

Supplementary information

The ES model describes both non-stem-like and stem-like populations

The ES model makes only the most fundamental assumptions (of a possible source, of possible division in the population, and of loss from the population), that should be true for any population and, thus, is very general. With a few substitutions, we can also express the system of equations (equations 2a-d) in the classical form of stem cells [1], i.e.,

$$\begin{aligned}\frac{dN_1}{dt} &= \sigma + (p_1 - d_1)N_1 \\ &= \sigma - (1 - \alpha)d_1N_1 - \alpha d_1N_1 - p_1N_1 + 2p_1N_1 \\ &= \sigma - (1 - \alpha)d_1N_1 + (2\beta - 1)\delta N_1\end{aligned}\tag{S1a}$$

$$\begin{aligned}\frac{dN_2}{dt} &= 2\alpha d_1N_1 + (p_2 - d_2)N_2 \\ &= 2(1 - \beta)\delta N_1 + (p_2 - d_2)N_2\end{aligned}\tag{S1b}$$

where $\delta = p_1 + \alpha d_1$ is the total division rate of the precursor population, and $\beta = \frac{p_1}{\delta} < 1$ is the fraction of the daughter cells that replenish the precursor population. The fraction $(1 - \beta)$ of daughter cells differentiate into the next population, N_2 . Therefore, the ES model is general enough to describe both non-stem-like and stem-like populations.

The initial gain and loss rates are well-described even when the POI is very short-lived

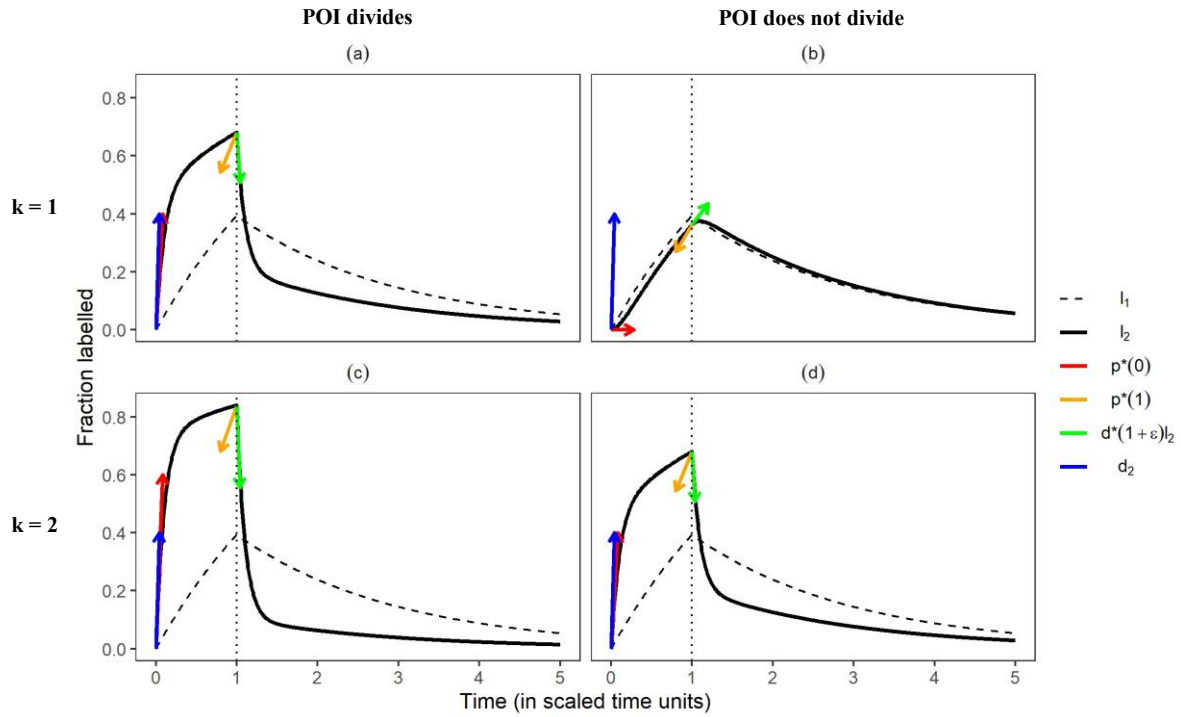


Figure S1: The estimated labelling rates when the POI is very short-lived. The loss rate of the POI was chosen to be 20 times higher than that of the precursors ($d_1 = 0.5$, $d_2 = 10$). Note that the slopes are calculated as the label (gain or) loss rate multiplied by the fraction of (un)labelled cells, for example, $d^*(1 + \epsilon)l_2$. The division rate of the POI, p_2 , was either set to 5 (in (a) and (c)) or 0 (in (b) and (d)).

