Supplementary Material for

Calibration, Selection and Identifiability Analysis of a Mathematical Model of the *in vitro* Erythropoiesis in Normal and Perturbed Contexts

Ronan Duchesne[1,2], Anissa Guillemin[1], Fabien Crauste[2,3], and Olivier Gandrillon[1,2]

[1]Laboratoire de Biologie et Modélisation de la Cellule,École Normale Supérieure de Lyon, 46 allée d'Italie, Lyon.
[2]Équipe Dracula, Inria, 56 Boulevard Niels Bohr, Villeurbanne
[3]Institut Mathématiques de Bordeaux, CNRS UMR5251, Université de Bordeaux, Talence, France

All the datasets and pieces of code analysed and generated during the current study are available in a public github repository, at `https://github.com/rduchesn/Dynamic_Model_Erythropoiesis`.

1 Solutions of the dynamic models

Here, we give the solutions of the three dynamic models presented in systems (1) to (3) of the main text, which are represented graphically in figure 2. In all the following, we will note Y_0 the initial condition at time $t = 0$ for variable Y.

1.1 SB model

System (1) is written in matrix form as:

$$\begin{pmatrix} \frac{dS}{dt}(t) \\ \frac{dB}{dt}(t) \end{pmatrix} = \begin{pmatrix} \rho_S - \delta_{SB} & 0 \\ \delta_{SB} & \rho_B \end{pmatrix} \begin{pmatrix} S(t) \\ B(t) \end{pmatrix}.$$

We thus have immediate access to the eigenvalues of the system, and can solve it analytically. There are two possible cases, whether the two eigenvalues of the matrix are equal or not.

The dynamics of the self-renewing cells are described by the equation:

$$S(t) = S_0 e^{(\rho_S - \delta_{SB})t}.$$

Concerning the dynamics of the differentiated cells, we thus have :

$$\frac{dB}{dt}(t) - \rho_B B(t) = \delta_{SB} S_0 e^{(\rho_S - \delta_{SB})t}.$$

Then, there are two cases depending on the value of ρ_B:

1. If $\rho_B \neq \rho_S - \delta_{SB}$, then B writes as:

$$B(t) = b_1 e^{\rho_B t} + b_2 e^{(\rho_S - \delta_{SB})t},$$

with $b2 = \frac{\delta_{SB} S_0}{\rho_S - \delta_{SB} - \rho_B}$ and $b_1 = B_0 - b_2$.

2. If $\rho_B = \rho_S - \delta_{SB}$, then B writes as:

$$B(t) = (b_1 + b_2 t) e^{\rho_B t},$$

with $b_2 = \delta_{SB} S_0$ and $b_1 = B_0$.

1.2 S2B model

System (2) is written in matrix form as:

$$\begin{pmatrix} \frac{dS_{LM1}}{dt}(t) \\ \frac{dS_{DM17}}{dt}(t) \\ \frac{dB}{dt}(t) \end{pmatrix} = \begin{pmatrix} \rho_{LM1} & 0 & 0 \\ 0 & \rho_{DM17} - \delta_{SB} & 0 \\ 0 & \delta_{SB} & \rho_B \end{pmatrix} \begin{pmatrix} S_{LM1}(t) \\ S_{DM17}(t) \\ B(t) \end{pmatrix}.$$

Again, the eigenvalues of the matrix of the system are its diagonal coefficients. The dynamics of the self-renewing cells still follow an exponential dynamic:

$$\begin{cases} S_{LM1}(t) = S_0 e^{\rho_{LM1} t}, \\ S_{DM17}(t) = S_0 e^{(\rho_{DM17} - \delta_{SB})t}, \end{cases}$$

and the differentiated cells follow the same dynamics as in the SB model (by replacing ρ_S by ρ_{DM17}).

