# Modelling cell lifespan and proliferation: is likelihood to die or to divide independent of age?

## Mark R. Dowling[1,2], Dejan Milutinović[3] and Philip D. Hodgkin[1,†]

[1] *The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3050, Australia*
[2] *School of Physical Sciences, The University of Queensland, Queensland 4072, Australia*
[3] *Theoretical Biology, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands*

In cell lifespan studies the exponential nature of cell survival curves is often interpreted as showing the rate of death is independent of the age of the cells within the population. Here we present an alternative model where cells that die are replaced and the age and lifespan of the population pool is monitored until a steady state is reached. In our model newly generated individual cells are given a determined lifespan drawn from a number of known distributions including the lognormal, which is frequently found in nature. For lognormal lifespans the analytic steady-state survival curve obtained can be well-fit by a single or double exponential, depending on the mean and standard deviation. Thus, experimental evidence for exponential lifespans of one and/or two populations cannot be taken as definitive evidence for time and age independence of cell survival. A related model for a dividing population in steady state is also developed. We propose that the common adoption of age-independent, constant rates of change in biological modelling may be responsible for significant errors, both of interpretation and of mathematical deduction. We suggest that additional mathematical and experimental methods must be used to resolve the relationship between time and behavioural changes by cells that are predominantly unsynchronized.

**Keywords:** cell lifespan; cell proliferation; lymphocyte half lives; lognormal

## 1. INTRODUCTION

Similar cells vary in their responses, even when placed under identical environmental or stimulatory conditions. This inherent variation presents a challenge to the development of mathematical descriptions of cell behaviour. Perhaps the most common strategy is to treat the variation as an error and take the average behaviour of the population as the 'correct' outcome. A more realistic approach is to attribute cell differences to some internal stochastic behaviour such that a rate of change, for example likelihood to die, is constant over time and therefore the survival curve is a decaying exponential function. While the assumption that a rate of change is constant is taken partly for mathematical convenience, there are many experiments consistent, at least superficially, with the projected exponential. These include measurement of variations in pre-replicative phases of cell cycle times (Smith & Martin 1973), loss of cells from the naïve lymphocyte pool (Fulcher & Basten 1994, 1997; Tough & Sprent 1994; Tanchot & Rocha 1997), loss of plasma cells after an immune response (Slifka *et al.* 1998), as well as cell death *in vitro* (Gett & Hodgkin 2000; Deenick *et al.* 2003).

One area where the assumption of time independence has been common is in studies of lymphocyte lifespan and proliferation. For example loss and replacement of naïve non-dividing lymphocytes in blood or peripheral lymphoid organs can be followed by labelling newly formed, and therefore recently divided, cells with the Thymidine analogue Bromodeoxyuridine (BrDU; Osmond 1991). Curves showing replacement of BrDU negative murine B cells are well fitted by either one or two exponentials (Fulcher & Basten 1994, 1997). Human T lymphocyte half-lives have been estimated by loss of cells with permanent chromosome damage following radiation exposure. These curves are again well fit by one, or two, exponentials (Mclean & Michie 1995; Ramalho *et al.* 1995). Similarly the half life of non-dividing antigen specific plasma cells from bone marrow have also been estimated using exponential decay (Slifka *et al.* 1998). While the assumptions implicit in fitting exponentials are not always stated in these studies, both Fulcher & Basten (1997) and Tanchot & Rocha (1998) argue that the exponential nature of the replacement curves is indicative of time independent, stochastic processes governing cell survival.

Despite this success, recent rigorous quantitative experimental analyses of cells triggered to either die or to divide have revealed data that is inconsistent with the simple exponential. In each case the variation is better described by a time-sensitive, non-homogeneous distribution similar to the lognormal (Deenick *et al.* 2003; Tangye *et al.* 2003; Hawkins *et al.* in preparation). This result argues that the stochastic events governing variation intrinsic to cells are not applied constantly, but can in some way be timed, and presumably regulated positively and negatively by incoming cell signals.

There are many mathematical and philosophical implications raised by changing from a time-independent probability of change to a time and history sensitive distribution such as the lognormal. An important difference is the need to consider the age of cells, or cell cohorts, over time. This led us to search for novel mathematical results that may be useful for studying age-sensitive systems. Given the more recent experimental results favouring time-sensitive distributions we wondered if the data supporting the fitting of exponential functions could be accommodated within a new framework, where cells have a determined lifespan, and if so, whether experiments could be devised to distinguish them. In solving the problem we develop a general solution for equilibrium states and apply it to a number of source distributions.

## 2. DYING POPULATION WITH EXTERNAL REPLACEMENT

The system to be studied is illustrated in figure 1. Cells are generated at an external source and used to replace cells lost by attrition from a mature cell pool. We initially ignore any homeostatic process, such as division, for replacing cells outside of the generative source. We assume that newly generated cells vary in their lifetimes, $l$, with the distribution of lifetimes denoted by $L(l)$ and referred to as the source distribution.

It is not important for our purpose to know why lifetimes vary among cells, however we note two general scenarios. In the first the age of a cell does not affect its likelihood of dying; the rate of cell loss is constant for the population, and 'stochastic' for the individual. This age-independent model for cell death is analogous to atoms undergoing radioactive decay and yields survival curves that decay exponentially. The second is to assume that cells have a lifespan capable of being determined at birth. In this scenario cells undergo internal changes that effectively act as a 'clock' that inexorably leads to a cell dying at some point. Individual cells will vary in lifespan according to some distribution (as in figure 1). By this view external factors such as cytokines and chemokines may also act on the lifespan clock to vary the precise time to eventual death, and presumably contribute variability to the lifespan of the population. As an example T lymphocytes exported to the periphery from the thymus (the generative source) have their lifespan prolonged by low-level self-antigen recognition (Kirberg *et al.* 1997; Tanchot *et al.* 1997; Nesic & Vukmanovic 1998; Ernst *et al.* 1999). As the exact form
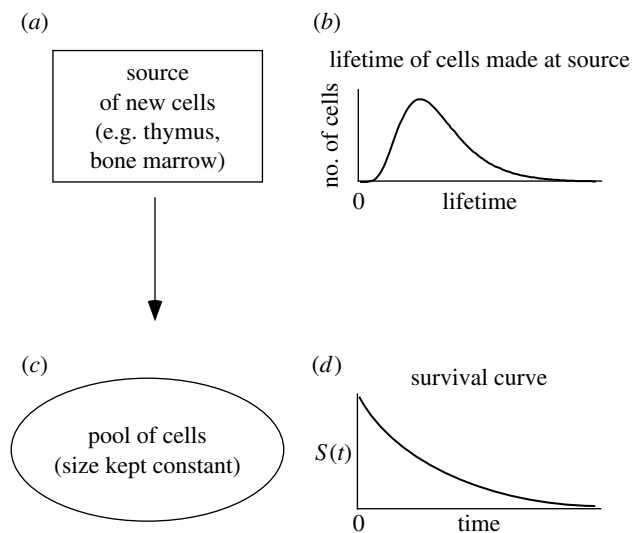


Figure 1. In the system studied, cells are generated at a source (*a*) with a defined lifetime that follows a distribution such as that shown in (*b*). The cells are non-dividing. The pool of cells (*c*) is in continual flux as cells that die are replaced from the source. We are interested in determining the survival curve shown in (*d*). That is, at any time what fraction of the total population, $S(t)$, will remain alive at a time $t$ later?

of this lifespan variation is unknown we sought a general solution here, but consider lognormal and gamma distributions as candidate possibilities for lymphocytes.

