How are estimated cellular turnover rates influenced by the dynamics of a source population?

Arpit C. Swain[1,2,3,*], José A. M. Borghans[2], Rob J. de Boer[1]

[1]Theoretical Biology, Utrecht University, Utrecht, The Netherlands

[2]Center for Translational Immunology, University Medical Center Utrecht, Utrecht, The Netherlands

[3]Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY, United States

[*]Address correspondence to: swainarpit@gmail.com

Abstract

Estimating production and loss rates of cell populations is essential but difficult. The current state-of-the-art method to estimate these rates involves mathematical modelling of deuterium labelling experiments. Current models typically assume that the labelling in the precursors of the population of interest (POI) is proportional to the deuterium enrichment in body water. This assumption is not always true and could have a significant effect on the rates estimated from labelling experiments. Here we quantify the effect different turnover rates of the precursors could have on the estimated proliferation and loss rates of a POI by explicitly modelling the dynamics of the precursors. We find that the labelling curve of the POI only reflects its own turnover rate if either the turnover rate of the precursors is sufficiently fast, or the contribution from the precursors is sufficiently small. We describe three other realistic scenarios where the labelling of the POI depends on both its own (proliferation and loss) rates and on the turnover rate of its precursors. Uniquely identifying the turnover rate of a POI, therefore, requires measurements (or knowledge) on the contribution of label from its precursors.

Introduction

The estimation of the average production and loss rates of a cell population, and therefore of the expected lifespan of its cells, has remained challenging. Most estimation procedures either interfere with the true dynamics of the populations (as in adoptive transfer experiments or in-vitro cultures) or are toxic for the cells (for example during in-vivo labelling with compounds such as BrdU or radioactive diisopropylfluorophosphate ($DF^{32}P$)) [1,2]. In recent decades, naturally occurring heavy isotopes (such as ^{2}H, ^{13}C) have been proposed as alternative labelling compounds that are stable and non-toxic, and that should not alter the behaviour of the cells [3,4].

The current state-of-the-art technique to quantify the dynamics of cell populations involves the use of deuterium (^{2}H) as the labelling agent. Deuterated water or glucose is ingested by human volunteers or mice and its incorporation in the DNA of a population of interest (POI) is followed over time. The kinetics of label uptake (upon labelling) and loss (after labelling is stopped) from the DNA of the cells of interest reveals information about the turnover of the POI, which is deciphered using mathematical models [4,5]. These models often simplify all cellular processes into just gain and loss of cells in the POI [5].

It is common practice, in studies addressing maintenance mechanisms of the POI, to distinguish between the gain of cells in the population due to i) cell division and ii) maturation from a precursor population. These are typically considered as two separate processes and are modelled as two independent parameters. Typically, the gain of label from the precursor population has been modelled implicitly, with the assumption that the enrichment of the precursor population mimics that of the body water, i.e., the precursors were assumed to turn over rapidly. In such a case, the rate at which the POI gains label is dictated by its own turnover rate [6]. Here, we show that the labelling dynamics of the source (precursor population) can have a significant effect on the labelling curve of the POI. Depending on the turnover rates of the precursors and the POI, the labelling curve of the POI could reflect its own rates, its precursors' rates, or their combination. Thus, to uniquely identify the kinetics of a POI that is (partly) maintained by a source, we propose to always consider (through prior knowledge or additional measurements) the labelling curve of its immediate precursors.

Implicit source (IS) model

Any population is maintained due to gain of new cells (through a source from a precursor or through proliferation) and loss of existing cells (through differentiation into another population, migration or death). Consider a population, N, which is at equilibrium (denoted by $\bar{N}$), has a source of σ cells/day, a *per capita* proliferation rate, p/day and a *per capita* loss rate, d/day (equation 1a). The kinetics of the number of cells, N, the number of labelled DNA fragments, L, and fraction of labelled DNA, l, within this population can be described as:

$$\frac{dN(t)}{dt} = \sigma + (p - d)N(t) \tag{1a}$$

$$\frac{dL(t)}{dt} = \sigma D(t) + pD(t)\bar{N} - dL(t) \tag{1b}$$

$$\frac{dl(t)}{dt} = d(D(t) - l(t)) \tag{1c}$$

where

$$\bar{N} = \frac{\sigma}{(d - p)} \tag{1d}$$

2

$$D(t) = cU(t) = \begin{cases} 1, & t < 1 \\ 0, & t \geq 1 \end{cases} \qquad (1e)$$

This model suggests that the fraction of the population that is labelled can be accurately tracked if just the loss rate, d, of the population is known (equation 1c). The turnover rate, d, is typically identifiable, as the number of parameters to be estimated reduces from 3 to only 1, when dealing with the labelled fraction of a population at steady state (equation 1a vs equation 1c). Importantly, the precursor population (source) is assumed to label and de-label instantaneously (equation 1b), i.e., the DNA of the cells arriving from the precursors (σ) is assumed to be maximally labelled.

In the model, $D(t)$ is a function denoting the probability that deuterium is incorporated into newly synthesized deoxyribose molecules ($U(t)$ is the level of deuterated water or glucose in plasma (or urine), and c is the intracellular amplification or dilution factor [7]). For simplicity, we treat $D(t)$ as a square pulse where the maximum is scaled to 1 so that the labelled fraction approaches 1 ($l(t) \rightarrow 1$) for very long labelling experiments ($t \rightarrow \infty$). We also scale time such that the labelling period (typically denoted as τ days) is 1 time unit. Scaling time scales the rates of the population. For example, if the memory T-cell pool (i.e., the POI) is labelled for 28 days in mice and has been estimated to have a turnover rate of 0.02/day, the labelling period becomes 1 (scaled unit of time, reflecting about a month) and the turnover rate of the POI becomes $0.02 \times 28 = 0.56$ per scaled unit of time (i.e., approximately per month). This scaling simplifies the presentation of the results considerably.

Explicit source (ES) model

While the previous model assumed that the precursor population labels and de-labels instantaneously, in fact, the dynamics of a source can have a major impact. This was illustrated by a study on the dynamics of neutrophils. The lifespan of blood neutrophils was estimated to be ~5 days when the precursors of neutrophils (in the bone marrow) were assumed to have a fast turnover, and ~0.5 day when the precursors were assumed to be slow, underlining the importance of explicitly considering the dynamics of the source [4,8].

To quantify the effect of the source's dynamics on the rate estimates of the POI, consider an experiment designed to find out the rates of a POI, N_2, that has an unobserved precursor population, N_1. The cells of the precursors, N_1, have a source of σ cells/scaled time unit (stu), from their immediate precursors, they divide at a rate of p_1/stu and are lost at a rate of d_1/stu. A fraction α of the cells leaving the precursor population mature into cells of the POI (i.e., the remaining $(1 - \alpha)$ fraction die or move elsewhere). Thus, the precursors obey the standard IS model (equation 1a). If the cells of the POI divide at a rate of p_2/stu, and are lost at a rate of d_2/stu, the kinetics of the precursors and the POI are written as:

$$\frac{dN_1(t)}{dt} = \sigma + (p_1 - d_1)N_1(t) \tag{2a}$$

$$\frac{dN_2(t)}{dt} = k\alpha d_1 N_1(t) + (p_2 - d_2)N_2(t) \tag{2b}$$

with,
$$\overline{N_1} = \frac{\sigma}{(d_1 - p_1)} \tag{2c}$$

$$\overline{N_2} = \frac{k\alpha d_1 \overline{N_1}}{(d_2 - p_2)} \tag{2d}$$

100

101 where $0 < \alpha \le 1$ and, $\overline{N_1}$ and $\overline{N_2}$ are the steady state values of the precursors and the POI,
102 respectively. If the differentiation of the precursors into the POI is accompanied by one cell
103 division, the parameter $k = 2$, otherwise $k = 1$. See the companion article (*Yan et al.,*
104 submitted) for a discussion on larger values of k. The POI is primarily maintained by the source
105 whenever $k\alpha d_1\overline{N_1} > p_2\overline{N_2}$ in equation 2b, which simplifies to the quite natural condition $p_2 <$
106 $d_2/2$. For simplicity, we no longer mention the units (stu) of the parameters as they are all
107 scaled with respect to the labelling period.

108 The number of labelled DNA strands in the precursors, L_1, and in the POI, L_2, in the ES model
109 obey:

$$\frac{dL_1(t)}{dt} = \sigma D(t) + p_1 D(t)\overline{N_1} - d_1 L_1(t) \tag{2e}$$

$$\frac{dL_2(t)}{dt} = \alpha d_1 L_1(t) + (k-1)\alpha d_1 \overline{N_1} D(t) + p_2 D(t)\overline{N_2} - d_2 L_2(t) \tag{2f}$$

110

111 Thus, if differentiation is accompanied by cell division, the POI gains label due to the inflow
112 of labelled DNA strands from the precursors as well as due to the generation of de-novo DNA
113 strands on replication of the labelled and unlabelled DNA strands. The kinetics of the fractions
114 of labelled DNA strands in the precursors, l_1, and the POI, l_2, then follow:

$$\frac{dl_1(t)}{dt} = d_1(D(t) - l_1(t)) \tag{2g}$$

$$\frac{dl_2(t)}{dt} = \frac{(d_2 - p_2)}{k}\left(l_1(t) + (k-1)D(t)\right) + p_2 D(t) - d_2 l_2(t) \tag{2h}$$

$$= \frac{d_2}{k}\left(l_1(t) + (k-1)D(t) - kl_2(t)\right) + \frac{p_2}{k}\left(D(t) - l_1(t)\right) \tag{2i}$$

115

116 Transforming the system from cell numbers into fractions reduces the system to a 3-parameter
117 model: d_1, p_2 and d_2 (equation 2i). Note that after a long labelling period (or in rapidly turning
118 over populations), $l_1(t) \to D(t) = 1$, and that regardless of the value of k, $l_2(t) \to 1$.

119

The ES model can perfectly describe the dynamics of a kinetically heterogeneous population

122 We generalized the IS model above by simply introducing a source population in the ES model
123 which allows us to model either a slow or a fast source population. The POI in the ES model,
124 however, is still modelled as a homogeneous population. Yet, the ES model can describe a
125 kinetically heterogeneous population perfectly because the kinetic heterogeneity model [6,9]
126 turns out to be a special case of the ES model. To show this, we compare the solutions of the
127 kinetically heterogeneous model (equation 3) and the ES model (equation 4).