1.3 SCB model

System (3) is written in matrix form as:

$$\begin{pmatrix} \frac{dS}{dt}(t) \\ \frac{dC}{dt}(t) \\ \frac{dB}{dt}(t) \end{pmatrix} = \begin{pmatrix} \rho_S - \delta_{SC} & 0 & 0 \\ \delta_{SC} & \rho_C - \delta_{CB} & 0 \\ 0 & \delta_{CB} & \rho_B \end{pmatrix} \begin{pmatrix} S(t) \\ C(t) \\ B(t) \end{pmatrix}.$$

Again, the eigenvalues of the matrix of the system are its diagonal coefficients, and we can solve it analytically. There are several possible cases, depending on what eigenvalues are equal.

For the self-renewing cells, we have the same solution as in the two other models:

$$S(t) = S_0 e^{(\rho_S - \delta_{SC})t}.$$

Concerning the dynamics of the committed cells, we thus have :

$$\frac{dC}{dt}(t) - (\rho_C - \delta_{CB})C(t) = \delta_{SC} S_0 e^{(\rho_S - \delta_{SC})t}.$$

There are two cases depending on the respective values of $\rho_C - \delta_{CB}$ and $\rho_S - \delta_{SC}$:

1. If $\rho_C - \delta_{CB} \neq \rho_S - \delta_{SC}$, then C writes as:

   $$C(t) = c_1 e^{(\rho_C - \delta_{CB})t} + c_2 e^{(\rho_S - \delta_{SC})t},$$

   with $c_2 = \frac{\delta_{SC} S_0}{(\rho_S - \delta_{SC}) - (\rho_C - \delta_{CB})}$ and $c_1 = C_0 - c2$.

2. If $\rho_C - \delta_{CB} = \rho_S - \delta_{SC}$, then C writes as:

   $$C(t) = (c_1 + c_2 t) e^{(\rho_S - \delta_{SC})t},$$

   with $c2 = \delta_{SC} S_0$ and $c_1 = C_0$.

Depending on the two previous cases, there are several possible solutions for the dynamics of differentiated cells:

1. If $\rho_C - \delta_{CB} \neq \rho_S - \delta_{SC}$, then:

$$\frac{dB}{dt}(t) - \rho_B B(t) = \delta_{CB} c_1 e^{(\rho_C - \delta_{CB})t} + \delta_{CB} c_2 e^{(\rho_S - \delta_{SC})t}.$$

Then, there are three possible sub-cases depending on the relative values of $\rho_S - \delta_{SC}$, $\rho_C - \delta_{CB}$ and ρ_B:

   (a) If $\rho_C - \delta_{CB} \neq \rho_S - \delta_{SC}$, $\rho_S - \delta_{SC} \neq \rho_B$ and $\rho_C - \delta_{CB} \neq \rho_B$, then B writes as:

$$B(t) = b_1 e^{\rho_B t} + b_2 e^{(\rho_C - \delta_{CB})t} + b_3 e^{(\rho_S - \delta_{SC})t},$$

   with $b_2 = \frac{\delta_{CB} c_1}{\rho_C - \delta_{CB} - \rho_B}$, $b_3 = \frac{\delta_{CB} c_2}{\rho_S - \delta_{SC} - \rho_B}$ and $b_1 = B_0 - b_2 - b_3$.

   (b) If $\rho_S - \delta_{SC} \neq \rho_C - \delta_{CB}$ and $\rho_B = \rho_C - \delta_{CB}$, then B writes as:

$$B(t) = (b_1 + b_2 t) e^{\rho_B t} + b_3 e^{(\rho_S - \delta_{SC})t},$$

   with $b_2 = \delta_{CB} c_1$, b3=$\frac{\delta_{CB} c_2}{\rho_S - \delta_{SC} - \rho_B}$ and $b_1 = B_0 - b_3$.

   (c) If $\rho_S - \delta_{SC} \neq \rho_C - \delta_{CB}$ and $\rho_B = \rho_S - \delta_{SC}$, then B writes as:

$$B(t) = (b_1 + b_3 t) e^{(\rho_S - \delta_{SC})t} + b_2 e^{(\rho_C - \delta_{CB})t},$$

   with $b_2 = \frac{\delta_{CB} c_1}{\rho_C - \delta_{CB} - \rho_B}$, $b_3 = \delta_{CB} c_2$ and $b_1 = B_0 - b_2$.