The question we ask relates to the changes in survival times of the cells in the mature cell pool: At a given time what is the distribution for remaining lifetimes in the population? We are particularly interested to know if the corresponding survival curve can be distinguished from the exponential function usually applied. We also aim to calculate the current distribution of total lifespans in the pool at any time, which will be different in general from the source distribution as longer-lived cells spend proportionally longer in the pool.

The two characteristics of a cell that we need to follow are its current age, $a$, and its pre-determined lifetime, $l$, which we will refer to as the *total lifetime* for clarity. We will find it more convenient to work with the *remaining lifetime*, $\lambda = l - a$, and total lifetime, $l$, as the two cell variables, which clearly contain the age information as well. We will be interested in how the joint probability density $p(\lambda, l, t)$ of remaining lifetimes and total lifetimes of the population evolves with time $t$. The probability density must be normalized at all times

$$\int_0^\infty \mathrm{d}\lambda \int_0^\infty \mathrm{d}l\, p(\lambda, l, t) = 1. \tag{2.1}$$

A cell's remaining lifetime cannot be greater than its total lifetime, therefore $p(\lambda, l, t) = 0$ when $\lambda > l$.

The two marginal distributions of $p(\lambda, l, t)$ give the distribution of remaining lifetimes and total lifetimes in the population, respectively

$$p_{\mathrm{rl}}(\lambda, t) = \int_0^\infty \mathrm{d}l\, p(\lambda, l, t), \tag{2.2}$$

$$p_{\mathrm{tl}}(l, t) = \int_0^\infty \mathrm{d}\lambda\, p(\lambda, l, t). \tag{2.3}$$

A common experiment in lifetime-distribution studies is to monitor the fraction of survivors remaining as the cells are allowed to die. We refer to the curve so-obtained as the *survival curve*, $S(t)$. The survival curve is determined only by the distribution of remaining lifetimes in the population

$$S(t) = 1 - \int_0^t \mathrm{d}\lambda\, p_{\mathrm{rl}}(\lambda, t'), \qquad (2.4)$$

where $t$ is measured from the time of interest, $t'$. We will be particularly interested in what the survival curve looks like in steady-state.

### 2.1. Age-independent model

Before presenting results for the full age-dependent model we first study the simpler model where the rate of death of cells is independent of their age. Let $p$ be the total number of cells in the population. Without replacement cells die at a rate $1/\tau$, where $\tau$ is the *time constant*, according to the differential equation

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -p/\tau. \qquad (2.5)$$

The solution to this differential equation is

$$p(t) = p(0)\exp(-t/\tau), \qquad (2.6)$$

i.e. the survival curve $S(t) = p(t)/p(0)$ is a decaying exponential.

The age-independent model with constant death rate is indistinguishable from an age-dependent model where cell lifespans are determined at birth and distributed according to an exponential distribution

$$L(l) = \exp(-l/\tau)/\tau. \qquad (2.7)$$

To see this, we note that the exponential distributions have the property of being *memoryless* (time-insensitive), that is, if dying cells are not replaced the distribution of remaining lifetimes of the remaining cells is an identical exponential distribution. If the dying cells are actually replaced from an external source, the newly generated cells have lifetimes with the same exponential distribution. Hence the distribution of remaining lifetimes does not alter or mature over time in time

$$p_{\mathrm{rl}}(\lambda) = \exp(-\lambda/\tau)/\tau, \qquad (2.8)$$

and the survival curve is always exponential

$$S(t) = \exp(-t/\tau). \qquad (2.9)$$

### 2.2. Age-dependent model

We now derive results for the model described above in which the lifespan of cells is determined at birth and distributed according to some known distribution. If that distribution is exponential we recover the age-independent model with a constant death rate. However if the distribution is 'time-sensitive', i.e. not exponential, the rate of death will depend on the distribution of remaining lifetimes in the population, and the survival curve can take forms other than the exponential.

We begin by deriving a partial differential equation that describes the evolution of the distribution of remaining lifetimes, $p_{\mathrm{rl}}(\lambda, t)$. As noted above, this marginal distribution is sufficient to determine the form of the survival curve, which is often observed in experiments. We will be particularly interested in the circumstances under which the survival curve is similar to or distinguishable from an exponential or double exponential, thus providing insight into possible experimental tests of our model.

Between time $t$ and $t + \Delta t$ a proportion $p_{\mathrm{rl}}(0, t)\Delta t$ of the cells die (that is those that have between 0 and $\Delta t$ remaining lifetime). These are replaced by new cells from the source with their total lifetime remaining. The proportion of these new cells that have a remaining lifetime between $\lambda$ and $\lambda + \Delta \lambda$ is $L(\lambda)\Delta\lambda$. Those cells that do not die reduce their remaining lifetime by $\Delta t$. Mathematically this means that

$$p_{\mathrm{rl}}(\lambda, t + \Delta t)\Delta\lambda$$
$$= p_{\mathrm{rl}}(0, t)\Delta t L(\lambda)\Delta\lambda + p_{\mathrm{rl}}(\lambda + \Delta t, t)\Delta\lambda. \qquad (2.10)$$

In reality it may take a finite time for the drop in population to be detected and the source to be stimulated to produce new cells. However, we assume these times to be relatively small compared with the width of the source distribution and so to a good approximation we may consider the replacement to be continuous. In this limit we may expand both sides of equation (2.10) in a Taylor series to first order in $\Delta t$ to obtain a differential equation for $p_{\mathrm{rl}}(\lambda, t)$

$$p_{\mathrm{rl}}(\lambda, t)\Delta\lambda + \frac{\partial p_{\mathrm{rl}}(\lambda, t)}{\partial t}\Delta t\Delta\lambda$$
$$= L(\lambda)p_{\mathrm{rl}}(0, t)\Delta t\Delta\lambda + p_{\mathrm{rl}}(\lambda, t)\Delta\lambda$$
$$+ \frac{\partial p_{\mathrm{rl}}(\lambda, t)}{\partial\lambda}\Delta t\Delta\lambda$$
$$\Rightarrow \frac{\partial p_{\mathrm{rl}}(\lambda, t)}{\partial t} = L(\lambda)p_{\mathrm{rl}}(0, t) + \frac{\partial p_{\mathrm{rl}}(\lambda, t)}{\partial\lambda}. \qquad (2.11)$$

The boundary condition we impose is that $p_{\mathrm{rl}}(\lambda \rightarrow \infty,\ t) = 0$, as is necessary for any probability density.[1] An intuitive schematic for the evolution of the distribution of remaining lifetimes under equation (2.11) is shown in the electronic supplementary material, section A.

