SI Appendix

Estimating cell number over time

The total number of CD8+ T cells (T) in an animal at a given age (a) was measured from pooled lymph node and spleen cells of different ages, and the cell number over time, then fitted using a logistic curve as below:

$$T(a) = \frac{KT_0e^{ga}}{K+T_0(e^{ga}-1)} \quad (1)$$

in which,

$T(a) =$ Total CD8 T cell level at age a

$T_0 =$ Initial number of total CD8 + T cells at birth

$g =$ Rate of increase of total CD8 + T cells

$K =$ Stable level of total CD8 + T cells.

Similarly, we also fitted the background level of RFP expression (B) (as a percent of CD8+ T cells) with age using a logistic curve.

$$B(a) = \frac{PB_0e^{ba}}{P+B_0(e^{ba}-1)} \quad (2)$$

in which,

$B(a) =$ Background %RFP expression level at age a

$B_0 =$ Initial %RFP expression at birth

$b =$ Rate of increase of %RFP expression

$P =$ Stable (plateau) level of %RFP expression.

Then, using this following relationship

$$\%\ labeled\ CD8 = \frac{\#\ labeled\ CD8}{\#\ total\ CD8} + \%\ background\ RFP \quad (3)$$

and the estimated background level and total CD8+ T cell numbers at any age, the corrected number of labeled CD8+ T cells is:

$$\#\ labeled\ CD8 = (\%\ labeled\ CD8 - \%\ background\ ) * \#\ total\ CD8 \quad (4)$$

Dynamics of labeled cells

Let *L* denote the number of labeled CD8+ T cells at any time. If survival of these cells is exponential, decaying at rate *k*, then the number of labeled cells remaining at time *t* will be;

$$L(t) = L_0 e^{-kt} \qquad (5)$$

However, inspection of the curves suggests there may be a higher initial rate of cell loss (*k*) that slows over time (Figure 2B). This might either slow with age (slower decay in older animals), or with time since production of the cell (slower decay for older cells). Thus, we fit a model where the initial decay rate *(k_0)* decreases at rate *d* either with time since production of the cell in the thymus (*t*), or the age of the animal (*a*).

If the loss rate of cells slows with time since cell production, then the number of labeled cells should decrease as;

$$L(t) = L_0 e^{-\left(ke^{-\delta t}\right)t} \qquad (6)$$

However, if the initial loss rate slows with the age of the animal, we have:

$$L(t) = L_0 e^{-\left(ke^{-\delta a}\right)t} \qquad (7)$$

The first model implies that all cells behave the same regardless of the age at which they were produced (ie: they only change with time-since-production – equation 6). Whereas the second model implies that cell behavior does not change with the age of the cell (time-since-production), but is instead dependent only on the age of the animal (equation 7). It is also possible that both age of the animal and age of the cell may affect behavior. In this case we can fit a model where cells stamped at different ages have different initial decay rates (for example, allow different decay rates - k_1 to k_5, for the 5 host ages at which cells were stamped). Similarly, the rate of slowing down in the initial decay rate might be different for each age group of hosts at which cells were stamped ($\delta_1 - \delta_5$). Finally, we also consider a model where the initial rate of decay changes with age of the host; and this initial rate slows down with time (model 8, see Supplementary Table 1).