d_1	d_2	p_2	k	$p^*(0)$	$p^*(1)$	$d^*(1 + \epsilon)$
0.5	10	5	1	5	0.68	6.33
			2	7.5	0.80	8.43
		0	1	0	0.46	-1
			2	5	0.68	6.33

Table S1: The true and calculated rates corresponding to the simulations shown in Figure S1. The rates (expressed as stu) are scaled with respect to the labelling period.

The best description of the labelling in the POI by the phenomenological p^*d^* model

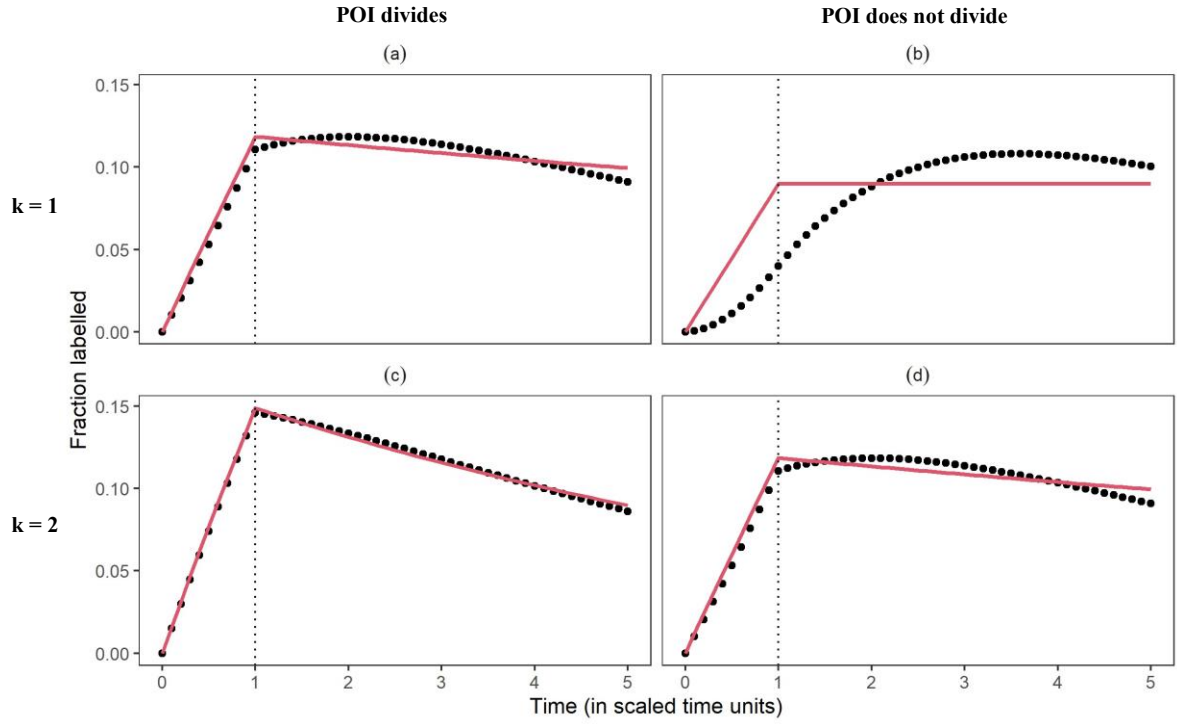


Figure S2: The best fits of the phenomenological model (equation 8 in the main text) to the labelling curve of the POIs shown in Figure 1. The black circles show the data, and the red trajectory shows the best fit.