After evolving for some transient time according to the above differential equation, we expect the cell population to settle down to some *equilibrium* or *steady-state* distribution $p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda)$ (independent of time). To find this distribution we set $\partial p_{\mathrm{rl}}/\partial t = 0$ in equation (2.11)

$$0 = L(\lambda)p_{\mathrm{rl}}^{\mathrm{ss}}(0) + \frac{\partial p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda)}{\partial\lambda}$$
$$\Rightarrow p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda) = -p_{\mathrm{rl}}^{\mathrm{ss}}(0)\int_0^\lambda \mathrm{d}\lambda'\, L(\lambda') + C,$$

---

[1] With this condition it is straightforward to check that $\frac{\mathrm{d}}{\mathrm{d}t}\int_0^\infty \mathrm{d}\lambda\, p_{\mathrm{rl}}(\lambda, t) = 0$. In other words the differential equation conserves normalization—if we start with a normalized probability density it will remain so for all time. This is as expected since we derived the equation by assuming all cells that die are replaced.
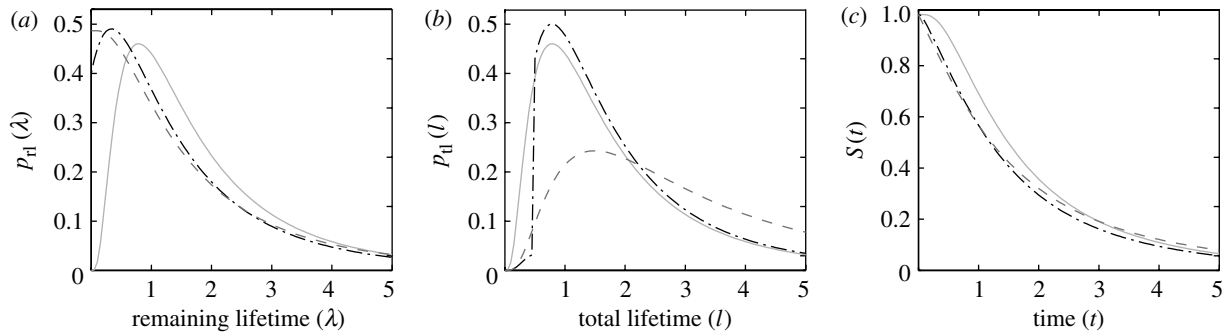
Figure 2. Dying population with replacement from lognormal source distribution—$m = 0.4$ and $s = 0.8$. Initial ($t = 0$—solid line), intermediate ($t = 0.5$—dot-dashed line) and steady-state ($t \to \infty$—dashed line) curves are shown to illustrate the approach to equilibrium. (*a*) Distribution of remaining lifetimes in the population, $p_{\mathrm{rl}}(\lambda, t)$, where $p_{\mathrm{rl}}(\lambda, \infty) = p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda)$. (*b*) Distribution of total lifetimes in the population, $p_{\mathrm{tl}}(l, t)$, where $p_{\mathrm{tl}}(\lambda, \infty) = p_{\mathrm{tl}}^{\mathrm{ss}}(l)$. (*c*) Survival curves, $S(t)$, obtained from the corresponding remaining-lifetime distributions of (*a*).

where $C$ is the constant of integration. Setting $\lambda = 0$ we see that $C = p_{\mathrm{rl}}^{\mathrm{ss}}(0)$.

Finally, the requirement that $p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda)$ be normalized, $\int_0^\infty \mathrm{d}\lambda\, p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda) = 1$ determines $p_{\mathrm{rl}}^{\mathrm{ss}}(0) = 1/\mu$, where

$$\mu = \int_0^\infty \mathrm{d}\lambda \left[ 1 - \int_0^\lambda \mathrm{d}\lambda'\, L(\lambda') \right] = \int_0^\infty \mathrm{d}\lambda\, \lambda L(\lambda),$$

is the mean of the source distribution. Therefore we have

$$p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda) = \frac{1 - \int_0^\lambda \mathrm{d}\lambda'\, L(\lambda')}{\mu}. \qquad (2.12)$$

In steady-state, cells with a particular total lifetime, $l$, will be completely unsynchronized in birth times, that is their ages or remaining lifetimes will be uniformly spread from 0 to $l$. Therefore we expect the steady-state distribution of total lifetimes to have the form

$$p_{\mathrm{tl}}^{\mathrm{ss}}(l) = \frac{l L(l)}{\mu}, \qquad (2.13)$$

where the mean lifetime, $\mu$, serves as the normalization of this marginal distribution as well.

In fact, it is possible to write a differential equation to describe the evolution of the full distribution of remaining lifetimes and total lifetimes, $p(\lambda, l, t)$, as described in the electronic supplementary material, section B. We are able to solve this full model in steady state, which proves equation (2.13) and is consistent with equation (2.12) for the marginal distributions.

From equation (2.13) it is clear that with the population at steady-state the distribution of total lifetimes is skewed towards longer lifetimes compared with the source distribution. This is a general feature that occurs independent of the particular source distribution and is due to the fact that longer-lived cells spend proportionally longer in the population before dying.

We now explicitly study the dynamics and steady-state solutions of our model for particular choices of source distribution. For the dynamics, i.e. the transition to steady state, we use numerical methods as described in the electronic supplementary material, section C. As alluded to earlier, we are particularly interested in the lognormal distribution as it is ubiquitous in nature and has relevance to recent experimental studies. We contrast the lognormal results with those of a gamma distribution, which has an interpretation as a more-sophisticated form of age-independent model.

*2.2.1. Lognormal distribution.* The lognormal distribution is defined as

$$L(l) = \frac{1}{\sqrt{2\pi} s l} \exp\left( -\frac{(\log(l) - m)^2}{2s^2} \right), \qquad (2.14)$$

where $m$ and $s$ are the mean and standard deviation of the *natural logarithms* of the lifetimes, which are normally distributed. The mean $\mu = \int_0^\infty \mathrm{d}l\, l L(l)$ and standard deviation $\sigma = \sqrt{\int_0^\infty \mathrm{d}l\, (l - \mu)^2 L(l)}$ of the lifetimes are related to the mean $m$ and standard deviation $s$ of the *logarithms* of the lifetimes via

$$\mu = \exp(m + s^2/2),$$

$$\sigma = \sqrt{\exp(2(m + s^2)) - \exp(2m + s^2)}.$$

Lognormal distributions can arise from many small multiplicative random events, through the multiplicative version of the central limit theorem. The distribution has a characteristic 'long tail', as seen in figure 2 for example. Many processes in biology are known to follow lognormal distributions.