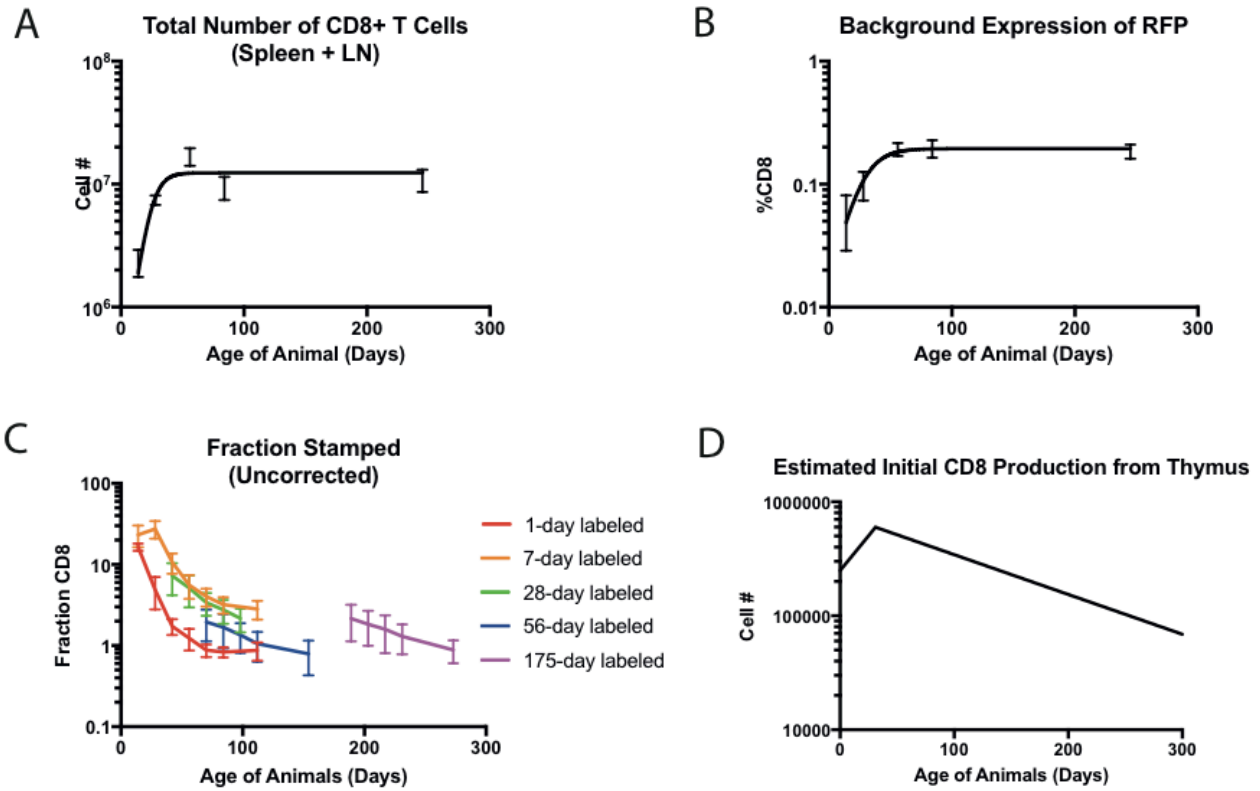
For model comparison, first the model with the lowest AIC was chosen as a baseline, then we subtracted this AIC_{min} from all AIC. A model with a difference in AIC greater than 10 was ignored. We excluded the first measurement from the mice labelled at 7 days of age, since their first sample was taken at one week after tamoxifen treatment when labeled cells were still emerging from the thymus.

Allowing for variation in background staining and total cell number

A caveat of our experimental system is that while labeled cells were tracked over time in individual animals, total cell number (and background staining) was analyzed in a separate cohort (because of the need to analyze spleen and lymph nodes at every time point). Thus, the proportion of labeled cells was corrected based on the mean total cell number and background staining in a separate cohort of untreated animals. To account for potential variation in this mean, we performed 1000 bootstraps (pick one with replacement) of this background and total cell number data, each time refitting our functions of cell number and background with age. Since the total cell and background were sparsely sampled, we also performed a leave-one-out analysis to randomly exclude one time point (Supplementary Figure 2). We then used each new bootstrapped dataset to re-estimate corrected cell numbers and refitted our models. Supplementary Figure 3A and 3C shows the distribution of AIC difference from this procedure. Overall, the results are consistent with our reported main result (Supplementary Figure 3B and 3D - model 5 as the best model, but we picked model 9 by parsimony), The distribution of the estimated parameters from model 9 (with background and total cell variation included) were presented in Supplementary Figure 4A (total CD8 population) and 4B (naïve population).