2. If $\rho_S - \delta_{SC} = \rho_C - \delta_{CB}$, then:

$$\frac{dB}{dt}(t) - \rho_B B(t) = \delta_{CB}(c_1 + c_2 t) e^{(\rho_S - \delta_{SC})t}.$$

Then, there are two possible sub-cases depending on the relative values of $\rho_S - \delta_{SC}$, $\rho_C - \delta_{CB}$ and ρ_B:

   (a) If $\rho_S - \delta_{SC} = \rho_C - \delta_{CB}$ and $\rho_B \neq \rho_S - \delta_{SC}$, then B writes as:

$$B(t) = b_1 e^{\rho_B t} + (b_2 + b_3 t) e^{(\rho_S - \delta_{SC})t},$$

   with $b_3 = \frac{\delta_{CB} c_2}{\rho_S - \delta_{SC} - \rho_B}$, $b_2 = \frac{\delta_{CB} c_1 - b_3}{\rho_S - \delta_{SC} - \rho_B}$ and $b_1 = B_0 - b_2$.

   (b) If $\rho_S - \delta_{SC} = \rho_C - \delta_{CB} = \rho_B$, then B writes as:

$$B(t) = (b_1 + b_2 t + b_3 t^2) e^{\rho_B t},$$

   with $b_2 = \delta_{CB} c_1$, $b_3 = \frac{\delta_{CB} c_2}{2}$ and $b1 = B_0$

2 Convergence of the estimation

In both estimation steps, we minimized the -log likelihood with the Truncated Newton's algorithm implemented in scipy, with a maximum number of function evaluations of 10^6. We used random sampling of the initial guesses for parameter values to assure convergence to the global optimum.

Figure S1 shows the distance to the minimal log-likelihood (sorted from highest to lowest) over large samples of initial guesses (200 initial guesses for the first step, 1000 initial guesses for the second one) for our SCB model with proportional error. It shows that with a relatively small sample, the estimated likelihood is already quite close to its minimal value and that increasing the sample size doesn't result in a better fit.

In order to balance the quality of the fit with the computational cost of the estimation, we used 100 different initial guesses for the first estimation step and 500 initial guesses for the second one for all of our models.

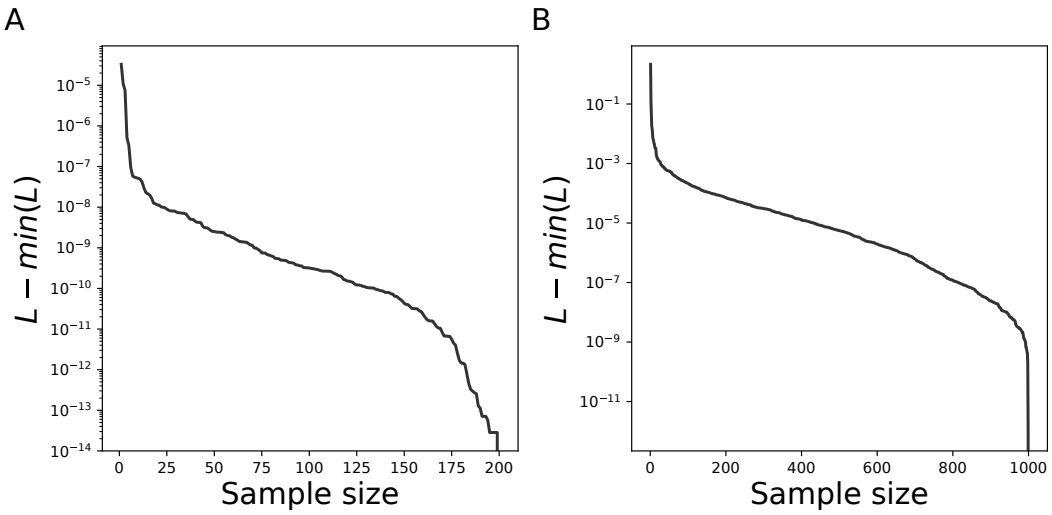

Figure S1: Influence of the initial guess sample size on the estimation. A: Sorted distance to the optimal likelihood over 200 runs of the estimation of ρ_S and b_1 (SCB model with proportional error). B: Sorted distance to the optimal likelihood over 1000 runs of the estimation of ρ_C, δ_{CB}, ρ_B and b_2 (SCB model with proportional error).