As described in §2.2 we calculate the steady-state distribution of remaining lifetimes

$$p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda) = \frac{\mathrm{e}^{-(m+s^2/2)}}{2} \left[ 1 - \mathrm{erf}\left( \frac{\log(\lambda) - m}{\sqrt{2} s} \right) \right], \qquad (2.15)$$

of total lifetimes

$$p_{\mathrm{tl}}^{\mathrm{ss}}(l) = \frac{1}{\sqrt{2\pi} s} \exp\left( -\frac{(\log(l) - m)^2}{2s^2} - m - s^2/2 \right), \qquad (2.16)$$

and the equilibrium survival curve

$$S(t) = \frac{1}{2} \left[ 1 - \mathrm{e}^{-(m+s^2/2)} t \left( 1 - \mathrm{erf}\left( \frac{\log(t) - m}{\sqrt{2} s} \right) \right) \right.$$
$$\left. - \mathrm{erf}\left( \frac{\log(t) - (m + s^2)}{\sqrt{2} s} \right) \right], \qquad (2.17)$$
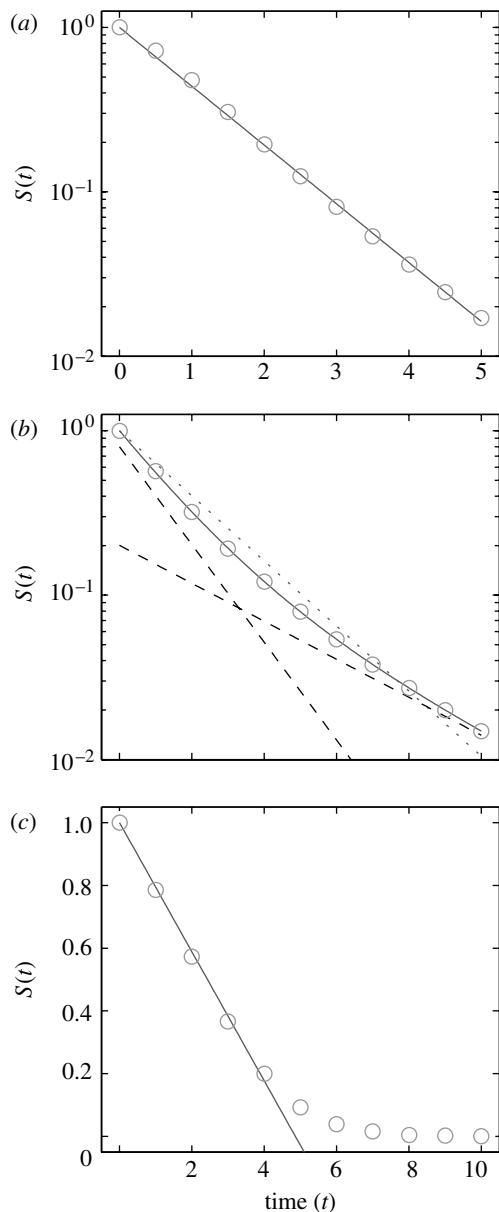
Figure 3. Fits to equilibrium survival data for varying lognormal source parameters. The circles are ideal data generated from our model—i.e. points on the survival curve, $S(t)$—and the solid lines are fits (exponential, double exponential or linear). (a) $m = 0.4$ and $s = 0.6$: the survivor data are fit quite well with a single exponential—a straight line on the semilog graph—with $\tau = 1.21$ (solid line). (b) $m = 0.4$ and $s = 0.8$: the survivor data are fit very well by a double exponential (solid line) with $A = 0.80$, $\tau_1 = 1.46$ and $\tau_2 = 3.76$. The two dashed lines are the two exponentially dying components. The double-exponential fit is much better than can be achieved by a single exponential (dotted line) with $\tau = 2.19$. (c) $m = 1.5$ and $s = 0.3$: the survivor data initially follows a linear decay (solid line—fit to data until $t = 4$).

where $\mathrm{erf}(\cdot)$ is the *error function* defined as the area under the normal distribution from 0 to $x$,

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{d}x' \exp(-x'^2).$$

We illustrate the distributions and survival curve obtained from our model in figure 2 for a particular choice of parameters, $m = 0.4$ and $s = 0.8$. As well as

the analytic steady-state solutions, above, we illustrate the approach to equilibrium, starting with a lognormal population, obtained from numerical simulation as described in the electronic supplementary material, section C. The distribution of remaining lifetimes approaches equilibrium rather quickly, whereas the distribution of total lifetimes takes considerably longer.

It is important to note that in contrast to the age-independent model, the distribution of remaining lifetimes with lognormal replacement changes in time. Even so, the steady-state distribution that is reached retains the long tail of the underlying distribution, which has important implications for the survival curve. This is quite a different phenomenon to accumulating longer-lived cells in the population, which will occur regardless of the underlying distribution, even the exponential.

The survival curve in steady-state can look very different for different parameter values. In figure 3 we illustrate three distinct shapes for the survival curves with different parameters. In each case we plot ideal survival data (circles) generated from our model, i.e. points on the survival curve, and fit common functions to it.

In figure 3a, $m = 0.4$ and $s = 0.6$, we obtain a good fit to the survival data with a single exponential until such time as only about 1% of the original population remains alive. If we were to fit for longer times we would find that the single exponential would begin to deviate from the survival data, and we would require a double exponential to obtain a good fit. A single exponential, which is usually the first function that one would attempt to fit to any survival curve, fits well when $s$ is approximately the same, or only slightly larger than $m$.

In figure 3b, $m = 0.4$ and $s = 0.8$ (as in figure 2), we show that a double exponential

$$S(t) \simeq A \exp(-t/\tau_1) + (1 - A)\exp(-t/\tau_2),$$

is sometimes required to obtain a good fit to the survival data even at intermediate times when a large fraction of the original population remains alive. The double exponential nature of the survival data is clearly demonstrated by the two distinct linear regions on the log scale. Such a curve might otherwise be interpreted as evidence for two 'subpopulations', one dying-off exponentially with time constant $\tau_1$ ('short-lived') and the other with time constant $\tau_2$ ('long-lived'). However in our model the survival curve arises from a single generative mechanism—there is no true division into two subpopulations. Indeed if we fit to the survival data over a longer time we would obtain a different split ratio, $A$, and different time constants. A double-exponential survival curve typically arises when the standard deviation, $s$, of the logarithms of the lifetimes is much larger than the mean, $m$, that is when the distribution has the characteristic long tail as seen in figure 2.

Finally, in figure 3c, $m = 1.5$ and $s = 0.3$, we see that the decay can look linear for short times. This situation corresponds to a narrow distribution, i.e. when $s$ is much smaller than $m$ that can look qualitatively similar to a normal (Gaussian) distribution. In such cases the

equilibrium distribution of remaining lifetimes is flat initially, due to the cells' ages being unsynchronized together with only little variance in cell lifetime. See figure 5 for an illustration of this type of distribution.

It is interesting that the survival curve obtained from the lognormal age-dependent model in steady-state can often be closely approximated by a single exponential survival curve. Hence experimental evidence for an exponential survival curve is not able to distinguish the lognormal 'fate determined' source population from a simple random loss from the mature pool (age-independent death). A double-exponential survival curve can also be explained in both models: as two subpopulations with different half-lives, or derived from a lognormal source distribution with a large variance. However in contrast to the empirical conclusion of two subpopulations required in the age-independent analysis, the lognormal age-dependent model offers an appealing mechanism for the apparent appearance of two subpopulations when no true distinction exists.

It may be possible to acquire very accurate data with many time points to help make a distinction between the two explanations. To do so would require monitoring the population in the long-time (small-remaining-population) limit. If the lognormal explanation is correct the best-fit double exponential will change when fit to data over a longer time. If there is a true division into two subpopulations with different survival rates then the best-fit double exponential should remain the same.