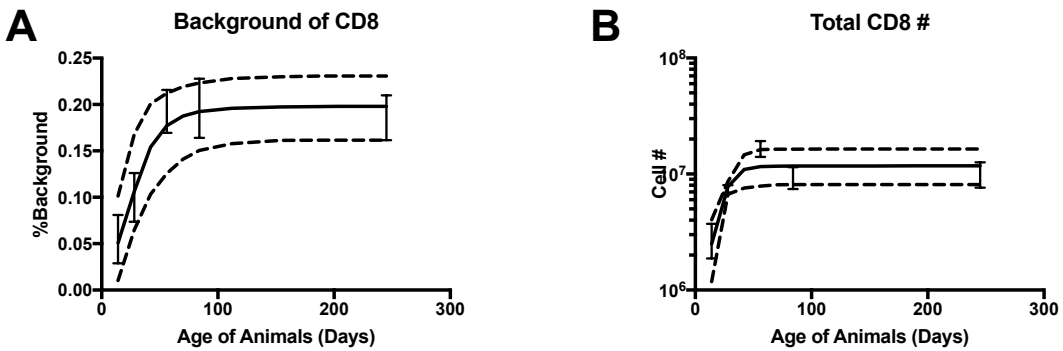
Supplementary Figure 1

(A) Total $CD8^+$ T cells from spleen and lymph nodes (cervical, mesenteric, and inguinal lymph nodes) were collected from timestamp mice at 14d, 28d, 56d, 84d, and 245d (n = 7 – 8 per time point). (B) The background level of RFP expression with age measured in mice that were never given tamoxifen. (C) Persistence of labeled $CD8^+$ T cells made at various ages. Shown here is stamped $CD8^+$ T cells as a percentage of total $CD8^+$ T cells over time for the 5 groups of mice (labeled at 1 day, 7 days, 28 days, 56 days, and 185 days). (D) Estimated $CD8^+$ T cell production with age using a piecewise exponential function.