128 The solution of the kinetically heterogeneous model [9] is:

$$l(t) = \beta\left(1 - e^{-\delta_1 t}\right) + (1 - \beta)\left(1 - e^{-\delta_2 t}\right) \tag{3}$$

129

130 where δ_1 and δ_2 are the turnover rates of the first (slow) and the second (fast) sub-populations,
131 respectively, because we define $\beta \in [0,1]$ for the fraction of the population that is made up of
132 the slow sub-population (without loss of generality).

133 During the labelling phase, the solution of the ES model is:

$$l_1(t) = 1 - e^{-d_1 t} \tag{4a}$$

$$l_2(t) = 1 - \frac{(d_2 - p_2)}{(d_2 - d_1)} e^{-d_1 t} - \frac{(p_2 - d_1)}{(d_2 - d_1)} e^{-d_2 t} \tag{4b}$$

$$= \alpha(1 - e^{-d_1 t}) + (1 - \alpha)(1 - e^{-d_2 t}) \tag{4c}$$

134

135 where $\alpha = \frac{(d_2 - p_2)}{(d_2 - d_1)} \in (-\infty, +\infty)$. We, here, consider the $k = 1$ case of the ES model because
136 the $k = 1$ case is the more general model and includes the $k = 2$ case (see the sub-section
137 *division-linked differentiation* below). The $k > 2$ case is studied in the companion paper (*Yan*
138 *et al.*, submitted).

139 The solutions of the POI in the ES model (equation 4c) and the kinetically heterogeneous model
140 (equation 3) are identical. As $\alpha \in (-\infty, +\infty)$ has a wider range than $\beta \in [0,1]$, the ES model
141 can perfectly describe any kinetically heterogeneous population but the converse is not always
142 true. Further, if $k = 2$ and $p_1 > 0$, the ES model can also represent a population of stem cells,
143 N_1, dividing asymmetrically into themselves (N_1) and into the following population (N_2) (see
144 SI). Thus, the ES model is the more general model and can provide alternate explanations of
145 many other mechanistic models.

5

146   Although this equality makes it difficult to recognize the true mechanism underlying a labelling
147   curve, the ranges of α and β help us identify scenarios where the mechanism is unambiguous.
148   To find scenarios where the dynamics of the ES model is identical to that of the kinetic
149   heterogeneity model, we compare the relationship between the parameters of the ES and the
150   kinetic heterogeneity models. As $\beta > 0$, we only consider $\alpha > 0$, meaning that $d_2 > d_1$, i.e.,
151   that the POI is faster than its source population. Comparing equations 3 and 4c, we find that
152    $\alpha = \beta$, and

$$\alpha = \frac{(d_2 - p_2)}{(d_2 - d_1)}$$

$$\Rightarrow p_2 = \alpha d_1 + (1 - \alpha)d_2 = \beta \delta_1 + (1 - \beta)\delta_2 = \bar{\delta} \tag{5}$$

153

154   when $d_1 = \delta_1$ and $d_2 = \delta_2$. That is, α in the ES model gives the fraction of the kinetically
155   heterogeneous population that is made up of the slow sub-population, and the proliferation rate
156   of the POI, p_2, in the ES model is the same as the average turnover rate, $\bar{\delta}$, estimated for the
157   kinetically heterogeneous population. For example, deuterium labelling of CD8$^+$ memory T
158   cells [10] can be described equally well with either the kinetic heterogeneity model (as was
159   shown by Westera et al. [10]) or the ES model. See the sub-section discussing **the case study**
160   **on Westera et al. (2013)** below for a numerical example confirming these equalities.

161   As the POI in both the models are made up of two exponentials, the turnover rates of the source
162   and the POI in the ES model have to be the same as those of the slow and fast sub-populations
163   in the kinetic heterogeneity model, respectively. Therefore, based on its deuterium labelling
164   dynamics, a kinetically homogeneous population with a source cannot be distinguished from a
165   kinetically heterogeneous population whenever the POI is faster than its source population (i.e.,
166   when $d_2 > d_1$, meaning $\alpha > 0$). In cases where $d_2 < d_1$, one can, therefore, be certain that the
167   population is not kinetically heterogeneous.

168

The labelling curve of the POI reflects its own or its precursor's turnover rates in special cases

169
170

171   In the ES model, the POI can have a source (of potentially recently divided cells), can divide,
172   and has a loss rate. Trivially, the labelling curve of the ES model is defined by the turnover rate
173   of the POI, d_2, if the source into the POI is negligible, i.e., if $p_2 \to d_2$ (in equation 2i).
174   Therefore, in this section, we discuss the label gain and loss rates in two special cases of the
175   ES model where the POI does have a significant source from the precursors: 1) a rapidly turning
176   over precursor, and 2) a rapidly turning over, non-dividing POI. Throughout this article, we
177   use label gain rate (and label loss rate) to refer to the rate at which a population gains (and
178   loses) labelled DNA.

179   If the precursors turn over rapidly, e.g., $d_1 \geq 10$, the unlabelled cells in the precursor
180   population would be rapidly replaced by labelled cells. The precursors can, in such a case, be

approximated by the deuterium availability in plasma, i.e., $l_1(t) \approx D(t)$ (equation 2g). The labelling in the POI (equation 2i) will then boil down to:

$$\frac{dl_2(t)}{dt} \approx d_2(D(t) - l_2(t)) \tag{6}$$

which is the familiar IS model, showing that the enrichment of the POI is determined by just the turnover rate of the POI, d_2. Thus, the gain and loss rates estimated from labelling experiments where the precursors turn over rapidly (relative to the labelling period) indeed reflect the turnover rate of the POI [6]. For example, as rapidly dividing thymocytes are the precursors of slowly dividing naive T cells, the estimated lifespan should correctly reflect the true lifespan of naive T cells [11].

If the only source of label into the POI is due to the flow of labelled DNA from the precursors (i.e., $k = 1$ and $p_2 = 0$), the labelling dynamics in the POI (equation 2i) obeys:

$$\frac{dl_2(t)}{dt} = d_2(l_1(t) - l_2(t)) \tag{7}$$

If the cells of the POI turn over rapidly (e.g., $d_2 \gg 1$), the enrichment of the POI mimics that of the precursors (equation 7). Therefore, measuring only the POI will, in fact, reveal the turnover rate of the precursors. The neutrophils are a prime example of this case, as it is unlikely that the slow labelling rate of mature neutrophils in blood reflects their own rate. Their rate of label accrual probably reflects the slow turnover of their precursors in the bone marrow [4,8]. Note that when $k > 2$, most precursors become labelled upon arriving in the POI (see the companion paper by *Yan et al.* (submitted)). These special cases showcase that the labelling curve of the POI sometimes reflects the turnover rate of the POI and sometimes that of its precursors.

The label gain and loss rates of the POI do not equal its turnover rate if the POI and the precursors have comparable turnover rates

Above we estimated the label gain and loss rates in a few extreme cases where either the source into the POI is negligible, or the precursors or the POI are very short-lived (i.e., the precursors or the POI become labelled very rapidly). In many labelling experiments, the duration of the labelling period is comparable to the POI's lifespan, to avoid scenarios where the POI either hardly labels or becomes labelled very rapidly during the labelling period. This is important as an accurate estimation of the POI's rates requires the collection of enough informative data points during the labelling period. As estimating the label gain and loss rates in these cases is not straight-forward, we derive first-order approximations for the POI's label gain and loss rates.

Denoting the label gain and loss rates as $p^*(t)$ and $d^*(t)$, respectively, the labelling curve of a POI can be written as (from equation 1c):

$$\frac{dl(t)}{dt} = p^*(t)D(t) - d^*(t)l(t) \tag{8}$$

The label gain rate reflects the dynamics of the entire population [5], as de-novo labelled DNA molecules are made if a cell (with either labelled or unlabelled DNA molecules) divides during the labelling phase. The label loss rate, however, reflects the loss rate of the labelled sub-population only [5], as the POI loses label only when a cell with labelled DNA is replaced by a cell with unlabelled DNA.

In the ES model introduced above, the fraction labelled DNA in the precursors is determined only by its loss rate, d_1, whereas the fraction labelled DNA in the POI is dependent not just on its own rates but also on the rates of its precursor population (equation 2h). The rate of label gain in the POI, $p^*(t)$, during the labelling phase ($t \leq 1$, which means $D(t) = 1$) is:

$$p^*(t) = \frac{dl_2(t)}{dt} = \frac{(d_2 - p_2)}{k}(l_1(t) + k - 1) + p_2 - d_2 l_2(t) \tag{9a}$$

The rate of label gain depends on both the precursor and the POI, as $p^*(t)$ also depends on $l_1(t)$. The initial label gain rate of the POI (i.e., when $t = 0$) is:

$$p^*(0) = \frac{d_2(k - 1) + p_2}{k} \tag{9b}$$

To calculate the rate of label gain later during the labelling phase, the expression of label gain rate (equation 9a) needs to be further simplified. If the label gain can be approximated well by a linear increase, the expression for the gain rate (equation 9a) can be simplified by noticing that $l_1(t) = 1 - e^{-d_1 t} \approx d_1 t$, and $l_2(t) \approx p^*(t)t$. During the labelling phase ($t < 1$), the rate of label gain in the POI can then be approximated by:

$$p^*(t) \approx \frac{d_2(d_1 t + k - 1) + p_2(1 - d_1 t)}{k(1 + d_2 t)} \tag{9c}$$

The approximate rate of label gain at the end of the labelling phase (when $t \to 1$) of the POI, then, becomes

$$p^*(1) \approx \frac{d_2(d_1 + k - 1) + p_2(1 - d_1)}{k(1 + d_2)} \tag{9d}$$

8

239 If we are in a regime where the gain of label in the POI can be approximated well by a straight
240 line, these expressions become similar, i.e., $p^*(t) \approx p^*(0) \approx p^*(1)$ (see Figures 1 and 2, and
241 Table 1 below).