*2.2.2. Gamma distribution.* To contrast the lognormal results we consider the gamma distribution as a possible source distribution, $L(l) = \gamma(\alpha, \tau, l)$. The gamma distribution is defined as

$$\gamma(\alpha, \tau, l) = \frac{l^{\alpha-1} \exp(-l/\tau)}{\Gamma(\alpha)\tau^\alpha}, \qquad (2.18)$$

where $\Gamma(\alpha) = \int_0^\infty \mathrm{d}x\, x^{\alpha-1} \mathrm{e}^{-x}$ is the gamma function ($\Gamma(n) = (n-1)!$ when $n$ is an integer), and $\alpha \in [1, \infty)$, $\tau \in (0, \infty)$ are parameters. The parameter $\tau$ is a time-constant that plays an analogous role to $\tau$ in the exponential distribution. The mean and standard deviation of the gamma distribution are given by

$$\mu = \alpha\tau,$$
$$\sigma = \sqrt{\alpha}\tau.$$

When $\alpha$ is an integer, $n$, the distribution can be interpreted as the time taken for a cell to pass through $n$ 'compartments' (for example phases of the cell cycle), where the time spent in each compartment is exponentially distributed with the same time constant $\tau$. This compartment approach is common in modelling biological process and is a more-sophisticated form of age-independent model. One might hope to capture most of the features of a lognormal age-dependent model with a gamma model as then one could resort to the simpler age-independent approach to death rates. However we will show that the gamma distribution does not ever produce a double-exponential survival curve as we obtained from the lognormal distribution,

so there is at least one unique feature of our lognormal age-dependent model.

For a gamma source distribution with $\alpha = n$ we derive the steady-state distributions of remaining lifetimes and total lifetimes, and the survival curve as described in §2.2

$$p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda) = \frac{1}{n}\sum_{j=1}^{n}\gamma(j, \tau, \lambda), \qquad (2.19)$$

$$p_{\mathrm{tl}}^{\mathrm{ss}}(l) = \gamma(n+1, \tau, l), \qquad (2.20)$$

and

$$S(t) = 1 - \frac{1}{n}\sum_{j=1}^{n}\frac{\bar{\Gamma}(j, t/\tau)}{(j-1)!}, \qquad (2.21)$$

where

$$\bar{\Gamma}(\alpha, x) = \int_0^x \mathrm{d}x'\, x'^{\alpha-1}\exp(-x'),$$

is the incomplete gamma function.

The fact that $p_{\mathrm{tl}}^{\mathrm{ss}}(l)$ is itself a gamma distribution with $\alpha = n+1$ illustrates that the total lifetime distribution is skewed towards longer lifetimes compared with the source distribution. This is true even for the $\alpha = 1$ (exponential distribution) case, which is equivalent to the age-independent model, despite the fact that the remaining lifetime distribution does not change in time.

The remaining-lifetime distribution $p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda)$ can be seen to be a sum of gamma distributions with $\alpha = j$ in the sum, where $j = 1 \ldots n$. This distribution can be understood in terms of a compartment model by noting that our model corresponds to differential equations for the compartment populations, $p_k$, $k = 0 \ldots n-1$ of the form

$$\frac{\mathrm{d}p_k(t)}{\mathrm{d}t} = \frac{1}{\tau}(p_{k-1}(t) - p_{k+1}(t)), \qquad (2.22)$$

that is cells passing from one compartment to the next at rate $1/\tau$, where the indices are interpreted modulo $n$, so dying cells (i.e. those leaving compartment $n-1$) are replaced in compartment 0. The steady-state solution of these differential equations is clearly

$$p_k^{\mathrm{ss}} = 1/n, \quad k = 0 \ldots n-1,$$

i.e. all $n$ compartments equally populated with $1/n$ of the cells. The distribution of remaining lifetimes is given by $p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda)$ in equation (2.19) because the cells have anywhere from 1 to $n$ compartments yet to traverse ($j = n-k$), and those in each compartment have remaining lifetimes given by a gamma distribution.

In figure 4a,b we plot the initial and equilibrium distributions obtained from a gamma source distribution with $\alpha = 2$, $\tau = 1$. Superficially, the gamma distribution can look very similar to a lognormal, especially for short time, for example the gamma distribution plotted here looks similar to the lognormal with $m = 0.4$ and $s = 0.8$ plotted in figure 2. However, lognormal distributions typically have longer tails than gamma distributions due to $\log(l)$ appearing in the exponent of equation (2.14) compared with $l$ in equation (2.18). This feature has important consequences for the type of survival curve that can be obtained with the gamma.
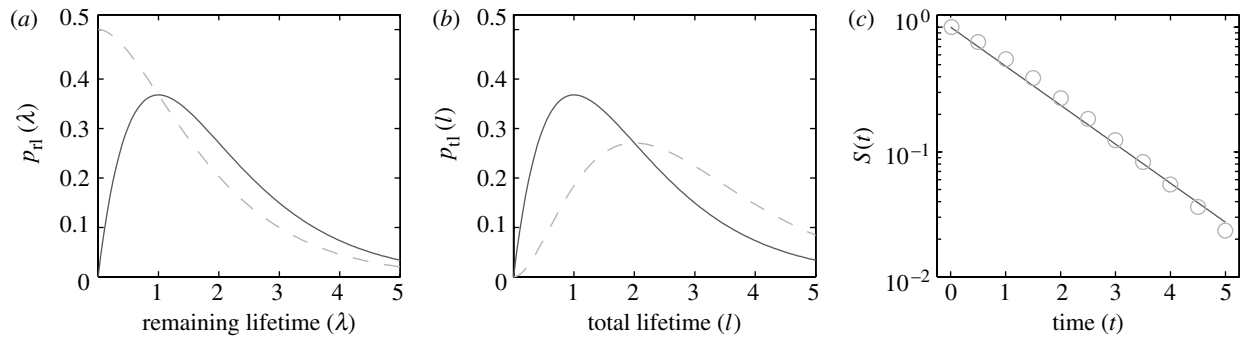
Figure 4. Dying population with replacement from gamma source distribution—$\alpha = 2$ and $\tau = 1$ corresponding to mean and standard deviation of $\mu \simeq 2$, $\sigma \simeq 1.41$. (*a*) Distribution of remaining lifetimes in the population, $p_{\rm rl}(\lambda, t)$. The solid line is the initial gamma distribution ($t = 0$), and the dashed line is the steady-state distribution, $p_{\rm rl}^{\rm ss}(\lambda)$ ($t \to \infty$). (*b*) Distribution of total lifetimes in the population, $p_{\rm tl}(l, t)$. The solid line is the initial gamma distribution ($t = 0$), and the dashed line is the steady-state distribution, $p_{\rm tl}^{\rm ss}(l)$ ($t \to \infty$). (*c*) Exponential fit to ideal equilibrium survival data (circles—generated as points on $S(t)$). A double exponential will not achieve a better fit of this data because $\mathrm{d}^2 \log(S(t))/\mathrm{d}t^2 < 0$ (concave down on the log scale).