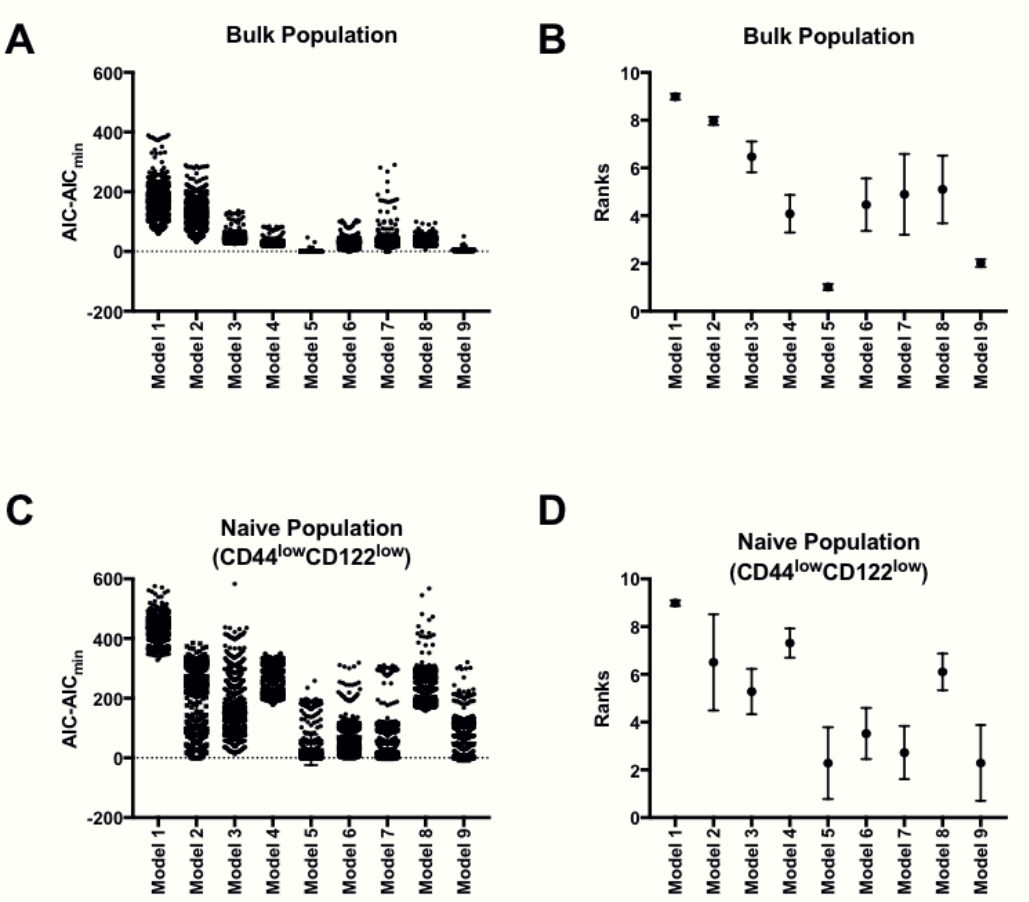
Supplementary Figure 2

Quantifying uncertainty from background staining (A) and total $CD8^+$ T cells (B). Animals were sampled at day 14 (n=7), 28 (n=8), 56 (n=7), 84 (n=9), and 245 (n=8). The error bars represent the standard deviation from each time point across different animals. We sampled one observation for each time point randomly (with replacement), then did a non-linear least-squares fitting to a logistic curve. Moreover, since we only sparsely sampled these animals to quantify the background staining and total cell number, we also randomly leave one time point out during the least-squares fitting. We did this procedure 1000 times, and the lower and upper dashed lines represent the lower 2.5^{th} percentile and 97.5^{th} percentile, respectively, of the fitting results. The solid line represents the mean of the fitting results.