242 Additionally, the above expressions can be used to determine the behaviour of the precursors
243 relative to the POI. If the labelling curves of both the precursors and the POI can be
244 approximated reasonably by a straight line, the population with the faster labelling dynamics
245 can be identified just by comparing the initial gain rate of these populations. Hence, the gain
246 in label in the POI (during the labelling phase) is faster than that in its precursor if

$$l_2(t) > l_1(t) \Leftrightarrow p^*(0)t > d_1 t \Leftrightarrow p^*(0) > d_1 \tag{10}$$

247

248 As opposed to the rate of label gain, $p^*(t) = \frac{dl_2(t)}{dt}$, which is defined on the total population,

249 the label loss rate, $d^*(t) = \frac{1}{l_2(t)}\frac{dl_2(t)}{dt}$, is defined on the labelled fraction (see equation 8). So,

250 the rate of label loss in the POI, $d^*(t)$, in the de-labelling phase (i.e., when $D(t) = 0$ in

251 equation 2h) is:

$$d^*(t) = -\frac{\frac{(d_2 - p_2)}{k}l_1(t) - d_2 l_2(t)}{l_2(t)} \tag{11a}$$

$$= d_2 - \frac{(d_2 - p_2)}{k}\frac{l_1(t)}{l_2(t)} \tag{11b}$$

252

253 Again, as $l_1(t) \approx d_1 t$ and $l_2(t) \approx p^*(t)t$, the above expression simplifies to

$$d^*(t) \approx d_2 - \frac{(d_2 - p_2)}{k}\frac{d_1}{p^*(t)} \tag{11c}$$

254

255 It is important to keep in mind that the approximated downslope is accurate for only a short
256 period of time in the de-labelling phase as the approximations of $l_1(t)$ and $l_2(t)$ are not
257 accurate for $t > 1$.

258 As $p^*(0) = 0$ in some cases, for example when $p_2 = 0$ and $k = 1$, we use the approximation
259 $p^*(t) \approx p^*(1)$ for defining $d^*(t)$. Thus, the initial rate at which label is lost in the beginning
260 of the de-labelling phase is

$$d^*(1 + \varepsilon) \approx d_2 - \frac{(d_2 - p_2)}{k}\frac{d_1}{p^*(1)} \tag{11d}$$

261

262 where $\varepsilon > 0$ is small.

263   The approximations of the gain and loss rates of label in the POI show that these rates are not
264   dictated only by the turnover rate of the POI but also by the division rate of the POI and the
265   turnover rate of the precursors. Next, we use the above simplified expressions for the gain and
266   loss rate of label in the POI to find qualitative relationships between the estimated and true
267   label gain or loss rates.

268

The rate of label uptake in the POI is generally lower than its true turnover rate

271   Models that consider an implicit source (like the IS model above in equation 1 or the implicit
272   kinetic heterogeneity model [5]) predict that, in a population that is at steady state, the rate of
273   label gain in the POI, $p^*(t)$, represents the average turnover rate of the POI (d in the IS model
274   and p^* in the implicit kinetic heterogeneity model [5]). To test the validity of this prediction,
275   we graphically compare the label gain and loss rates of the POI found from the analytical
276   predictions (equations 9b-d and 11c) and from numerical estimations, using the implicit kinetic
277   heterogeneity model (equation 8 with p^* and d^* instead of $p^*(t)$ and $d^*(t)$, respectively) with
278   actual labelling curves of the ES model (Figures 1, 2 and S1, Table 1) in different scenarios.

279

Case I: <u>When the turnover rates of the precursors and the POI are comparable</u>

281   We first consider scenarios where the lifespans of the precursors and the POI are comparable
282   to the duration of the labelling period (e.g., $0.1 \leq d_1 < 1$, $0.1 \leq d_2 < 1$), for example,
283   labelling studies that track memory T cells, which might have a source from less-differentiated
284   memory T-cell phenotypes, in blood [12]. The numerical simulations explore examples where
285   the precursors and the POI have comparable lifespans, e.g., the considered lifespans are either
286   2-fold or 5-fold longer than the duration of the labelling period (Figures 1 and 2). Further, we
287   show cases where the POI may (left column in figures, $p_2 > 0$) or may not (right column in
288   figures, $p_2 = 0$) divide.

289

Precursors are faster than the POI. If the precursors turn over faster than the POI, i.e., $d_1 >$
291    $d_2 > p_2$, the labelling in the POI, during the labelling phase, cannot be higher than that in the
292   precursors ($p^*(0) < d_1$ in equation 10, Figure 1).

293   If the differentiation of the precursors into the POI is not accompanied by division (i.e., if $k =$
294   1), the initial rate of label gain in the POI, $p^*(0)$, is given by its division rate, p_2 (equation 9b,
295   Figures 1a-b, Table 1). As the predicted initial label gain rate, $p^*(0) = p_2$, can have any value
296   between 0 and d_2 (Figures 1a-b), the initial slope is not the average turnover rate of the POI,
297   as was previously assumed. If the cells in the POI do not divide (i.e., if the gain of label in the
298   POI is only due to the influx of labelled precursors), the labelling curve of the POI has an initial

299 delay during which the gain of label in the POI is zero (Figure 1b). This initial delay in the gain
300 of label is given by $p^*(0) = p_2 = 0$, which is infinitely smaller than the expected d_2.

301 After the initial phase, the labelling curve is best described by $p^*(1)$, which depends on all
302 three parameters of the model (equation 9d). The slope, $p^*(t)$, increases over time, starting at
303 the POI's division rate, p_2, up to $p^*(1)$, where the gain rate of label in the POI is higher than
304 its own division rate, and is approaching its turnover rate, d_2 (Table 1). The slope of the
305 labelling curve approaches the turnover rate of the POI as cells of the POI that are lost are
306 replaced by labelled cells from the higher enriched precursors. Numerical simulations confirm
307 that the predicted slopes describe the true labelling curve faithfully (Figures 1a-b) and are
308 similar to each other if the POI is a dividing population (Figure 1a).

309 If the differentiation of precursors into the POI is accompanied by cell division, the initial rate
310 of label gain in the POI cannot be zero (i.e., if $k = 2$ in equation 9b, Figures 1c-d). Like before
311 (the case when $k = 1$), the approximated gain rate increases from the initial $p^*(0)$ to $p^*(1)$
312 due to the higher enrichment of the incoming precursors. The label gain rate of the POI is closer
313 (relative to the case when $k = 1$) to, but still smaller than, its true turnover rate (Figures 1c-d).
314 Thus, if the precursors are faster than the POI, the label gain rate in the POI increases over the
315 labelling period due to an influx of precursors into the POI, provided $d_1, d_2 < 1$ and regardless
316 of whether differentiation is accompanied by division (Table 1).

Condition	d_1	d_2	p_2	k	$p^*(0)$	$p^*(1)$	$d^*(1 + \epsilon)$	p^*	d^*
Faster precursors	0.5	0.2	0.1	1	0.10	0.13	-0.20	0.12	0.04
				2	0.15	0.15	0.03	0.16	0.13
			0	1	0	0.08	-1	-	-
				2	0.10	0.13	-0.20	0.12	0.04
Slower precursors	0.2	0.5	0.25	1	0.25	0.20	0.25	0.24	0.27
				2	0.38	0.27	0.41	0.37	0.40
			0	1	0	0.07	-1	-	-
				2	0.25	0.20	0.25	0.24	0.27

317

318 **Table 1:** The true proliferation and turnover rate (d_1, d_2, and p_2), and the predicted label gain
319 and loss rates (i.e., $p^*(0)$, $p^*(1)$, and $d^*(1 + \epsilon)$ using equations 9b-d, 11c) corresponding to
320 the simulations shown in Figures 1 and 2. The parameters (p^* and d^*) of the best fits (shown
321 in Figures S2 and S3) of the phenomenological model (equation 8) to the labelling data of the
322 POI. The estimated label gain rate initially (see $p^*(0)$) reflects the proliferation rate of the POI
323 (when $k = 1$) and then changes to reflect the influence of the precursor's rate (captured by
324 $p^*(1)$). The estimates of p^* and d^* are not reported for the cases in which the phenomenological
325 model fails to describe the data (see Figures S2 and S3). Note that the estimated loss rate,
326 $d^*(1 + \epsilon)$, gives the rate at which the labelled population loses labelled DNA soon after the
327 end of the labelling period (ϵ is a very small positive number). The negative sign indicates that
328 the population, in fact, gains labelled DNA for a short period after the end of labelling (see
329 equation 8). The rates (expressed as stu) are scaled with respect to the labelling period.

330

331   ***Precursors are slower than the POI***. If the POI is faster than the precursors, i.e., $d_2 > d_1$, the
332   labelling in the POI can be either faster or slower than the labelling in the precursors (equation
333   10, Figures 2a-b, Table 1). The labelling of the POI is faster than that of the precursors if the
334   division rate of the POI is faster than the turnover rate of the precursors, i.e., $p_2 > d_1$. If the
335   cells in the POI do not divide, the labelling in the POI can be faster than that in the precursors
336   only if the precursors go through division-linked differentiation and the POI is at least 2-fold
337   faster than the precursors, $d_2 > 2d_1$ (equation 10; also see equations 14 and 15).

338   Some properties stay unchanged compared to the case where precursors are faster than the POI.
339   For example, if the precursors do not divide while differentiating into the POI ($k = 1$), the
340   initial rate of label uptake reflects the division rate of the POI, $p^*(0) = p_2$ (equation 9b). This
341   implies that a non-dividing POI would initially have a zero labelling rate (Figure 2b), which
342   should increase over time to reflect the turnover rate of the precursor population. If the
343   precursors go through division-linked differentiation ($k = 2$), the initial rate of label gain in
344   the POI reflects both its division and death rates (equation 9b). The approximated slopes ($p^*(0)$
345   and $p^*(1)$) are, in all cases, markedly lower than the true turnover rate of the POI, d_2. The
346   slope declines over time during the labelling period, starting at a value defined by its own
347   (division and loss) rates and moving towards the turnover rate of the precursors. The low
348   enrichment of the precursors dilutes the enrichment in the POI, causing a decline in the slope
349   of the labelling curve. Therefore, the POI's label gain rate initially reflects its own division and
350   death rates, and is only influenced by the rates of its precursors later in the labelling period.

351

352   Finally, to confirm these analytical results, we defined two constant parameters p^* and d^*
353   instead of $p^*(t)$ and $d^*(t)$, respectively, in the phenomenological implicit kinetic
354   heterogeneity model (equation 8) to numerically estimate the label gain and loss rates of the
355   POI in all cases (i.e., whether the precursors are faster or slower than the POI) by non-linear
356   parameter fitting. In all cases where the model was able to describe the artificial data, the
357   estimated label gain rate, p^*, lay between the analytically predicted initial and final label gain
358   rates ($p^*(0)$ and $p^*(1)$, respectively) (Table 1). The estimated label loss rate, d^*, and the
359   predicted initial label loss rate, $d^*(1 + \epsilon)$, however, differed considerably. As the estimated
360   label loss rate, d^*, is based upon the labelling dynamics of the POI during the entire de-labelling
361   phase, it need not be similar to the initial gain or loss of label after the end of the labelling
362   period (i.e., $d^*(1 + \epsilon)$, Table 1). Note that the phenomenological model (equation 8) is unable
363   to describe the labelling curve if the precursor population differentiates without division into a
364   non-dividing POI (see Table 1, Figures S2 and S3).

