetate succinimidyl ester. *Immunol. Cell Biol.* **77**(6), 516–522 (1999).
7. Lee, H. Y. & Perelson, A. S. Modeling T cell proliferation and death in vitro based on labeling data: Generalizations of the Smith-Martin cell cycle model. *Bull. Math. Biol.* **70**, 21–44 (2008).
8. Lyons, A. B. & Parish, C. R. Determination of lymphocyte division by flow cytometry. *J. Immunol. Methods* **171**(1), 131–137 (1994).
9. Wellard, C., Markham, J. F., Hawkins, E. D., & Hodgkin, P. D. The Cyton model for lymphocyte proliferation and differentiation. In *Mathematical Models and Immune Cell Biology* 107–120 (Springer, New York, 2011).
10. Dowling, M. R. *et al.* Stretched cell cycle model for proliferating lymphocytes. *Proc. Natl. Acad. Sci.* **111**(17), 6377–6382 (2014).
11. Duffy, K. R. & Subramanian, V. G. On the impact of correlation between collaterally consanguineous cells on lymphocyte population dynamics. *J. Math. Biol.* **59**(2), 255–285 (2009).
12. Duffy, K. R. *et al.* Activation-induced B cell fates are selected by intracellular stochastic competition. *Science* **335**(6066), 338–341 (2012).
13. Hawkins, E. D., Turner, M. L., Dowling, M. R., Van Gend, C. & Hodgkin, P. D. A model of immune regulation as a consequence of randomized lymphocyte division and death times. *Proc. Natl. Acad. Sci. USA* **104**, 5032–7 (2007).
14. Hawkins, E. D., Markham, J. F., McGuinness, L. P. & Hodgkin, P. D. A single-cell pedigree analysis of alternative stochastic lymphocyte fates. *Proc. Natl. Acad. Sci.* **106**(32), 13457–13462 (2009).
15. Kinjyo, I. *et al.* Real-time tracking of cell cycle progression during CD8$^+$ effector and memory T-cell differentiation. *Nat. Commun.* **6**, 6301 (2015).
16. Markham, J. F., Wellard, C. J., Hawkins, E. D., Duffy, K. R. & Hodgkin, P. D. A minimum of two distinct heritable factors are required to explain correlation structures in proliferating lymphocytes. *J. R. Soc. Interface* **7**, 1049–1059 (2010).
17. Wellard, C., Markham, J., Hawkins, E. D. & Hodgkin, P. D. The effect of correlations on the population dynamics of lymphocytes. *J. Theor. Biol.* **264**(2), 443–449 (2010).
18. Pandit, A. & De Boer, R. J. Stochastic inheritance of division and death times determines the size and phenotype of CD8+ T cell families. *Front. Immunol.* **10**, 436 (2019).
19. Smith, J. A. & Martin, L. Do cells cycle?. *Natl. Acad. Sci.* **70**, 1263–1267 (1973).
20. Hogan, T. *et al.* Clonally diverse T cell homeostasis is maintained by a common program of cell-cycle control. *J. Immunol.* **190**, 3985–3993 (2013).
21. Takahashi, M. Theoretical basis for cell cycle analysis I. Labelled mitosis wave method. *J. Theoret. Biol.* **13**, 202–211 (1966).
22. Takahashi, M. Theoretical basis for cell cycle analysis II. Labelled mitosis wave method. *J. Theoret. Biol.* **18**, 195–209 (1968).
23. Weber, T. S., Jaehnert, I., Schichor, C., Or-Guil, M. & Carneiro, J. Quantifying the length and variance of the eukaryotic cell cycle phases by a stochastic model and dual nucleoside pulse labelling. *PLoS Comput. Biol.* **10**(7), e1003616 (2014).
24. Cheon, H. *et al.* Cyton2: A model of immune cell population dynamics that includes familial instructional inheritance. *Front. Bioinform.* **50**, 1 (2021).
25. Duffy, K. R. & Hodgkin, P. D. Intracellular competition for fates in the immune system. *Trends Cell Biol.* **22**(9), 457–464 (2012).
26. Kendall, D. G. On the role of variable generation time in the development of a Stochastic birth process. *Biometrika* **35**, 316–330 (1948).
27. Luzyanina, T. *et al.* Computational analysis of CFSE proliferation assay. *J. Math. Biol.* **54**(1), 57–89 (2007).
28. León, K., Faro, J. & Carneiro, J. A general mathematical framework to model generation structure in a population of asynchronously dividing cells. *J. Theor. Biol.* **229**(4), 455–476 (2004).
29. Zilman, A., Ganusov, V. V. & Perelson, A. S. Stochastic models of lymphocyte proliferation and death. *PLoS ONE* **5**(9), e12775 (2010).
30. Yates, C. A., Ford, M. J. & Mort, L. M. A multi-stage representation of cell proliferation as a Markov process. *Bull. Math. Biol.* **79**, 2905–2928 (2017).
31. Chao, H. X. *et al.* Evidence that the human cell cycle is a series of uncoupled, memoryless phases. *Mol. Syst. Biol.* **15**(3), e8604 (2019).
32. Vittadello, V. T., McCue, S. W., Gunasingh, G., Haass, N. K. & Simpson, M. J. Mathematical models incorporating a multi-stage cell cycle replicate normally-hidden inherent synchronization in cell proliferation. *J. R. Soc. Interface* **16**, 20190382 (2019).
33. Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. H. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202 (2008).
34. He, Q. M. *Fundamentals of Matrix-analytic Methods* Vol. 365 (Springer, New York, 2014).
35. Ganusov, V. V., Milutinovic, D. & De Boer, R. J. IL-2 regulates expansion of CD4$^+$ T cell populations by affecting cell death: Insights from modeling CFSE data. *J. Immunol.* **179**, 950–957 (2007).
36. Gómez-Corral, A. & López-García, M. Maximum population sizes in host-parasitoid models. *Int. J. Biomath.* **6**(02), 1350002 (2013).
37. Gómez-Corral, A. & López-García, M. Maximum queue lengths during a fixed time interval in the M/M/c retrial queue. *Appl. Math. Comput.* **235**, 124–136 (2014).
38. Moler, C. & Van Loan, C. Nineteen dubious ways to compute the exponential of a matrix. *SIAM Rev.* **20**(4), 801–836 (1978).
39. Moler, C. & Van Loan, C. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev.* **45**(1), 3–49 (2003).
40. De Boer, R. J. & Perelson, A. S. Estimating division and death rates from CFSE data. *J. Comput. Appl. Math.* **184**(1), 140–164 (2005).
41. Luzyanina, T. B. & Bocharov, G. A. Markov chain Monte Carlo parameter estimation of the ODE compartmental cell growth model. *Math. Biol. Bioinform.* **13**(2), 376–391 (2018).
42. Revy, P. *et al.* Functional antigen-independent synapses formed between T cells and dendritic cells. *Nat. Immunol.* **2**, 925–931. https://doi.org/10.1038/ni713 (2001).
43. Pereira, J. P., Girard, R., Chaby, R., Cumano, A. & Vieira, P. Monoallelic expression of the murine gene encoding Toll-like receptor 4. *Nat. Immunol.* **4**(5), 464–470 (2003).