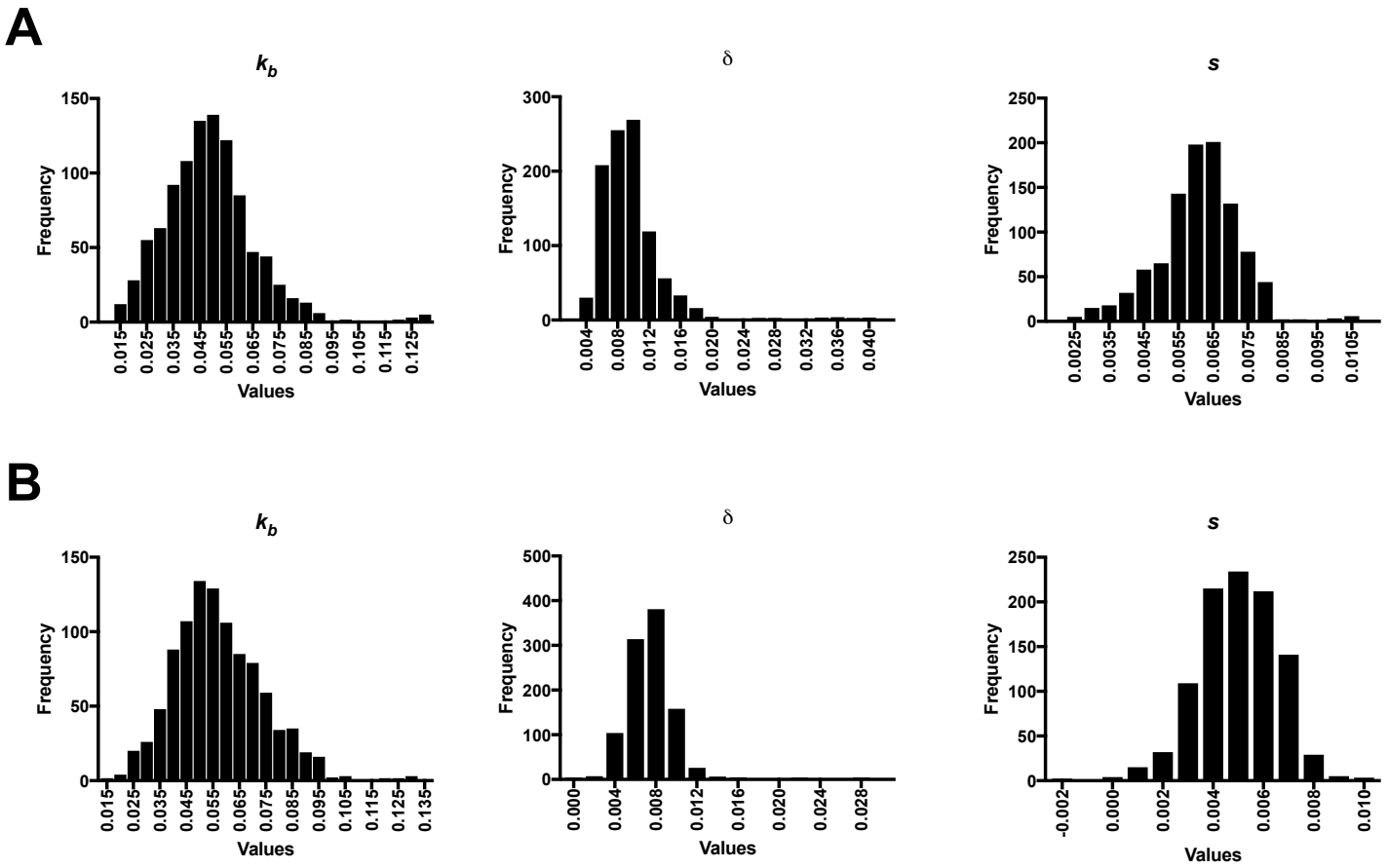
Supplementary Figure 3

The distribution of AIC-AIC_{min} after propagating uncertainty introduced when fitting the background and total CD8 T cell numbers. We did this to the total bulk CD8 T cell population (A) and naïve phenotype population (C). We also ranked these AIC-AIC_{min} values (error bars represent SD, n = 1000) in the bulk population (B) and naïve population (D). We found that model 5 and model 9 were equally good, but we picked model 9 due to parsimony.



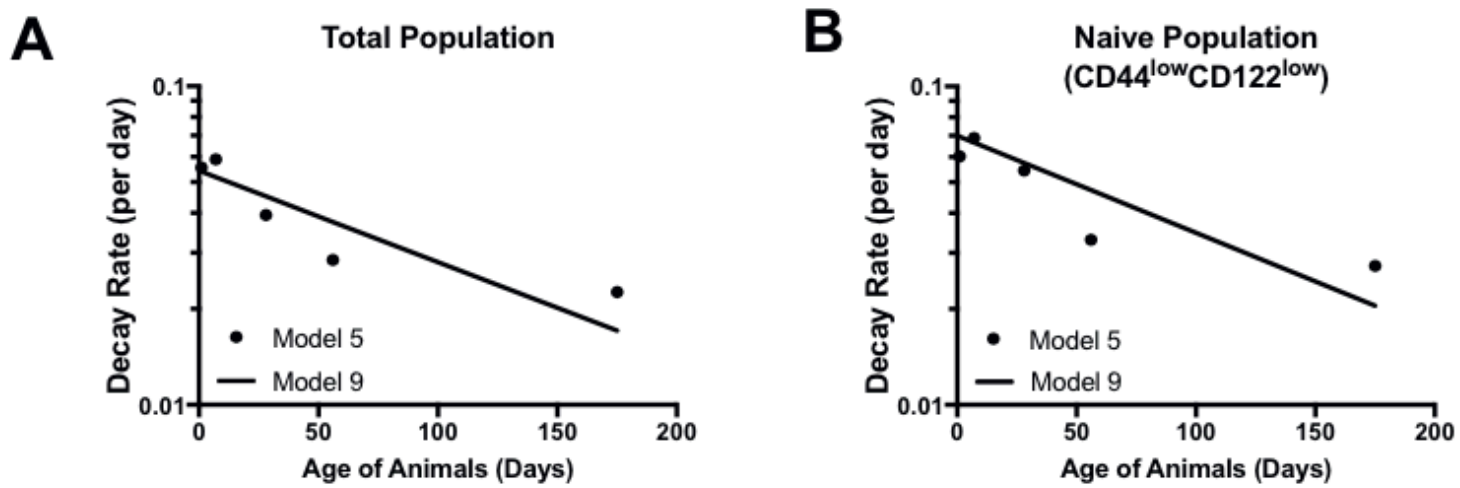
Supplementary Figure 4

The distribution of estimated parameters (k_b, δ, and s) from model 9 by fitting to the bulk population (A) and naïve population - $CD44^{low}$, $CD122^{low}$ (B) when the variation from background level and total cell number is included.