365

366   **Case II: <u>The turnover rates of the precursors and the POI are not comparable</u>**

367   If the POI is much faster than the precursors, the label gain rate approaches the turnover rate
368   of the precursors, which was highlighted above as a well-known special case [4,8]. Note that
369   the approximated slopes describe the labelling curve for a short period of time, as linear
370   approximations do not give good descriptions of $l_1(t)$ and $l_2(t)$ when $d_1, d_2 > 1$ (see Figure

S1 and Table S1). Similarly, if the precursors are faster than the POI, the label gain rate in the POI approaches its own turnover rate as the label gain rate in the POI cannot be higher than its turnover rate, which was also reported above as a special case. Therefore, the approximated rates give a good description of the gain and loss of label in a population.
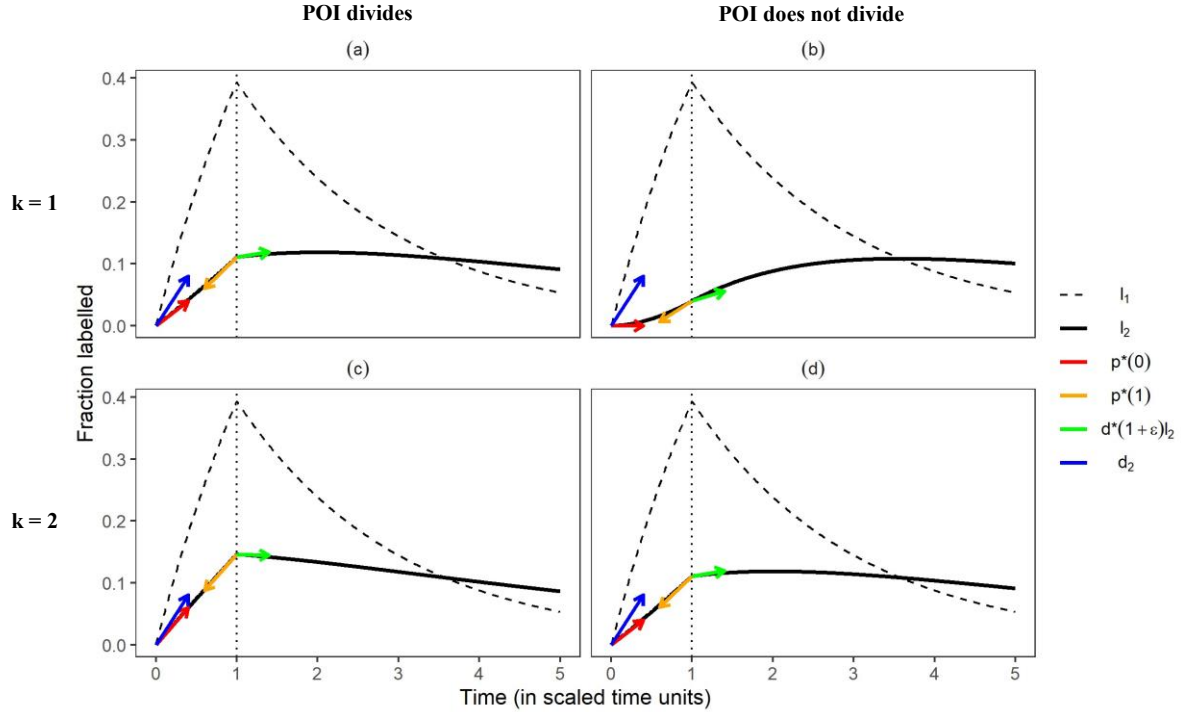
Figure 1: Precursors are faster than the POI. The rate of label gain is lower than the POI's turnover rate, d_2 (blue vector), if the precursors turn over faster than the POI. The coloured vectors represent the slopes identified in the legend and the arrows give the direction of extrapolation of these slopes. Note that the slopes are calculated as the label gain or loss rate multiplied by the fraction of unlabelled or labelled cells, respectively, for example, $d^*(1 + \epsilon)l_2$. The loss rate of the precursors, d_1, and the POI, d_2, were set to 0.5 and 0.2, respectively. The division rate of the POI, p_2, was either set to 0.1 (in (a) and (c)) or 0 (in (b) and (d)). Note that the dynamics in (a) and (d) are identical (see the sub-section ***Division-linked differentiation*** below for the explanation).

13

POI divides   **POI does not divide**
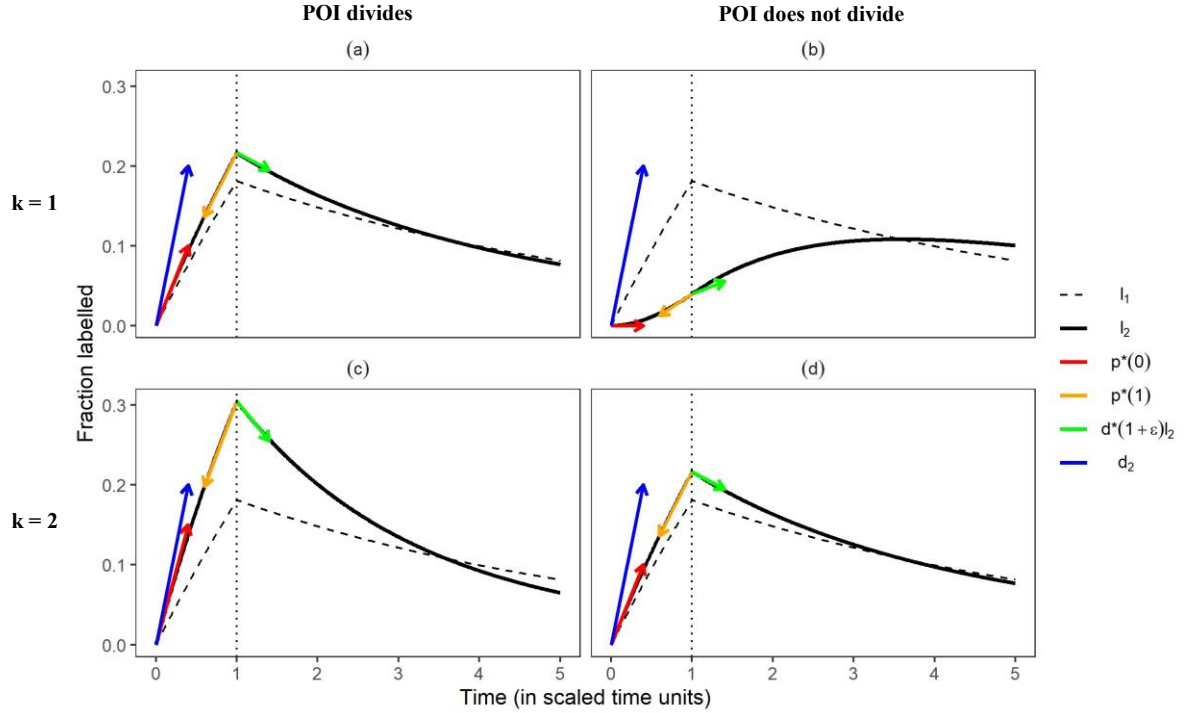
Figure 2: Precursors are slower than the POI. The rate of label gain is also lower than the POI's true turnover rate, d_2 (blue vector), if the precursors turn over slower than the POI. The coloured vectors represent the slopes identified in the legend and the arrows give the direction of extrapolation of these slopes. Note that the slopes are calculated as the label gain or loss rate multiplied by the fraction of unlabelled or labelled cells, respectively, for example, $d^*(1+\epsilon)l_2$. The loss rate of the precursors, d_1, and the POI, d_2, were set to 0.2 and 0.5, respectively. The division rate of the POI, p_2, was either set to 0.1 (in (a) and (c)) or 0 (in (b) and (d)). Note that the dynamics in (a) and (d) are identical, which can also be seen in Table 1 (see the sub-section ***Division-linked differentiation*** below for details).

The label gain rate can maximally be the turnover rate of the POI. The examples above showed that the rate at which the POI gains label was invariably lower than the POI's turnover rate. Here we show that this is always the case, provided that the lifespans of the POI and the precursors are comparable to the duration of the labelling period.

If the precursors are not fast, i.e., if $d_1 < 1$ (and hence, $l_1(1) < 1$), both the slopes, $p^*(0)$ and $p^*(1)$, are smaller than the expected value, d_2. Defining $p_2 = \alpha_2 d_2$, where $0 \leq \alpha_2 < 1$, from equation 9b we can obtain:

$$p^*(0) = \frac{d_2\big(k - (1-\alpha_2)\big)}{k} < d_2 \qquad (12a)$$

And from equation 9d, we can obtain:

14

$$p^*(1) \approx \frac{d_2\big(k + (d_1 - 1)(1 - \alpha_2)\big)}{k(1 + d_2)} < d_2 \tag{12b}$$

because $\frac{\big(k + (d_1 - 1)(1 - \alpha_2)\big)}{k(1 + d_2)} < 1$, when $d_2 > \frac{(d_1 - 1)(1 - \alpha_2)}{k}$, which always holds true as $d_1 < 1$. Thus, the rate at which the POI gets labelled is always lower than its turnover rate, d_2, provided $d_1, d_2 < 1$. Interpreting the label gain rate as the true turnover rate would, therefore, overestimate the POI's lifespan whenever the precursors are not very fast and/or play a significant role in the maintenance of the POI.

The initial label loss in the POI is accurately predicted

We have shown that the calculated label gain rates predict the enrichment in the precursors and the POI accurately in most cases. As the relative enrichment of populations at the end of the labelling period influences the rate at which label is lost (equation 11b, Swain et al., manuscript in preparation), the initial rate at which label is lost, $d^*(1)$, can be approximated well by using the estimated gain rates (equation 11c).

If the lifespans of the POI and the precursors are comparable to the labelling period, the estimated gain and loss rates give a good description of the labelling curves (see the red and green arrows in Figures 1 and 2). The estimated loss rates predict whether the POI start to lose or continue to accrue label at the end of the labelling phase (Table 1). If the precursors are more enriched than the POI, the labelling of the POI continues after the labelling period in most cases (4 out of 5 cases, see Figures 1a,b,d and Figure 2b). In these cases, the gain of label in the POI is primarily due to the inflow of highly labelled precursor cells (also see the subsection ***Timing of the peak***). In the other four cases (see Figure 1c and Figures 2a,c,d), the label gain in the POI was driven by proliferation.