For small $\alpha$ the survival curve derived from the gamma distribution can be well-approximated by an exponential. In figure 4c we show survival data taken from our model and fit with an exponential on the log scale, for $\alpha = 2$, $\tau = 1$. The fit is reasonable for $\alpha = 2$ but becomes worse for larger $\alpha$. Importantly, because $\mathrm{d}^2 \log(S(t))/\mathrm{d}t^2 < 0$ a double exponential cannot give a better fit than a single exponential in this case. Hence we cannot interpret the equilibrium population as being composed of two subpopulations, one short-lived and one long-lived. In fact, it is possible to show that for any $\alpha = n$ an integer, $\mathrm{d}^2 \log(S(t))/\mathrm{d}t^2 < 0$, so under no circumstances is survival curve obtained from a gamma source distribution better approximated by a double exponential. This feature is in stark contrast to the lognormal, which produces survival curves that appear double-exponential when the standard deviation is large.

## 3. DIVIDING POPULATION

We now turn to a related problem where times to cell division (rather than death) vary according to a lognormal distribution in a continuously growing cell population. We are interested in finding the distribution of 'time until next division' in the growing population, where cells are dividing completely out of synchrony, given that we know the distribution of division times of newly divided cells. We will denote this distribution $D(\cdot)$ and refer to it as the *division-time distribution*—it plays a role analogous to the source distribution in §2. The two daughter cells produced by each division are assumed to have division times drawn from the distribution that are completely uncorrelated with each other, or with the parent.

This problem is very similar to the problem of cell death with replacement from an external source considered in §2. Cell division can be thought of as cell death with two replacement cells (the daughters) instead of one. Analogously, the two cell variables of interest are its age, $a$, and *total division time*, $d$, however we will find it more convenient to work with total division time, $d$, and *time to next division*, $\delta = d - a$. The full problem is to find the evolution of

the *population density*, $P(\delta, d, t)$, describing the total size of the population and the distribution of $\delta$ and $d$ within it, given the division-time distribution. Because the population will grow in time, unlike in §2, $P(\delta, d, t)$ is not normalized to one (which is why we have capitalized it and refer to the population density rather than the probability density)—the normalization

$$N(t) = \int_0^\infty \mathrm{d}\delta \int_0^\infty \mathrm{d}d \, P(\delta, d, t) \qquad (3.1)$$

is the total population at time $t$.

The marginals

$$P_{\rm tnd}(\delta, t) = \int_0^\infty \mathrm{d}d \, P(\delta, d, t),$$

$$P_{\rm td}(d, t) = \int_0^\infty \mathrm{d}\delta \, P(\delta, d, t)$$

contain information of the total population size and the distributions of time to next division

$$p_{\rm tnd}(\delta, t) = P_{\rm tnd}(\delta, t)/N(t), \qquad (3.2)$$

and total division time

$$p_{\rm td}(d, t) = P_{\rm tnd}(d, t)/N(t), \qquad (3.3)$$

respectively.

The full model of division is presented in the electronic supplementary material, section B. However, analogously to §2 we are most interested in the marginal density of time to next division and choose to focus on it here. We may derive a differential equation for its evolution

$$\frac{\partial P_{\rm tnd}(\delta, t)}{\partial t} = 2D(\delta)P_{\rm tnd}(0, t) + \frac{\partial P_{\rm tnd}(\delta, t)}{\partial \delta}. \qquad (3.4)$$

We will look for solutions to this equation of the form $P_{\rm tnd}(\delta, t) = N(t)p_{\rm tnd}^{\rm ss}(\delta)$. Such a solution describes a growing population in which the distribution of time to next division has reached a steady state, $p_{\rm tnd}^{\rm ss}(\delta)$.

From its definition we can derive a differential equation for the normalization

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \int_0^\infty \mathrm{d}\delta P_{\mathrm{tnd}}(\delta, t)$$

$$= \int_0^\infty \mathrm{d}\delta \frac{\partial P_{\mathrm{tnd}}(\delta, t)}{\partial t}$$

$$= \int_0^\infty \mathrm{d}\delta \left[ 2D(\delta)P_{\mathrm{tnd}}(0, t) + \frac{\partial P_{\mathrm{tnd}}(\delta, t)}{\partial \delta} \right]$$

$$= P_{\mathrm{tnd}}(0, t), \qquad (3.5)$$

where the third line follows from equation (3.4), and for the fourth line we have used the fact that the population density must vanish at infinity, $P_{\mathrm{tnd}}(\delta \to \infty, t) = 0$.

For a solution of the form $P_{\mathrm{tnd}}(\delta, t) = N(t)p_{\mathrm{tnd}}^{\mathrm{ss}}(\delta)$, equation (3.5) becomes

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = kN(t), \qquad (3.6)$$

with solution

$$N(t) = N(0)\exp(kt), \qquad (3.7)$$

where $k = p_{\mathrm{tnd}}^{\mathrm{ss}}(0)$ is a constant that will be fixed below. That is, the total size of the population grows exponentially in a steady-state solution. This exponential growth of the total population size is exactly what would have been obtained in a model where the likelihood to divide is independent of age, i.e. constant across the population. The difference in our model is that the distribution of time to next division will not be exponential as it is in the age-independent model.

With $N(t)$ of the form equation (3.7), equation (3.4) reduces to

$$\frac{\mathrm{d}p_{\mathrm{tnd}}^{\mathrm{ss}}(\delta)}{\mathrm{d}\delta} - kp_{\mathrm{tnd}}^{\mathrm{ss}}(\delta) + kD(\delta) = 0. \qquad (3.8)$$

We solve this differential equation using an integrating factor

$$p_{\mathrm{tnd}}^{\mathrm{ss}}(\delta) = ke^{k\delta}\left(1 - 2\int_0^\delta \mathrm{d}\delta' D(\delta')e^{-k\delta'}\right), \qquad (3.9)$$

where the constant of integration was determined by self-consistency of the two sides of the equation at $\delta = 0$. The constant $k = p_{\mathrm{tnd}}^{\mathrm{ss}}(0)$ is determined by

$$\int_0^\infty \mathrm{d}\delta' D(\delta')e^{-k\delta'} = 1/2, \qquad (3.10)$$

as is necessary for $p_{\mathrm{tnd}}^{\mathrm{ss}}(\delta \to \infty) = 0$.

In the electronic supplementary material, section B, we derive an expression for the marginal distribution of total division times in steady state, $p_{\mathrm{td}}^{\mathrm{ss}}(d)$, by solving the differential equation for the full distribution.

Analogously to the survival curve of §2, we define the *undivided curve* as

$$U(t) = 1 - \int_0^t \mathrm{d}\delta p_{\mathrm{tnd}}(\delta, t'), \qquad (3.11)$$

which is the fraction of the population that remains undivided at a time $t$ after the time of interest, $t'$.

We choose to illustrate our results for a lognormal division-time distribution as there is recent experimental evidence for its relevance (Deenick *et al.* 2003; Tangye *et al.* 2003). Results for a gamma division-time distribution can also be derived and contrasted with the lognormal, as in §2.