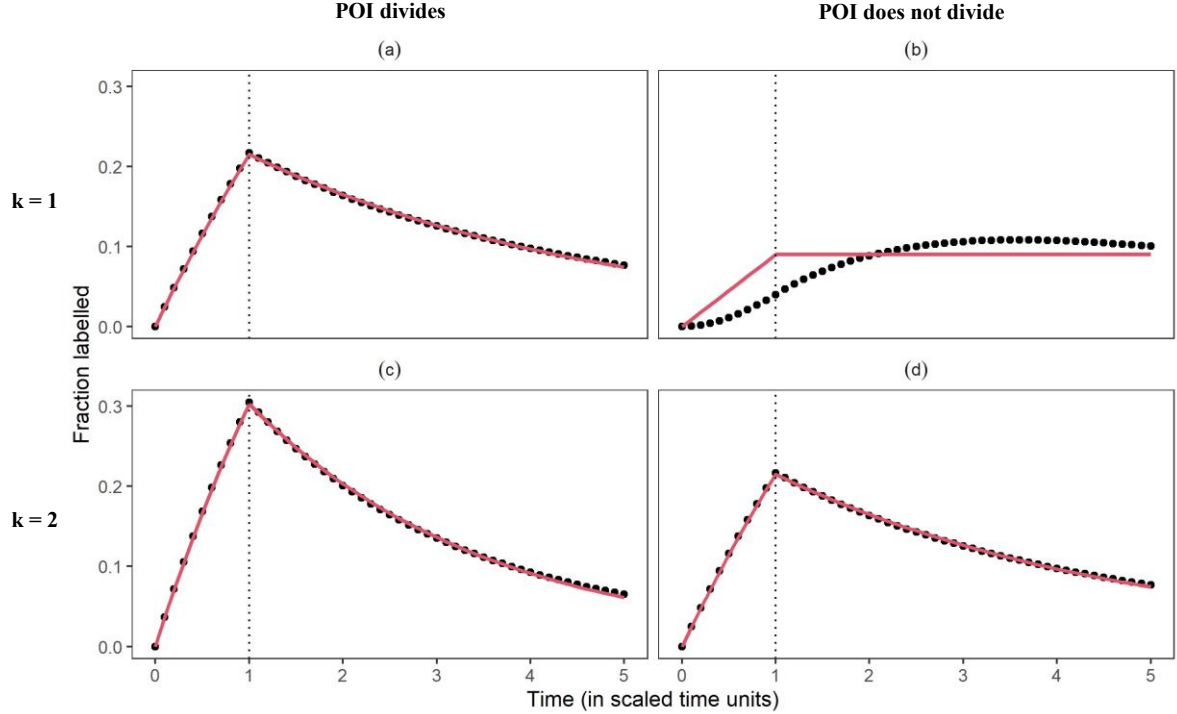
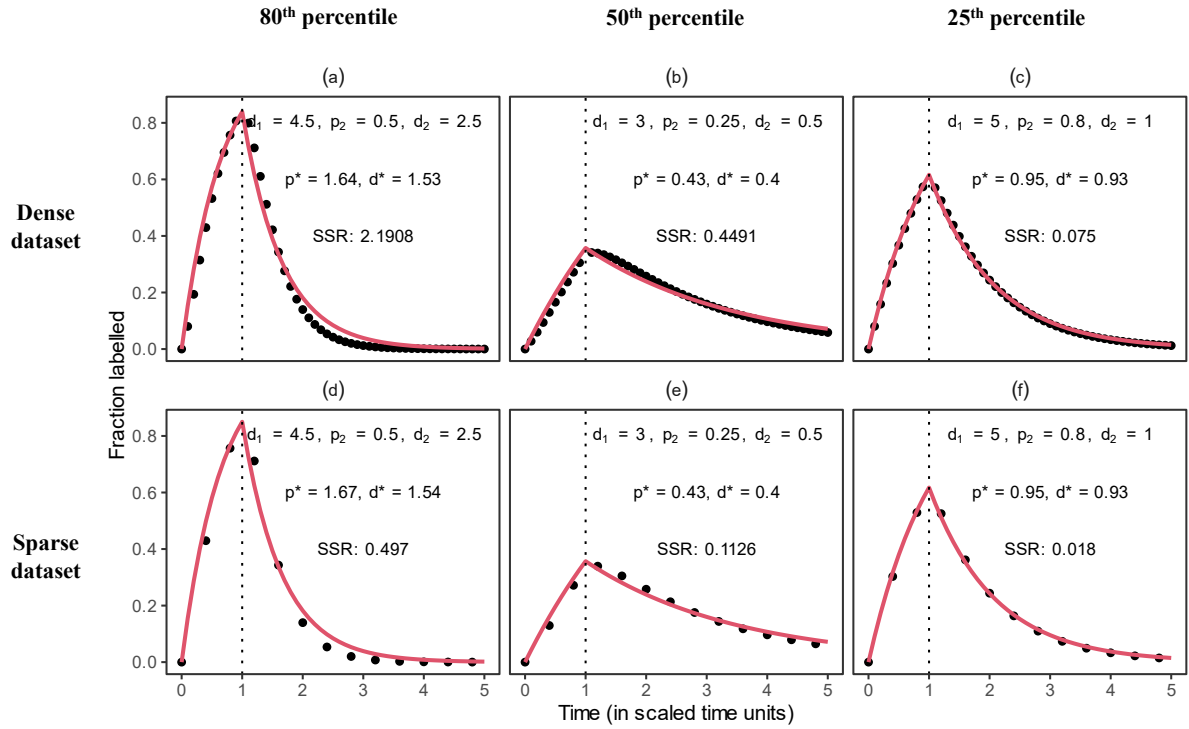