Note that even if the cells in the POI are very short-lived, the estimates of the initial loss rate accurately informed on whether the POI gained or lost label after the labelling period (Table S1). So, the estimated rates do provide a faithful description of the de-labelling curve, but only for a very short period of time (Figure S1). Thus, the predicted label gain rate, $p^*(1)$, can be used to find a good approximation of the initial loss rate (equation 11c).

The POI can have four different labelling behaviours

Unlike the IS model where the labelling of the POI is determined only by the death rate of the POI (equation 1c), in the ES model the POI can have four qualitatively different slopes at which the POI gains and loses label (summarized in Table 2). The labelling of the POI is determined by:

438   1) its own turnover rate, d_2, if the precursors turn over rapidly. In this case, the labelling
439 in the precursors approximates the label availability in the plasma. The ES model approximates
440 the IS model, and the rates of labelling and de-labelling are primarily driven by the POI's loss
441 rate. An example of this case are naive T cells, that have the rapidly dividing thymocytes as
442 their precursors [11]. Note that if $k \gg 2$, the ES model approximates the IS model (see the
443 companion paper by *Yan et al.* (submitted)).

444   2) the turnover rate of the precursors, d_1, if the POI is fast. An example would be mature
445 neutrophils in blood, which are thought to mimic the labelling curve of their precursors in the
446 bone marrow [4,8]. The POI, here mature neutrophils, are probably much faster than their
447 precursors in the bone marrow. They attain the enrichment of the precursors as they are
448 replaced (by the cells flowing in from the precursors, post-mitotic pool) much faster than the
449 change in the enrichment of the precursors.

450   3) its division rate, p_2, if differentiation is not accompanied by division ($k = 1$). This
451 is because the increase in enrichment due to the division in the POI would initially be the major
452 contributor compared to the label flowing in from the precursors. Note that the precursors and
453 the POI should have lifespans that are comparable to the length of the labelling period. A likely
454 example for such a case would be central memory T cells undergoing homeostatic
455 differentiation into effector memory T cells [13,14].

456   4) the average of its division and death rates, $(p_2 + d_2)/2$, if the precursors go through
457 division-linked differentiation ($k = 2$). Here, along with the label gain due to the division in
458 the POI, the source from the precursors also has a significant contribution of labelled cells as
459 the precursor cells also divide and pick up label when they differentiate into the POI. Thus, if
460 a naive T cell were to undergo division while differentiation into memory T cells [15], the gain
461 of label by memory T cells would reflect both its division and death rates.

462 These four different labelling behaviours can be distinguished from each other if the turnover
463 rates of the precursors is known. Therefore, it is essential to have information on the precursors'
464 dynamics. Since the labelling in the POI depends on the enrichment of its immediate precursors
465 only (equation 2h), it is sufficient to have knowledge or data on the immediate precursors (and
466 not the precursors' precursor) to distinguish the rates of the POI from that of its precursors. So,
467 even a phenomenological description of the precursors' labelling curve, $l_1(t)$, would suffice to
468 correctly estimate the turnover rate of the POI. Therefore, it is essential and sufficient to
469 measure and model the immediate precursor population of the POI to reliably interpret the
470 estimated rates.

Scenarios	$p^*(0)$	$p^*(1)$	$d^*(1 + \varepsilon)$	Constraints
General cases				
Differentiation without division	p_2	$\dfrac{d_2 d_1 + p_2(1 - d_1)}{(1 + d_2)}$	$d_2 - (d_2 - p_2)\dfrac{d_1}{p^*(1)}$	$k = 1$

Division-linked differentiation	$\dfrac{(d_2+p_2)}{2}$	$\dfrac{d_2(d_1+1)+p_2(1-d_1)}{2(1+d_2)}$	$\dfrac{d_2}{2}-\dfrac{(d_2-p_2)}{2}\dfrac{d_1}{p^*(1)}$	$k=2$
Special cases				
Precursors turn over about as fast as deuterated water/glucose **OR** **The source is negligible**	d_2	d_2	d_2	$1 \ll d_1$ OR $d_1 \approx 0$
Rapidly turning over, non-dividing POI with precursors that do not divide while differentiating	d_1	d_1	d_1	$0 < d_1 < d_2$ $1 \ll d_2$ $p_2 = 0$ $k = 1$

Table 2: A POI with a precursor population has 4 qualitatively different gain and loss rate scenarios. Knowledge about the precursor is crucial to identify the turnover rate of the POI.

Numerical tests reveal the range of variations around the identified parameter relationships that is introduced by fitting simple models to data

Up until now, we have identified relationships between the parameters of the ES model (for example, p_2 and d_2) and the parameters of the phenomenological model (p^* and d^*), by analytically comparing equations. Next, we aimed to take the measurement noise in the labelling data and uncertainties introduced by fitting simplified models to that data into account.

We systematically varied the values of the three parameters of the ES model (d_1, p_2 and d_2) to generate a 1000 different parameter sets. Each parameter could have 10 different values: $d_1 = 0.5i, d_2 = 0.5i, p_2 = \alpha_2 d_2, \alpha_2 = 0.1i$, where $i \in \{1, \dots, 10\}$. These 1000 parameter sets were then used to make artificial data using the ES model (having three parameters). This dense and perfect data was subsequently fitted with the implicit kinetic heterogeneity model (equation 8, having two parameters, p^* and d^*). As we are interested in the average turnover rate, the estimated value of p^* was plotted against the true values of the three parameters of the ES model (Figures 3 and 4).

490 To test the analytical prediction of the label gain rate (equation 9b), we also plotted the
491 estimated p^* values against $p^*(0)$, which denotes the initial label gain rate of a population.
492 Note that in the companion paper (*Yan et al.*, submitted) $p^*(0)$ is referred to as the quantity
493 'production by division', which combines potential division-linked differentiation of the
494 precursors with the proliferation within the POI. Production by division is defined as $p_2 +$
495 $(d_2 - p_2)\frac{(k-1)}{k}$ (i.e., the coefficients of $D(t)$ in equation 2h), which is equal to $p^*(0)$.

496

497 ***If differentiation is not accompanied by cell division (i.e., $k = 1$).*** The relationships observed
498 in the numerical simulations (Figure 3) tend to be in line with our analytical results. The
499 estimated label gain rate, p^*, is always lower than the true turnover rate of the population of
500 interest, d_2 (Figure 3b), and is approximately equal to the true proliferation rate of the
501 population of interest, p_2 (Figure 3c). The initial label gain rate, $p^*(0)$, is equal to p_2 when
502 $k = 1$, and hence Figure 3c and 3d are identical. The numerical results do not always match
503 the analytical ones, as the relationship between p^* and p_2 has a wide range for low values of
504 p_2, and differs by more than 2-fold for many parameter sets (the red line indicates the
505 relationships $p^* = 2p_2$ and $p^* = 2p^*(0)$ in Figures 3c and 3d, respectively). This shows that
506 fitting a simplified model to the labelling data can introduce error in the value of the estimated
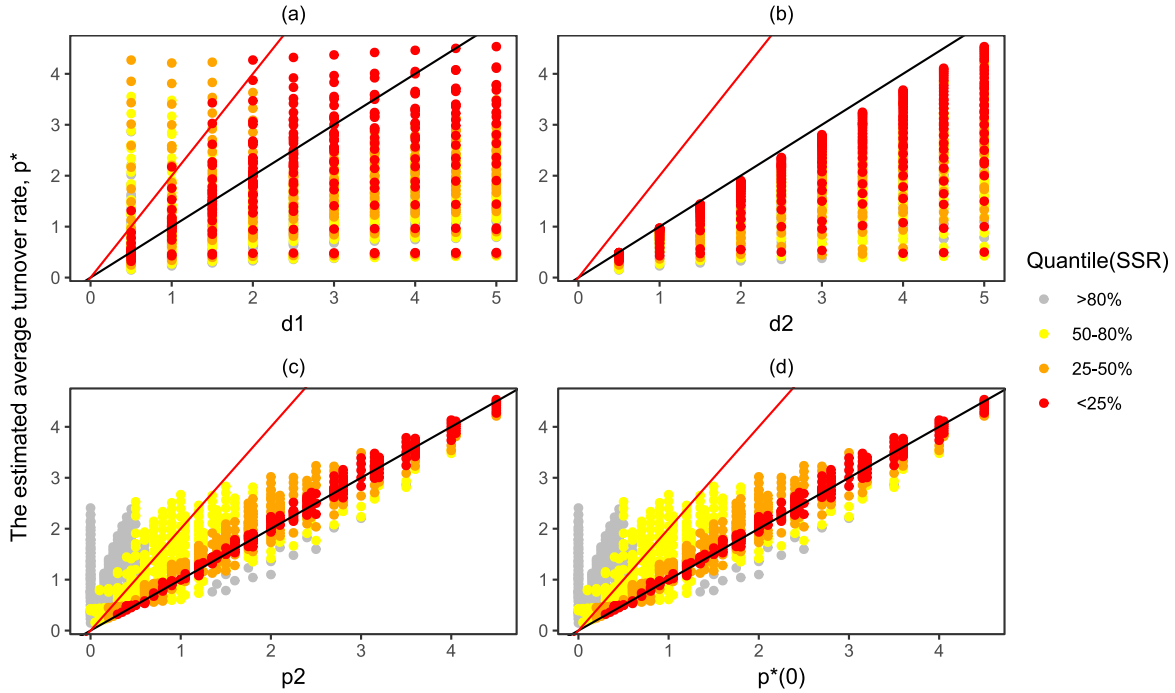507 parameters.

508

509 **Figure 3:** Plots showing the distribution of the estimated gain rate, p^*, as a function of the true
510 rates of the precursor population and the POI when $k = 1$. The black and red lines represent
511 slopes of 1 and 2, respectively. The lines with the slope of 1 and 2 show the line where the
512 value of p^* is equal to and twice that of the parameters on the x-axis, respectively. The colours

513   show four different groups of quantiles calculated based on the sum squared residual (SSR) of
514   the 1000 best fits. For example, the red circles show the 0-25% quantile group, which are the
515   top 250 best fits among the 1000 best fits.

516

517   In Figure 3, the implicit kinetic heterogeneity model was fitted to a dense, artificial dataset that
518   did not have any experimental noise. However, true deuterium labelling data are never free of
519   noise and are not as dense as our artificial data. Reassuringly, when we used a dense dataset
520   (Figure S4, top row), or a sparse dataset (Figure S4, bottom row), or when we added noise to
521   the data (Figures S5 and S6), the relationships between the parameters did not change.