When the division-time distribution is lognormal, equation (2.14), we were not able to find a closed-form expression for the integral $\int_0^\delta \mathrm{d}\delta' D(\delta')e^{-k\delta'}$ that appears in equations (3.9) and (3.10) ($\delta \to \infty$) in terms of standard elementary or special functions. We can, however, evaluate this integral numerically for specific values of $m$, $s$, $k$ and $\delta$ using standard numerical integration techniques. We can therefore determine $k = p_{\mathrm{tnd}}^{\mathrm{ss}}(0)$ and plot $p_{\mathrm{tnd}}^{\mathrm{ss}}(\delta)$ for specific values of $m$ and $s$.

In figure 5*a*–*c* we illustrate the steady-state distributions and undivided curve obtained for the dividing cell population. Also shown are the analogous distributions and survival curve for the dying population with external replacement. The division-time lognormal has $m = 1.5$ and $s = 0.3$, which were chosen to clearly illustrate the difference between the distributions obtained for the two cases. We see that the steady-state distributions are qualitatively similar in the two models, however they are slightly skewed towards the division-time lognormal for the dividing model compared with age-dependent death model. This effect is due to contribution of recently divided cells on the statistics of the growing population—it is a dynamic steady state compared with a static steady state.

Finally, figure 5*d* illustrates the growth of the overall size of the population, $N(t)$, with time in a numerical simulation. We see that after the initial transient time $N(t)$ grows exponentially, as expected from the steady-state solution. Also the marginal distributions obtained numerically in steady-state closely match the analytic expressions. Hence we conclude that the system does, in fact, evolve towards the steady-state solutions we derived above.

## 4. DISCUSSION

During lymphocyte homeostasis cells lost through attrition are replaced. Many models of lymphocyte homeostasis assume that the probability of a cell dying is constant across the population and independent of the age of the cell. This set of assumptions yields an exponential function to describe remaining lifetimes in the population. In experiments examining lymphocyte survival *in vitro*, cells removed from their homeostatic environment follow exponential decay to a first approximation (Gett & Hodgkin 2000; Deenick *et al.* 2003). However, more accurate experiments clearly show the rate of loss to be skewed and closely fit to a time-sensitive distribution closely approximated by the lognormal (Hawkins *et al.* in preparation). This result caused us to speculate that a similar timed, and possibly lognormally distributed lifespan *in vivo* might lead to a steady state where cell replacement was similar to the experimentally observed exponential. Our goal, therefore, was to
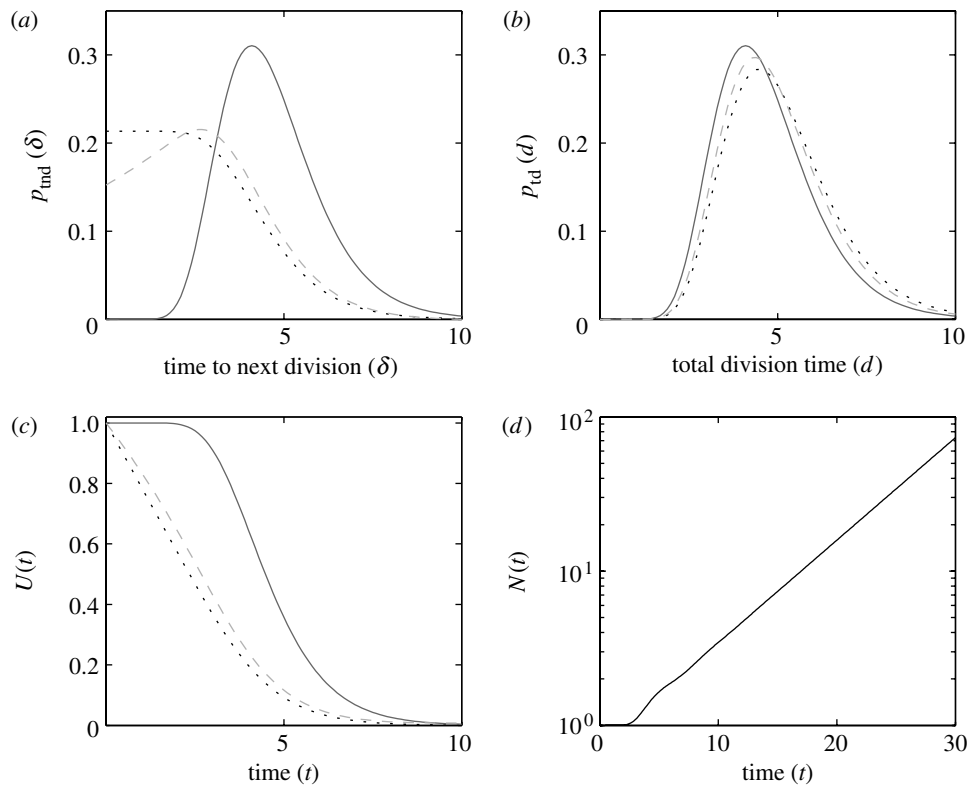
Figure 5. Dividing population with lognormal division-time distribution—$m = 1.5$ and $s = 0.3$. Initial ($t = 0$—solid line) and steady-state ($t \to \infty$—dashed line) curves are shown, as well as the analogous steady-state curves for the dying population (dotted lines), for comparison. (*a*) Distribution of time until next division, initial lognormal distribution and $p_{\mathrm{tnd}}^{\mathrm{ss}}(\delta)$, cf. remaining lifetime, $p_{\mathrm{rl}}^{\mathrm{ss}}(\lambda)$. (*b*) Distribution of total division time, initial lognormal distribution and $p_{\mathrm{td}}^{\mathrm{ss}}(d)$, cf. total lifetime, $p_{\mathrm{tl}}^{\mathrm{ss}}(l)$. (*c*) Undivided curves, $U(t)$, initial and steady-state, obtained from the corresponding time-to-next-division distributions of (*a*), cf. steady-state survival curve. (*d*) Growth of total population size, $N(t)$, with time in the dividing population. Note that the growth settles down to exponential within a few division cycles of most cells (straight line on the log scale), as expected in steady-state.

ask whether at steady state the survival curves, or equivalently remaining lifetime distributions, of each of the models can be distinguished.

A time-sensitive distribution has numerous consequences for development of models of homeostasis and cell dynamics. Our results show that under some circumstances the survival data from the age-dependent model can be closely fit by an exponential and therefore it will be difficult to make a distinction from the age-independent model experimentally. Hence fitting an exponential function to a survival curve does not ensure the underlying assumption of random loss in the pool is correct.

It is interesting that the lognormal steady state curve can sometimes be closely approximated by two exponentials, despite the single generative mechanism. Such data would usually be interpreted as evidence for two independent populations, each with their own survival kinetic. If the source distribution is a gamma distribution, which can often appear very similar to a lognormal, this feature is not observed, thus providing a quantitative test to distinguish the lognormal from similar distributions. This result also demonstrates that there are features of our age-dependent model that cannot be captured within an age-independent model, even a more-sophisticated one involving multiple compartments.