Figure S3: The best fits of the phenomenological model (equation 8 in the main text) to the labelling curve of the POIs shown in Figure 2. The black circles show the data, and the red trajectory shows the best fit.

36 Representative fits of the implicit source model (equation 8) for the $k = 1$ case



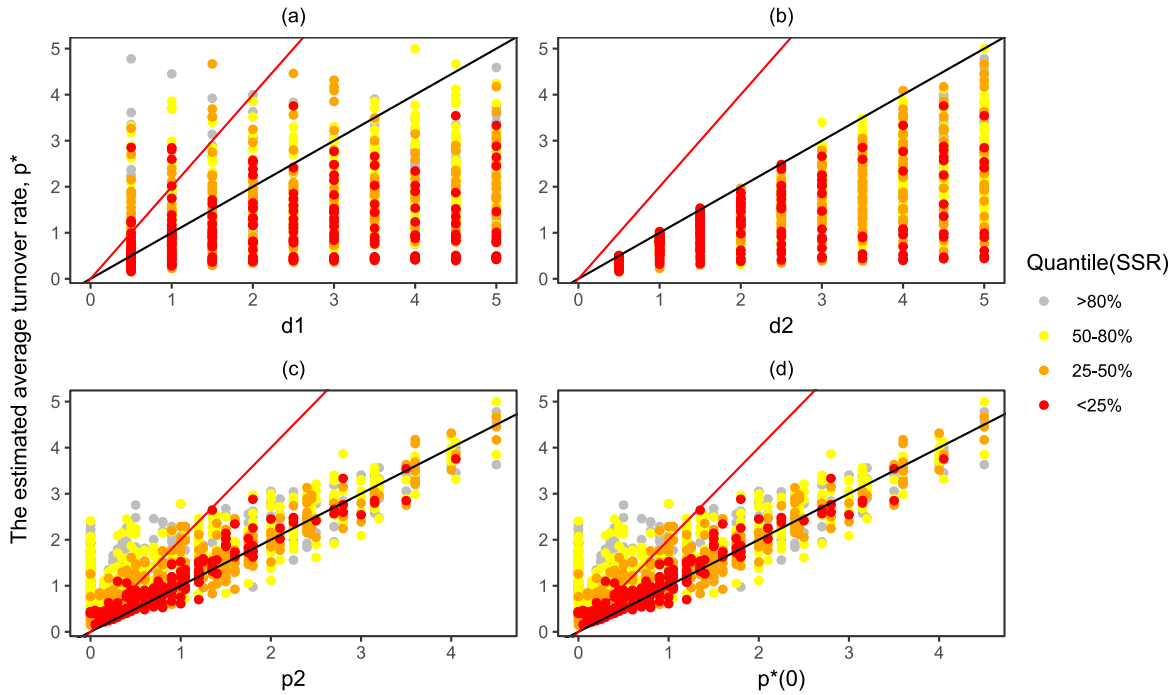
37

38 **Figure S4:** Examples of the best fits used to generate Figure 3. The top row shows fits to
 39 dense data and the bottom row shows fits to sparse data. Note that the fits and the estimates
 40 hardly change if the data is made sparser.

41

42 Correlation plots and representative fits of the implicit source model (equation
 43 8) for the $k = 1$ case based on noisy data

44 True experimental data is noisy. Here, we test whether the relationships between the true
 45 parameters and the estimated parameters are significantly different if the artificial dataset is
 46 'noisy'.



47

48 **Figure S5:** The same as Figure 3, except that each data point in the artificially generated
 49 dataset is replaced by two ‘noisy’ data points. The two noisy data points are drawn from a
 50 normal distribution with a mean that equals the original data point and a 20% standard
 51 deviation. The black and red lines represent slopes of 1 and 2, respectively.

52

53 The conclusions drawn from the noisy dataset are the same as those drawn from the ‘clean’
 54 dataset. However, the noisy dataset exaggerates the quantitative behaviour, i.e., the difference
 55 between the estimated p^* and the true value of p_2 is larger due to the noisy dataset (Figure
 56 S5d).