Supplementary Figure 5

Comparing model 5 and model 9 in the total (A) and naïve population (B). We found that model 5 is the best model for the bulk population (ΔAIC=0), while model 9 is the best model for the naïve population (ΔAIC=0). Model 5 has 5 different initial decay rates, one for each cohort, total of 6 parameters. Model 9 simplifies this by choosing a decay rate that varies constantly with age (3 parameters). Here, we compared model 5 (dots) vs model 9 (line) to demonstrate how the parametric form of model 9 can still capture the varying decay rates estimated in model 5 (A - total CD8 population; B – naïve population).



Supplementary Table 1. Model Comparison (Total Population and naïve phenotype)

Model	Description	Parameters	Equation	AIC-AIC_{min} (Total population)	AIC-AIC_{min} (Naive phenotype)
Model 1	Cells decay at the same rate independent of time of labeling and age.	Simple exponential decay model (eq 5) using the same decay rate for all 5 groups (one parameter, k)	$L(t) = L_0 e^{-kt}$	75.67	358.68
Model 2	Cells produced at different ages decay at different rates	Simple exponential decay model (eq 5), but each timestamp age group has a different decay rate (5 parameters, $k_1 - k_5$)	For groups 1-5, $L(t) = L_0 e^{-k_i t}$	40.57	215.69
Model 3	All groups decay at the same initial rate, but this rate slows with *time since production*.	Non-constant exponential model (eq 6) with the same initial decay rate for each group, and the same slowing down rate (with time since production) for all 5 groups (2 parameters, k and δ)	$L(t) = L_0 e^{-(ke^{-\delta t})t}$	51.21	56.3
Model 4	The rate of decay for cells produced at any age slows with the *age of the animal*	Non-constant exponential model (eq 7) with decay rate slowing down with age of animal for all 5 groups (2 parameters, k and δ)	$L(t) = L_0 e^{-(ke^{-\delta a})t}$	18.96	204.44
Model 5	There is a different initial decay rate for cells produced at different ages, but for all groups, this rate slows at the same rate with time since production.	Non-constant exponential model (eq 6) with different decay rate for each group, but the same slowing down rate for all groups (6 parameters, $k_1 - k_5$ and δ)	For groups 1-5; $L(t) = L_0 e^{-(k_i e^{-\delta t})t}$	0	4.69
Model 6	There is the same initial decay rate for cells produced	Non-constant exponential model (eq 6) with the same initial decay	For groups 1-5;	17.05	5.18

	at different ages, but this rate slows at a different rate (with time since production) for different age groups.	rate for each group, but different slowing down rate for all groups (6 parameters, k and δ_1 - δ_5)	$L(t) = L_0 e^{-(ke^{-\delta_i t})t}$		
Model 7	Different initial decay rate for cells produced at different ages, and this initial rate slows at a different rate (with time since production) for different age groups.	Non-constant exponential model (eq 6) with different decay rate and different slowing down rate for each group (10 parameters, k_1 – k_5 and δ_1 - δ_5).	For groups 1-5; $L(t) = L_0 e^{-(k_i e^{-\delta_i t})t}$	2.52	4.86
Model 8	There is a different initial decay rate for cells produced at different ages, but for all groups, this rate slows at the same rate with age of the animal	Non-constant exponential model (eq 6) with different decay rate for each group, but the same slowing down rate for all groups (6 parameters, k_1 – k_5 and δ)	For groups 1-5; $L(t) = L_0 e^{-(k_i e^{-\delta a})t}$	0.68	177.51
Model 9	The initial decay rate for cells produced at different ages is a function of the age of the animal, and this initial rate slows at the same rate (with time since production) for different age groups.	Non-constant exponential model (eq 6) with an initial rate of decay of newly produced cells at birth, and this decay rate is exponentially decreasing with the age of the animal. After production, the decay rate of cells slows down with time (at the same rate for each group) (3 parameters, k_b, s and δ)	$L(t) = L_0 e^{-(k(a)e^{-\delta t})t}$ where: $k(a) = k_b e^{-sa}$	0.11	0