522   Since the data points in these figures above are coloured by the quantile (based on the SSR of
523   the best fits) that those fits lay in, we can see that the deviations from the predictions are largely
524   due to the fits in the bottom 50% quantile (see Figures S4 and S6). However, as a quantile is a
525   relative measure, this does not necessarily mean that the fits in the bottom 50% quantile poorly
526   describe the data (see Figure S7). Thus, this numerical exercise not only confirms the analytical
527   results, but also identifies parameter values (for example, of p_2) where the estimated quantity
528   (p^*) is expected to have a large error.

529

530   ***If differentiation is accompanied by cell division (i.e., $k = 2$).*** The numerical results also tend
531   to be in line with the analytical results when the precursors go through division-linked
532   differentiation into the POI. The estimated label gain rate, p^*, always underestimates the true
533   turnover rate of the POI, d_2, as in the $k = 1$ case (Figure 4b). However, unlike the $k = 1$ case,
534   p^* always overestimates the true proliferation rate of the POI, p_2 (Figure 4c). This was bound
535   to happen as the estimate of p^* takes both the division-linked differentiation of the precursors
536   and the proliferation within the POI into account. The initial label gain rate, $p^*(0)$, is
537   approximately equal to the estimated p^* (Figure 4d). See Figure S7 for representative fits to
538   dense or sparse datasets.

Figure 4: Plots showing the distribution of the estimated gain rate, p^*, as a function of the true rates of the precursor population and the POI when $k = 2$. The black and red lines represent slopes of 1 and 2, respectively. The lines with the slope of 1 and 2 show the line where the value of p^* is equal to and twice that of the parameters on the x-axis, respectively. The colours show four different groups of quantiles calculated based on the sum squared residual (SSR) of the 1000 best fits. For example, the red circles show the 0-25% quantile group, which are the top 250 best fits among the 1000 best fits.

Qualitatively, Figures 4a-b look similar to Figures 3a-b, whereas Figures 4c-d seem to be a subset of Figures 3c-d. Thus, it appears that the $k = 2$ case is a subset of the $k = 1$ case. We will analytically confirm this below. Importantly, in both cases, the estimated p^* is consistently well-captured by the parameter $p^*(0)$, meaning that we can always estimate $p^*(0)$ from deuterium labelling data using the simple implicit kinetic heterogeneity model. However, the true turnover rate of a POI, d_2, cannot be inferred from $p^*(0)$ as the value of $p^*(0)$ depends on both p_2 and d_2, which weakens the significance of estimating this identifiable parameter.

The peaks of the labelling curves can help distinguish scenarios

Timing of the peak. Interestingly, it is possible to infer some properties of the precursors and the POI just by comparing both the height and the time of their peaks. The POI reaches its peak enrichment either precisely at, or sometime after, the end of the labelling period. For a model with random cellular processes (i.e., an ODE-based model), the peak of the POI will be *at the end* of the labelling phase when the precursors flowing into the POI have lower enrichment

than the cells in the POI, i.e., $l_2(1) > l_1(1) \Leftrightarrow p^*(0) > d_1$ (from equation 10). If precursors differentiate without dividing (i.e., when $k = 1$),

$$p^*(0) > d_1 \Leftrightarrow p_2 > d_1 \tag{13}$$

i.e., the peak will be at the end of the labelling phase when the division rate of the cells in the POI is higher than the loss rate of the precursors. Instead, if the differentiation of the precursors is accompanied by division (i.e., when $k = 2$),

$$p^*(0) > d_1 \Rightarrow \begin{cases} p_2 > d_1, \text{ or} \\ d_2/2 > d_1 > p_2 \end{cases} \tag{14}$$

meaning that the peak will be at the end of the labelling phase when the division rate of the POI is either higher than the loss rate of the precursors, $p_2 > d_1$, or is lower than the loss rate of the precursors, and the loss rate of the precursors is lower than half of the loss rate of the POI, $d_2/2 > d_1 > p_2$. Summarizing, the peak of the POI is at the end of the labelling period only when the enrichment of the POI is largely due to division (within the POI or by differentiating precursors).

There are two corollaries of the above equations (equations 13 and 14). One, a non-dividing POI always achieves its peak labelling after the stop of labelling. Two, the peak of the POI will be *after the end* of the labelling phase if the precursor cells flowing into the POI at and after the end of the labelling phase have a higher enrichment than the cells in the POI, i.e.,

$$l_1(1) > l_2(1) \Rightarrow d_1 > p^*(0)$$

$$\Rightarrow \begin{cases} d_1 > d_2, \text{ or} \\ d_2 > d_1 > p_2 \end{cases} \quad \text{when } k = 1 \tag{15}$$

$$\Rightarrow \begin{cases} d_1 > d_2, \text{ or} \\ 2d_1 > d_2 > d_1 > p_2 \end{cases} \quad \text{when } k = 2$$

Thus, the peak of the POI is after the end of the labelling phase only when non-dividing precursors differentiating into the POI are the major contributor of label in the POI. Therefore, the population architecture underlying a labelling curve can be partially realized by comparing the location of the peaks of the POI and its precursors (provided the dataset is sufficiently dense).

Division-linked differentiation. To further pinpoint the properties of the precursors and the POI, it is important to be able to distinguish between labelling curves where differentiation is accompanied by division from those where it is not. Unfortunately, whether a population goes through division-linked differentiation is, in general, unidentifiable (i.e., curves with $k = 1$ and $k = 2$ are indistinguishable). To demonstrate this, we compare the labelling equations of the

POI when $k = 1$ and $k = 2$ (equation 2h). Denoting the division rate of the POI, p_2, as p_{21} when $k = 1$, and as p_{22} when $k = 2$, we find

$$\frac{dl_2(t)}{dt}\bigg|_{k=1} = \frac{dl_2(t)}{dt}\bigg|_{k=2} \tag{16a}$$

$$\Rightarrow d_2(l_1 - l_2) + p_{21}(D(t) - l_1) = \frac{d_2(l_1 + D(t) - 2l_2) + p_{22}(D(t) - l_1)}{2} \tag{16b}$$

$$\Rightarrow p_{22} = 2p_{21} - d_2 \tag{16c}$$

where $p_{21} \geq d_2/2$. Therefore, the labelling curve of any population whose precursor goes through division-linked differentiation can also be explained by a scenario where the precursors differentiate without division (see Table 1 for examples). On the other hand, the labelling curve of a population whose precursors do not go through division-linked differentiation cannot be described with $k = 2$ if more than half of the production of the POI is due to the source, i.e., if $p_2 < d_2/2$. Thus, one can conclude that the influx into the POI is not accompanied by division if $p_2 < d_2/2$ in the estimates of the best fit. In a scenario where it is not known whether differentiation is linked with division, it is safe to conclude that the POI is primarily maintained by self-proliferation only if $p_2 \geq 3d_2/4$. Fortunately, the above transformation (equation 16) does not affect the turnover rate of the POI, d_2. Therefore, the turnover rate of the POI is identifiable if (and only if) the labelling in the precursors is known.

Intersecting labelling curves. Consider a simple scenario where precursors differentiate without division into a non-dividing POI, i.e., $k = 1$ and $p_2 = 0$ (see above). We have seen above that the peak enrichment in the POI will then be reached after the end of the labelling phase (where $D(t > 1) = 0$). More importantly, the peak occurs when $l_2(t) = l_1(t)$ (equation 4). Conversely, if $l_2(t) = l_1(t)$ when $t > 1$, equation 2i boils down to

$$k = 1 - \frac{p_2}{d_2} \tag{17}$$

As k takes only integer values and $p_2 < d_2$, the only value that satisfies the identity in equation 17 is $p_2 = 0$, implying $k = 1$. This provides a cardinal property of a chain of populations:

$$l_2(t) = l_1(t) \Leftrightarrow k = 1, p_2 = 0 \tag{18}$$

i.e., if during the de-labelling phase, a population reaches its peak when its labelling curve intersects that of its precursors, then this is proof that the POI is not dividing, and that its precursors differentiate into the population without division. This immediately implies that such a population ($k = 1, p_2 = 0$) has a lower peak than its precursor.

Case studies confirm that the underlying maintenance mechanisms can have enormous effect on the estimated turnover rate of the POI

The ES model can explain a wide variety of labelling curves, including the curves generated by the IS model. Both $p^* \leq d^*$ and $p^* > d^*$ labelling behaviour can be explained by the ES model (Table 2). This generality of the ES model is troubling, however. Most current estimates of cell lifespans are based on IS-like models that do not consider a precursor population, and the estimated rate at which label is gained is generally interpreted as the average turnover rate of the POI [5,6]. Since in the ES model, the same estimate could be reflecting the rate of the precursors (d_1), of the POI (d_2, p_2), or anything in between, this questions some of the current interpretations. We illustrate this problem with two examples.

Labelling of CD8[+] memory T cells by Westera et al. (2013)