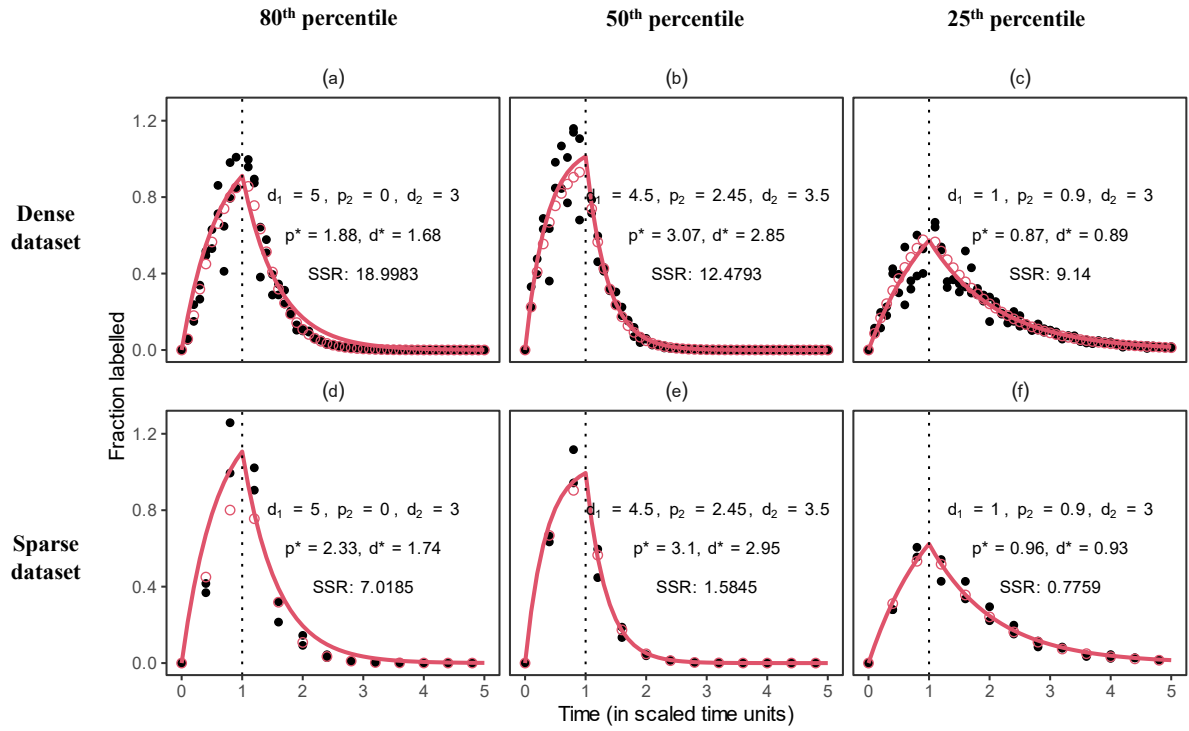


Figure S6: Examples of the best fits used to generate Figure S5. The top row shows fits to dense data and the bottom row shows fits to sparse data. Note that the fits and the estimates hardly change if the data is made more sparse. The open red circles show the data points without noise added to them, while the black bullet points are the data points after noise was introduced. There are two black bullet points corresponding to each open red circle. The model was fitted to the black bullet points.