The experimental implications for measuring turnover of peripheral naïve T and B lymphocytes, for

example by replacement with proliferating progenitors after labelling with BrDU (Fulcher & Basten 1994; Tough & Sprent 1994; Tanchot & Rocha 1997) are currently being tested. It is interesting to note that attempts to compartmentalize cells of different apparent lifespans to different origins or history will prove unproductive if lognormal replacement is operating.

It is also important to point out that T and B cell homeostasis results from proliferation of mature cells in addition to supply from a source such as the thymus. Hence, the system we describe here cannot serve as a complete model. Future models must combine a source mechanism (either exponential or, as we have defined here, a determined lifetime at steady state) with proliferation of mature cells. Such a model is significantly more complex as it must accommodate relative contributions from both mechanisms as well as take account of inheritance of lifetimes after cells divide in the periphery. Further experiments must be performed to distinguish the many theoretical possibilities. A unique population that may closely follow our steady state proposal are the antibody secreting plasma cells that deposit into bone marrow following an immune response. These non-dividing cells arise after each immune response and hence have highly unsynchronized lifetimes. The turnover of this population may provide an excellent model for our steady state proposal.

Here we have focused on the lognormal as a possible distribution for replacement in lifespan studies. This is

based in part on experimental evidence but also because the lognormal is found ubiquitously in nature (Limpert *et al.* 2001), and commonly observed in variation of cell behaviour. For example cell-surface-receptor expression levels are typically lognormally distributed. Lognormal variability is indicative of a series of multiplicative events providing some potential insight into the mechanism or manner of regulating and timing cell responses, such as division or death. For example a series of geometric steps as noted for triggering division or the onset of apoptosis could account for lognormal variation at the population level. Receptor-mediated signals that alter expression levels of key proteins in the geometric chain would be expected to alter the mean time to an event, while ensuring the overall pattern of variation is consistently lognormal. Further molecular analyses of how variation is summed to trigger a timed behavioural change may prove a fruitful area of research.

The exponential has proved to be a convenient workhorse of many mathematical models. Our demonstration here that the lognormal steady state is, under many situations, indistinguishable from the exponential, despite the quite different underlying philosophy (that is, age independence or not) should challenge the automatic adoption of constant rates of change in cell behaviour where cells are unsynchronized in starting time. For example we extended our lognormal framework to the related problem of variation in cell division times within a growing population, again motivated by our recent careful measurement of the time of entry of cells into division (Deenick *et al.* 2003; Tangye *et al.* 2003). Moreover our ability to obtain analytical results demonstrates the somewhat surprising feature that time and age-dependent approaches may not be prohibitively complicated for modelling complex, multi-dimensional systems, as may otherwise have been thought. We anticipate our analytic formulae will prove useful for extracting quantitative features of cell proliferation and allow model parameters ($m$ and $s$ for the lognormal) to be determined from directly monitoring entry to division by unsynchronized populations.

We suggest that the automatic adoption of constant rates of change by cells as a starting point for developing mathematical models may lead to errors in biological modelling that limit the value of the resulting model. Lognormal steady-state arguments may well apply to other biological systems where constant (in time and across populations or subpopulations) rates are observed. We believe it will be profitable to further explore time-sensitive regulation of cells, although additional new techniques will need to be developed.

## REFERENCES

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

Ernst, B., Lee, D. S., Chang, J. M., Sprent, J. & Surh, C. D. 1999 The peptide ligands mediating positive selection in the thymus control T cell survival and homeostatic proliferation in the periphery. *Immunity* **11**, 173–181. (doi:10.1016/S1074-7613(00)80092-8.)

Fulcher, D. A. & Basten, A. 1994 Reduced lifespan of anergic self-reactive B cells in a double-transgenic model. *J. Exp. Med.* **179**, 125–134. (doi:10.1084/jem.179.1.125.)

Fulcher, D. A. & Basten, A. 1997 Influences on the lifespan of B cell subpopulations defined by different phenotypes. *Eur. J. Immunol.* **27**, 1188–1199.

Gett, A. V. & Hodgkin, P. D. 2000 A cellular calculus for signal integration by T cells. *Nat. Immunol.* **3**, 239–244. (doi:10.1038/79782.)

Hawkins, E., Turner, M., van Gend, C. & Hodgkin, P. D. In preparation. Generalised cyton theory and the cellular calculus of lymphocyte growth and survival.

Kirberg, J., Berns, A. & von Boehmer, H. 1997 Peripheral T cell survival requires continual ligation of the T cell receptor to major histocompatibility complex-encoded molecules. *J. Exp. Med.* **186**, 1269–1275. (doi:10.1084/jem.186.8.1269.)

Limpert, E., Stahel, W. A. & Abbt, M. 2001 Log-normal distributions across the sciences: keys and clues. *Biosciences* **51**, 341.

Mclean, A. R. & Michie, C. A. 1995 *In vivo* estimates of division and death rates of human T lymphocytes. *Proc. Natl Acad. Sci. USA* **92**, 3707–3711.

Nesic, D. & Vukmanovic, S. 1998 MHC class I is required for peripheral accumulation of CD8+ thymic emigrants. *J. Immunol.* **160**, 3705–3712.

Osmond, D. G. 1991 Proliferation kinetics and the lifespan of B cells in central and peripheral organs. *Curr. Opin. Immunol.* **3**, 179. (doi:10.1016/0952-7915(91)90047-5.)

Ramalho, A. T., Curado, M. P. & Natarajan, A. T. 1995 Lifespan of human lymphocytes estimated during a six year cytogenetic follow-up of individuals accidentally exposed in the 1987 radiological accident in Brazil. *Mutat. Res.* **331**, 47–54.

Slifka, M. K., Antia, R., Whitmire, J. K. & Ahmed, R. 1998 Humoral immunity due to long-lived plasma cells. *Immunity* **8**, 363–372. (doi:10.1016/S1074-7613(00)80541-5.)

Smith, J. A. & Martin, L. 1973 Do cells cycle? *Proc. Natl Acad. Sci. USA* **70**, 1263.

Tanchot, C. & Rocha, B. 1997 Peripheral selection of T cell repertoires: the role of continuous thymus output. *J. Exp. Med.* **186**, 1099–1106. (doi:10.1084/jem.186.7.1099.)

Tanchot, C. & Rocha, B. 1998 The organization of mature T-cell pools. *Immunol. Today* **19**, 575–579. (doi:10.1016/S0167-5699(98)01344-9.)

Tanchot, C., Lemonnier, F. A., Perarnau, B., Freitas, A. A. & Rocha, B. 1997 Differential requirements for survival and proliferation of CD8 naive or memory T cells. *Science* **276**, 2057–2062. (doi:10.1126/science.276.5321.2057.)

Tangye, S. G., Avery, D. T., Deenick, E. K. & Hodgkin, P. D. 2003 Intrinsic differences in the proliferation of naïve and memory human B cells as a mechanism for enhanced secondary immune responses. *J. Immunol.* **170**, 686–694.

Tough, D. F. & Sprent, J. 1994 Turnover of naïve and memory-phenotype T cells. *J. Exp. Med.* **179**, 1127–1135. (doi:10.1084/jem.179.4.1127.)