In the study by Westera et al. (2013) [10] murine CD8[+] memory T cells were measured in a deuterium labelling experiment. The mice were drinking deuterated water for 1, 4 or 8 weeks. Using IS-like models (where the labelling in the source is the same as that in the body water), it was concluded that the CD8[+] memory T-cell pool is heterogeneous, because the data was best described with the kinetic heterogeneity (KH) model [10]. Above, we showed that a population whose label gain rate is slower than its loss rate can also be described by the ES model (equations 3 and 4c, Table 2), and hence as alluded to previously [10], is not conclusive evidence for heterogeneity in the POI. Further, recent reports have claimed that the memory T-cell pool in laboratory mice is partly (~10%) maintained by a continuous influx from naive T cells [15]. Therefore, we re-analysed the data from Westera et al. by fitting the data both to the KH model (with 3 parameters: d_1, d_2 and α) and to the ES model with $k = 1$ (again with 3 parameters d_1, d_2 and p_2); see Figure 5. Expectedly, a homogeneous POI with a source (i.e., the ES model in equation 2) explains the Westera et al. data [10] equally well, but with very different estimates for the turnover rate of the POI (Figure 5, Table 3). The ES model estimated a turnover rate of $\bar{d} = d_2 = 0.24$/day (i.e., a lifespan of 4 days) of the POI, which is about 5-fold higher than the average turnover rate of $\bar{d} = 0.046$/day (i.e., a ~ 20-day lifespan) estimated using the KH model (Table 3). Thus, the presence of a source that does not get instantaneously labelled can have a remarkable influence on the estimated lifespan of the POI (Table 3). However, since the estimated division rate of the POI, i.e., $p_2 = 0.04$/day, would suggest that only a mere ~17% (calculated as p_2/d_2) of the POI is replaced by self-replication, one would require a much larger contribution from the naive T cell pool than the estimated ~10% [15]. Thus, although it is possible that CD8[+] memory T cells are maintained by an influx from a precursor population, a contribution of 83% is unlikely.
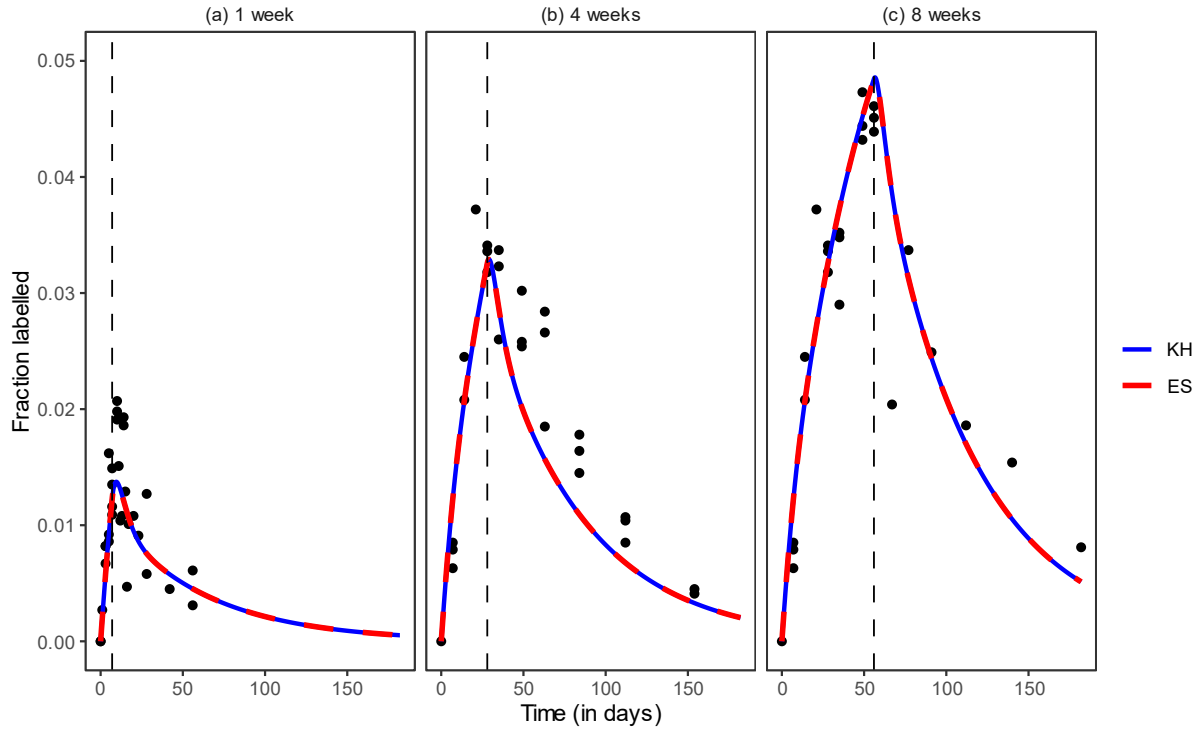
Figure 5: The best fits to the data from Westera et al. (2013) [10] with the KH model and the ES model with $k = 1$. The curves are on top of each other because the ES model mathematically encompasses the KH model. The data shows the labelling of the CD8$^+$ memory T cells in 1-week (a), 4-week (b) and 8-week (c) labelling experiments with deuterated water.

KH	d_1	d_2	α	$\bar{d} = \alpha d_1 + (1 - \alpha)d_2$	
	0.24	0.02	0.12	0.046	
ES	d_1	d_2	p_2	$\overline{N_2}/\overline{N_1}$	$\bar{d} = d_2$
	0.02	0.24	0.04	0.1	0.24

Table 3: The estimates of the models for the best fits shown in Figure 5. The size of the total population in the KH model can be scaled to any value. Note the large difference in the estimated turnover rates of the POI, $\bar{d}$, in the KH model and the ES model.

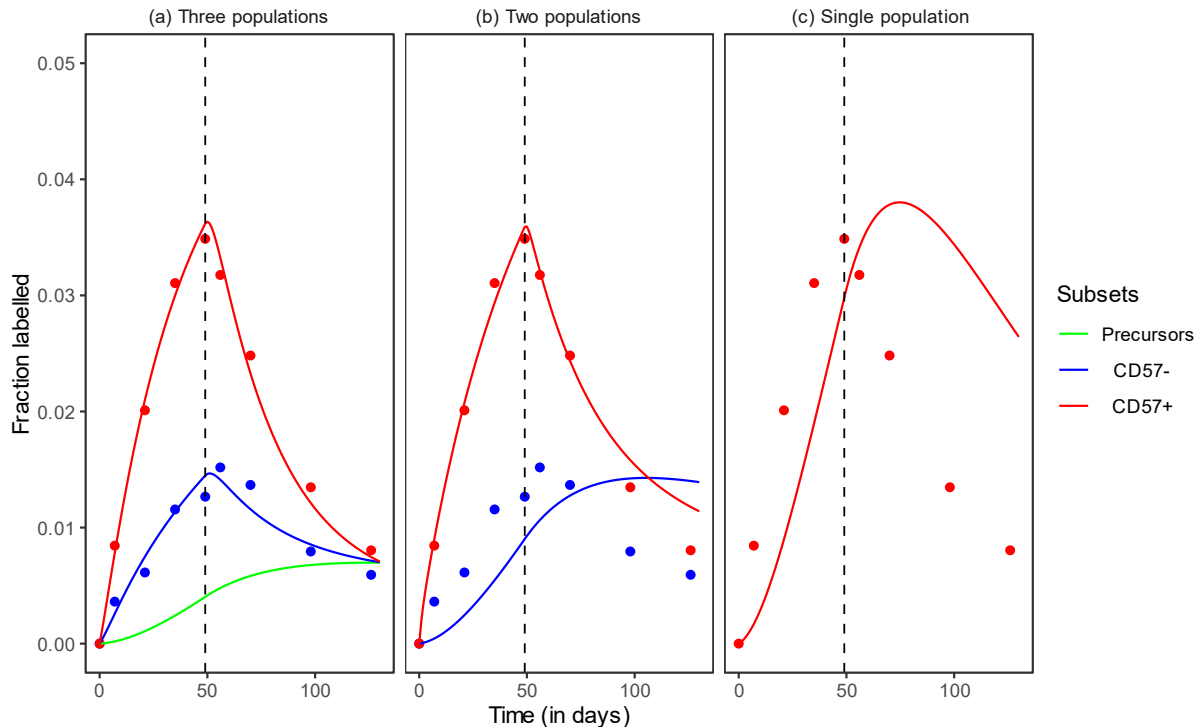### Labelling of CD57$^+$ memory T cells by Ahmed at al. (2020)

A dataset that seems uniquely suited for our analysis was generated by Ahmed et al. (2020) who measured the label gain and loss in both the POI, i.e., CD57$^+$CD4$^+$ memory T cells, and their precursor population, i.e., CD57$^-$CD4$^+$ memory T cells [12]. To find out the dynamics of the POI, they used a phenomenological chain of two populations describing both the POI and the precursors (see [12] for their model). Their analysis showed that the POI was maintained

largely (~ 95%) by division rather than by the source from the precursors (~5%). Hence, the POI and the precursors were maintained independently.

To re-analyse these data, we considered a three population chain model where the $CD57^+CD4^+$ cells are fed by the $CD57^-CD4^+$ population, that is in turn fed by an unknown precursor population. In the three-populations chain model, the parameters d_0, d_1 and d_2 define the turnover rates of the precursors of $CD57^-CD4^+$, $CD57^-CD4^+$ and $CD57^+CD4^+$ populations, respectively (see the legend of Figure 6). The labelling dynamics of both the $CD57^+CD4^+$ and the $CD57^-CD4^+$ populations were well-described by this three population chain model (Figure 6a). Both populations were estimated to have fast turnover rates and a similar size, with ~ 14% (p_1/d_1) of the $CD57^-CD4^+$ cells and ~ 33% (p_2/d_2) of the $CD57^+CD4^+$ cells being maintained by division. The description of the labelling data became significantly worse (Figures 6b,c) when the data was described either with a two population chain (ES) model where $CD57^-CD4^+$ are the precursors of $CD57^+CD4^+$ (Figure 6b), or with a single compartment (IS) model for the $CD57^+CD4^+$ cells (Figure 6c).

Notably, the turnover rate of the two POI, $CD57^-CD4^+$ and $CD57^+CD4^+$ memory T cells, estimated from a two population chain model and a three population chain model differed by 2-fold. This was due to the poor description of the precursors, $CD57^-CD4^+$ memory T cells. If the description of the labelling dynamics of the $CD57^-CD4^+$ memory T cells would have stayed the same, the estimates of the $CD57^+CD4^+$ memory T cells would have also remained unchanged.

Figure 6: The best fits of the three population ES model (a), the two population ES model (b), and the IS model (c) to the data from Ahmed et al. (2020) [12]. The data shows the labelling of $CD57^-CD4^+$ and $CD57^+CD4^+$ memory T cells in labelling experiments with deuterated

697   water. The estimates of the best fit of the three population ES model are: $d_0 = 0.001$/day, $d_1 =$
698   0.28/day, $d_2 = 0.4$/day, $p_1 = 0.04$/day, and $p_2 = 0.13$/day.

699

700   Note that both examples are from labelling experiments with deuterated water and the body
701   water enrichment approaches its asymptote slowly. Therefore, even a population that is entirely
702   self-renewing can have an initial delay in gaining label (Figure 6).

703   Both examples demonstrate that the estimated parameters can strongly depend on the
704   underlying model, and hence that one should always test whether the estimates change when a
705   precursor with a realistic turnover rate is added. In the first example, one could argue that the
706   previous KH model provides the best explanation, as the ES model requires an unrealistically
707   large contribution from the naive T-cell compartment. In the second example, the estimates
708   found from the ES model could be realistic, as they are in line with previous estimates stating
709   that as much as a quarter of the CD4$^+$ memory T-cell population could be maintained by a
710   source from naive T cells [15].

711

Discussion

713   Deuterium labelling is hailed as the current best technique to accurately measure the turnover
714   rates of a cell population *in vivo*. In this article, we share some insights into what can and what
715   cannot be deduced from deuterium labelling data using the current modelling approaches. We
716   show that models that fail to consider a slow source of label can severely underestimate the
717   true turnover rate of the POI. The rate at which the POI gains label is not simply its average
718   turnover rate but can be markedly influenced by the rate at which its precursors gain label. We
719   also show that typically one cannot tell from the labelling data whether cells divide while
720   differentiating into the POI. Therefore, unless one is convinced that 1) the precursors are much
721   faster, or 2) hardly play any role in maintaining the POI, a source should be modelled explicitly
722   (with the ES model) and ideally be matched with the labelling data of the precursors.