44. Deenick, E. K., Gett, A. V. & Hodgkin, P. D. Stochastic model of T cell proliferation: A calculus revealing IL-2 regulation of precursor frequencies, cell cycle time, and survival. *J. Immunol.* **170**(10), 4963–4972 (2003).
45. Mazzocco, P., Bernard, S. & Pujo-Menjouet, L. Estimates and impact of lymphocyte division parameters from CFSE data using mathematical modelling. *PLoS ONE* **12**(6), e0179768 (2017).
46. Anderson, D. & Burnham, K. Model selection and multi-model inference. *Second NY Springer* **63**, 10 (2004).
47. Burnham, K. P. & Anderson, D. R. Multimodel inference: Understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**(2), 261–304 (2004).
48. Callard, R. & Hodgkin, P. Modeling T-and B-cell growth and differentiation. *Immunol. Rev.* **216**(1), 119–129 (2007).
49. De Boer, R. J. & Perelson, A. S. Quantifying T lymphocyte turnover. *J. Theor. Biol.* **327**, 45–87 (2013).
50. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**(4), 403–434 (1976).
51. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**(25), 2340–2361 (1977).
52. Callard, R. E., Stark, J. & Yates, A. J. Fratricide: A mechanism for T memory-cell homeostasis. *Trends Immunol.* **24**(7), 370–375 (2003).
53. Hart, Y. *et al.* Paradoxical signaling by a secreted molecule leads to homeostasis of cell levels. *Cell* **158**(5), 1022–1032 (2014).

Acknowledgements

Author contributions

G.B. carried out model development and numerical work. G.B. prepared the first draft of the manuscript and figures.M.L.G. carried out supervision, model development and theoretical (mathematical and statistical) work.G.L. carried out supervision, model development and theoretical (mathematical and statistical) work.C.M.P. carried out supervision, model development and theoretical (mathematical and statistical) work. All authors wrote and reviewed the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.M.-P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.