# Supplementary Information: Summary of Mathematical Models

In **Table 1** we refer to the different mathematical models used to estimate T cell kinetics. Here we summarise the models.

In all equations below *bw/g* is the normalisation constant that scales from label enrichment in plasma to label enrichment in DNA for 2H2O and D2-glucose respectively (*bw*=3.2-4.9, *bg*=0.65-0.73) and *U(t)* is a function describing label enrichment in plasma.

**Kinetic heterogeneity model**

Underling cell population model for cell population *T* proliferating at rate *p* and disappearing (dying, differentiating, exiting blood long-term) at rate *d*:

 

At steady state *p*=*d.*

Fraction of labeled deoxyribose is *F:*

 

where *d\** is the disappearance rate of labeled cells (in general for a kinetically heterogeneous subpopulation *d*≠*d\* [17]*). Of the estimates reported in **Table 1** & **2**, this model was used by [13] and [16].

**Multi-exponential model**

Underling model for cell population consisting of two subpopulations *T1* and *T2* of size α and (1-α) respectively. Cells in *T1* proliferate at rate *p1* and disappear (die, differentiate, exit blood long-term) at rate *d1;* cells in *T2* proliferate at rate *p2* and disappear at rate *d2*:

 

At steady state *p1*=*d1, p2=d2*

Fraction of labeled deoxyribose in the whole population coming from subpopulation *T1* (*T2*) is *F1 (F2),* assuming steady state*:*

 

The sum *F1*+*F2* is fitted to the observed fraction of label in the population. Note, *α*, *p1* and *p2* are not identifiable which may cause problems for fitting routines; however the average proliferation *αp1+(1-α)p2* is identifiable. Of the estimates reported in **Table 1** & **2** this model was used by [27].

**Precursor model (implicit heterogeneity version)**

Underlying model for a cell population *TN* that proliferates at rate *pN*, disappears at rate *dN* and differentiates at rate *Δ*. Differentiation is accompanied by *k* rounds of clonal expansion. *TN* differentiate into *TS*. *TS* proliferate at rate *pS* and disappear at rate *dS*:

 

The fraction of labelled deoxyribose in the two populations is:

 

Where *R=TN/TS* (i.e. the relative size of the two populations), *dN\** is the disappearance rate of labelled *TN* cells and *dS\** is the disappearance rate of labelled *TS* cells.

**Precursor model (explicit heterogeneity version)**

Underlying model for a homogenous cell population *TN* that proliferates at rate *pN*, disappears at rate *dN* and differentiates at rate *Δ*. Differentiation is accompanied by *k* rounds of clonal expansion. *TN* differentiate into *TS* which is heterogeneous and consists of two subpopulations *TS1* and *TS2*. A fraction (1-*f)* differentiate into *TS1* and the remainder *(f)* differentiate into *TS2*. *TS1* proliferate at rate *pS1* and disappear at rate *dS1,* similarly for *TS2*:

 

Assuming steady state of the three populations *TN*, *TS1* and *TS2* independently, the fraction of labelled deoxyribose in the three populations is:

 

Where *R1=TN/TS1* (i.e. the relative size of the precursor population to the first subpopulation) and *R2=TN/TS2* (i.e. the relative size of the precursor population to the second subpopulation). Of the estimates reported in **Table 1** & **2** this model was used by [18].

**Relationship between the models**

The kinetic heterogeneity model and multi-exponential model both consider a single, closed population of cells with no input from a precursor population; cells can proliferate or disappear. Both models allow this population to be kinetically heterogeneous i.e. to consist of subpopulations with different kinetics. In the kinetic heterogeneity model this heterogeneity is dealt with implicitly by allowing *d*≠*d\**; in the multi-exponential model the heterogeneity is dealt with explicitly by describing the two subpopulations. The multi-exponential model suffers from parameter identifiability problems due to the strong correlation between the size of a subpopulation and its proliferation rate (i.e. a range of values of *α*, *p1* and *p2* give the same labelling curves and so are not distinguishable) this means that fitting routines will struggle to converge (there is no unique solution) and the individual parameters cannot be identified; however the average proliferation *αp1+(1-α)p2* is identifiable. In the absence of saturation the average proliferation from the multi-exponential model and the proliferation from the kinetic heterogeneity model are equivalent.

The precursor models (implicit heterogeneity version and explicit heterogeneity version) are the extensions of the kinetic heterogeneity model and the multi-exponential model respectively to describe an upstream precursor population differentiating into the target population of interest. The versions presented here were developed to describe naïve cells differentiating into TSCM cells. We showed that there was no reason to reject homogeneity for the naïve cell population but that there was strong evidence to reject it for the TSCM population [18] and so the models here assume that the precursor population is homogeneous but allow for heterogeneity (implicitly or explicitly) in the downstream TSCM population. It has been shown (Mackerodt unpub) that both models can be reparameterised to improve parameter identifiability but there are still problems with fitting the explicit version just to labelling data. In our experience it is necessary to include additional data (we used telomere length and cell numbers) to adequately constrain the parameters.