723   We found that the rate at which the POI gains label is always lower or equal to the true turnover
724   rate of the POI. Depending on the dynamics of its precursor, the labelling in the POI can follow
725   4 different scenarios. In cases where the lifespans of both the precursors and the POI are
726   comparable to the labelling duration, the rate of labelling varies with time and is not given by
727   the rates of any single population. So, we defined two label gain rates, one at the beginning of
728   the labelling period, $p^*(0)$, and the other at its end, $p^*(1)$. We show that these approximated
729   gain rates tend to be in good agreement with each other, and that both are smaller than the
730   turnover rate, d_2, of the POI. Thus, interpreting the label gain rate as the turnover rate can
731   markedly overestimate the lifespan of the measured population.

732   There are two cases where one can be certain that the estimated rate reflects a POI's true
733   turnover rate. First, when the source population's turnover rate is sufficiently fast, the labelling
734   in the POI is largely dictated by its own turnover rate. For example, thymocytes, that act as a

precursor to slowly turning-over naive T cells, turn over rapidly. Therefore, estimating the true turnover rate of naive T cells does not require measuring the labelling in the thymocytes. Second, when the population is largely maintained by proliferation (i.e., when the source into the POI is negligible). Memory T cells are a likely example of this scenario as they are thought to be largely maintained by proliferation. Therefore, although measuring the label incorporation in the precursors should be treated as the standard procedure, it may not be necessary in all deuterium labelling experiments. Note that there's also the case where precursor cells clonally expand during their differentiation into the POI cells (see the companion paper, *Yan et al.* (submitted)). As all the cells flowing into the POI, in this case, would be labelled, this scenario would approximate the first case mentioned above where the source is highly labelled.

Using two case studies as examples, we showed how estimates can become very different when the source population is not rapid. A study that focused on the labelling dynamics of the CD8$^+$ memory T cells was originally explained by a kinetically heterogeneous population and concluded that these cells had a ~20 day lifespan [10]. The ES model, in contrast, suggested that the CD8$^+$ memory T-cell pool could be homogeneous, where cells lived only ~4 days and were maintained by a much larger influx. Although some studies have estimated that circulating CD8$^+$ memory T cells could be short-lived [15,17], a very short-lived memory T-cell population that is primarily maintained by an influx from a large precursor population is not in line with many previous studies and can probably be discarded. Nevertheless, it remains important to study the possibility of a slow precursor, as in other cases this may provide a realistic description. In the second example that dealt with the maintenance mechanism of CD4$^+$CD57$^+$ memory T cells [12], the ES model suggested that these cells were only partly (~33%) maintained by division, as opposed to the original interpretation that suggested that they were largely maintained by self-replication. This is similar to previous conclusions on CD4$^+$ memory T cell pool, that are maintained by an influx from the naive CD4$^+$ T cell pool [15]. Additionally, in contrast to the original conclusion where the differentiation of CD4$^+$CD57$^-$ cells was accompanied by a clonal burst, the ES model showed that the labelling of both CD4$^+$CD57$^-$ cells and CD4$^+$CD57$^+$ cells can be well-described even if the differentiation of CD4$^+$CD57$^-$ cells is not accompanied by division. As very little is known about these two POIs, it is difficult to favour one of these solutions, underlying the importance of considering an explicitly modelled source when analysing deuterium labelling data.

In an independent study in the same issue as this article, *Yan et al.* (*Yan et al.*, submitted) also argue in favour of explicitly modelling the precursor population of the POI. Whereas we allow for maximally one cell division upon differentiation of a precursor cell into the POI (i.e., $k = 1$ or $k = 2$), *Yan et al.* also consider precursors that go through extensive clonal expansion ($k = 2^i$, where $i \geq 0$ in their paper, meaning that the expansion is instantaneous). Interestingly, regardless of this difference, our results are similar. This could be because the more general $k = 1$ case includes the $k > 1$ cases. We showed, here, that the case of $k = 1$ includes the $k = 2$ case. It would be interesting to see whether it also includes the $k > 2$ cases. Yan and colleagues find the same relationship between the estimated turnover rate of the POI and its true proliferation rate as we do, i.e., $p^* = p_2$ (when $k = 1$) and $p^* > p_2$ (when $k = 2$). They

777 also conclude that the turnover rate of the POI, d_2, is not equal to the estimate of p^*, however,
778 unlike us they only conclude that for the $k = 1$ case. They show that for $k > 1$, the turnover
779 rate of the POI is well-approximated by the estimated value of p^*, i.e., $p^* \approx d_2$. This is likely
780 due to the difference in our assumptions about maximally one division (in our case) and clonal
781 expansion (in *Yan et al.*), as most of their data samples could reflect cases where $k > 2$. If k is
782 high, the enrichment in the precursor cells flowing into the POI would be very high (close to
783 the maximum), essentially boiling down the ES model into the IS model, in which case it would
784 be expected that $p^* \approx d_2$. The agreement between the conclusions drawn by both studies
785 present a compelling case for explicitly modelling the precursor population of a POI that is at
786 homeostasis.

787 Here we show even the interpretation of deuterium labelling data, the current state-of-the-art
788 method to estimate a population's lifespan, can be prone to error. Strikingly, for several
789 scenarios, it is essential to have information on the dynamics of the precursors to not undermine
790 the reliability of the estimates of the POI found from deuterium labelling experiments.

791

Data and Code Availability

793 R scripts that were used to make the main and supplementary figures are available at
794 https://github.com/swainarpit/ExplicitSourceChain.

795

798

Author Contributions

800 ACS performed the mathematical modelling. ACS, JB and RdB designed the study. ACS, JB
801 and RdB analysed the data and wrote the manuscript. All authors contributed to the article and
802 approved the current version.

803

808

Competing Interests

References

1. Gratzner HG. 1982 Monoclonal antibody to 5-bromo- and 5-iododeoxyuridine: A new reagent for detection of DNA replication. *Science* **218**, 474–475. (doi:10.1126/SCIENCE.7123245)

2. CARTWRIGHT GE, ATHENS JW, WINTROBE MM. 1964 Analytical Review: The Kinetics of Granulopoiesis in Normal Man. *Blood* **24**, 780–803. (doi:10.1182/BLOOD.V24.6.780.780)

3. Macallan DC, Fullerton CA, Neese RA, Haddock K, Park SS, Hellerstein MK. 1998 Measurement of cell proliferation by labeling of DNA with stable isotope-labeled glucose: studies in vitro, in animals, and in humans. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 708–713. (doi:10.1073/PNAS.95.2.708)

4. Borghans JAM, Tesselaar K, de Boer RJ. 2018 Current best estimates for the average lifespans of mouse and human leukocytes: reviewing two decades of deuterium-labeling experiments. *Immunol. Rev.* **285**, 233–248. (doi:10.1111/IMR.12693)

5. Asquith B, Debacq C, Macallan DC, Willems L, Bangham CRM. 2002 Lymphocyte kinetics: The interpretation of labelling data. *Trends Immunol.* **23**, 596–601. (doi:10.1016/S1471-4906(02)02337-2)

6. De Boer RJ, Perelson AS. 2013 Quantifying T lymphocyte turnover. *J. Theor. Biol.* **327**, 45–87. (doi:10.1016/J.JTBI.2012.12.025)

7. Ahmed R *et al.* 2015 Reconciling Estimates of Cell Proliferation from Stable Isotope Labeling Experiments. *PLOS Comput. Biol.* **11**, e1004355. (doi:10.1371/JOURNAL.PCBI.1004355)

8. Lahoz-Beneytez J, Elemans M, Zhang Y, Ahmed R, Salam A, Block M, Niederalt C, Asquith B, Macallan D. 2016 Human neutrophil kinetics: modeling of stable isotope labeling data supports short blood neutrophil half-lives. *Blood* **127**, 3431. (doi:10.1182/BLOOD-2016-03-700336)

9. Ganusov V V., Borghans JAM, De Boer RJ. 2010 Explicit kinetic heterogeneity: Mathematical models for interpretation of deuterium labeling of heterogeneous cell populations. *PLoS Comput. Biol.* (doi:10.1371/journal.pcbi.1000666)

10. Westera L *et al.* 2013 Closing the gap between T-cell life span estimates from stable isotope-labeling studies in mice and humans. *Blood* **122**, 2205–2212. (doi:10.1182/BLOOD-2013-03-488411)

11. Vrisekoop N *et al.* 2008 Sparse production but preferential incorporation of recently produced naïve T cells in the human peripheral pool. *Proc. Natl. Acad. Sci. U. S. A.*

847   **105**, 6115–6120. (doi:10.1073/PNAS.0709713105/SUPPL_FILE/SD3.TXT)

848   12.   Ahmed R *et al.* 2020 CD57+ Memory T Cells Proliferate In Vivo. *Cell Rep.* **33**.
849   (doi:10.1016/J.CELREP.2020.108501)

850   13.   Restifo NP, Gattinoni L. 2013 Lineage relationship of effector and memory T cells.
851   *Curr. Opin. Immunol.* **25**, 556–563. (doi:10.1016/J.COI.2013.09.003)

852   14.   Sallusto F, Geginat J, Lanzavecchia A. 2004 Central memory and effector memory T
853   cell subsets: function, generation, and maintenance. *Annu. Rev. Immunol.* **22**, 745–763.
854   (doi:10.1146/ANNUREV.IMMUNOL.22.012703.104702)

855   15.   Gossel G, Hogan T, Cownden D, Seddon B, Yates AJ. 2017 Memory CD4 T cell
856   subsets are kinetically heterogeneous and replenished from naive T cells at high levels.
857   *Elife* **6**. (doi:10.7554/ELIFE.23013)

858   16.   Pillay J, Den Braber I, Vrisekoop N, Kwast LM, De Boer RJ, Borghans JAM,
859   Tesselaar K, Koenderman L. 2010 In vivo labeling with 2H2O reveals a human
860   neutrophil lifespan of 5.4 days. *Blood* **116**, 625–627. (doi:10.1182/BLOOD-2010-01-
861   259028)

862   17.   Bresser K *et al.* 2022 Replicative history marks transcriptional and functional disparity
863   in the CD8+ T cell memory pool. *Nat. Immunol.* **23**, 791–801. (doi:10.1038/S41590-
864   022-01171-9)

865