Representative fits of the implicit source model (equation 8) for the $k = 2$ case

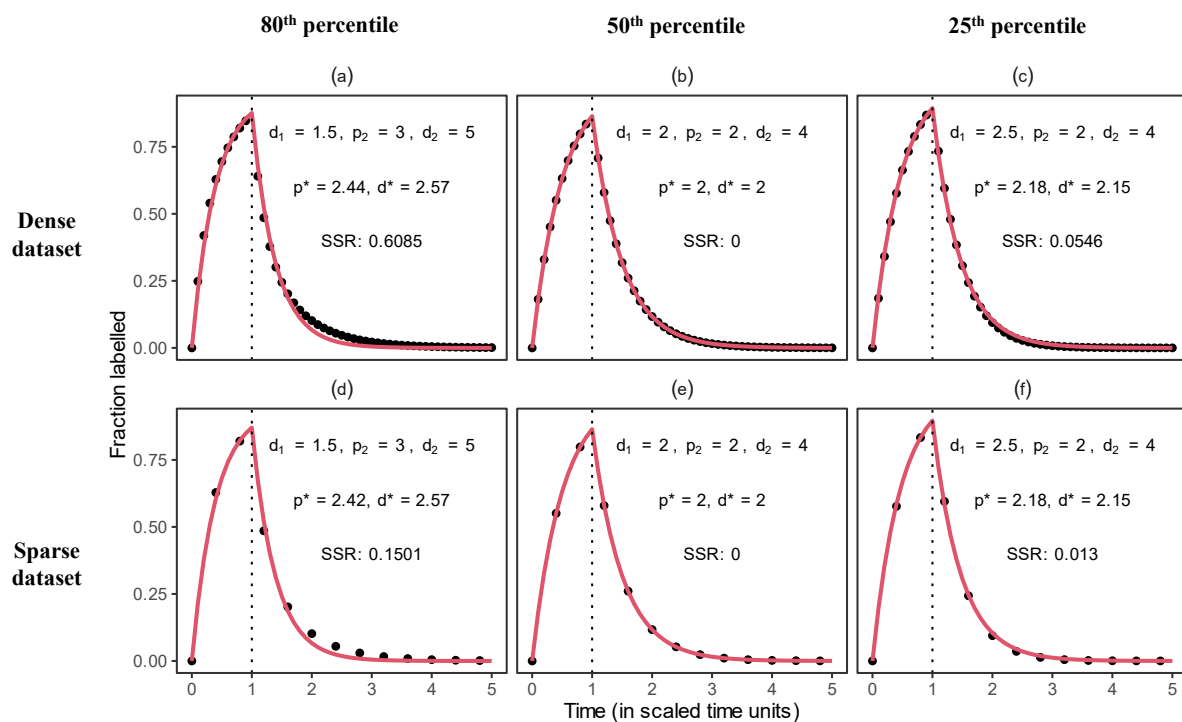


Figure S7: Examples of the best fits used to generate Figure 4. The top row shows fits to dense data and the bottom row shows fits to sparse data. Note that the fits and the estimates hardly change if the data is made more sparse.

References

1. Lander AD, Gokoffski KK, Wan FYM, Nie Q, Calof AL. 2009 Cell Lineages and the Logic of Proliferative Control. *PLOS Biol.* **7**, e1000015. (doi:10.1371/JOURNAL.PBIO.1000015)