3 Parameter values

The estimated parameter values for the 9 pairs of dynamic and error models are displayed in table S1

Dynamic	Error	ρ_1	δ_1	a_1	b_1	ρ_2	δ_2	ρ_B	a_2	b_2
SB	Constant	0.59	5.7	2.8×10^4	-	-	-	0.62	5.4×10^4	-
SB	Proportional	0.53	5.6	-	0.34	-	-	0.56	-	0.45
SB	Combined	0.53	5.6	0	0.34	-	-	0.56	0	0.45
S2B	Constant	0.59	5.7	2.8×10^4	-	0.48	0.033	1.4	1×10^4	-
S2B	Proportional	0.53	5.6	-	0.34	0.47	0.15	0.96	-	0.14
S2B	Combined	0.53	5.6	0	0.34	0.47	0.15	0.96	0x	0.14
SCB	Constant	0.59	5.7	2.8×10^4	-	0.47	0.038	1.4	1×10^4	-
SCB	Proportional	0.53	5.6	-	0.34	0.49	0.18	0.92	-	0.15
SCB	Combined	0.53	5.6	0	0.34	0.49	0.18	0.92	0	0.15

Table S1: Best-fit estimates of the parameters for the 9 pairs of error model and dynamic model. ρ_1 is the net proliferation rate of the first compartment of the model (*i.e.* ρ_S for the SB and SCB models, and ρ_{LM1} for the S2B model). δ_1 is the differentiation rate of this compartment (*i.e.* δ_{SB} for the SB model, δ_{SC} for the SCB model, and it is not defined for the S2B model). ρ_2 is the net proliferation rate of the second compartment, when it is different from the B compartment (*i.e.* ρ_{DM17} for the S2B model, ρ_C for the SCB model, and it is not defined for the SB model). δ_2 is its differentiation rate (*i.e.* δ_{SB} for the S2B model, δ_{CB} for the SCB model, and it is not defined for the SB model). ρ_B is defined in every model as in the SCB model.

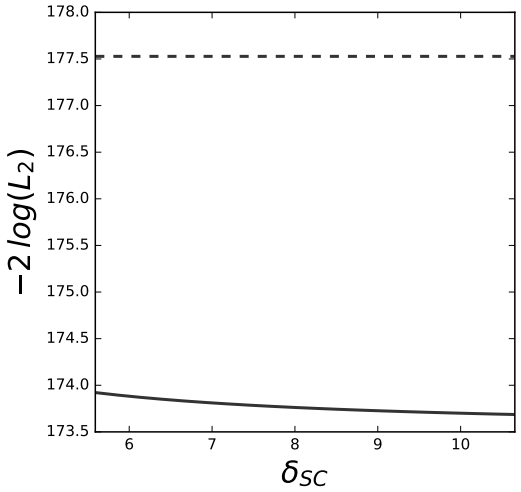

Figure S2: The choice of δ_{SC} does not impact the quality of the fit of the model. The straight line represents the minimum -log likelihood optimized in the last step of our estimation procedure as a function of the chosen value for δ_{SC}. The dashed line gives the χ^2 significance threshold of a likelihood-ratio test.

4 Importance of δ_{SC}

In the estimation procedure described in the Methods section, every parameter of the dynamic model is estimated according to the experimental data, except δ_{SC}, which is set to an arbitrary value between its two bounds (determined from the commitment experiment pictured on figure 1C). Though these bounds give precise limits to the values that δ_{SC} can take, setting it to different values might result in different optimal parameter values for the second estimation step of the procedure.

Figure S2 displays the likelihood of the second estimation step for the range of values that δ_{SC} can take. It does not vary significantly in this range, which means that the quality of the fit of the model is not influenced by the choice of the value of δ_{SC}. This leaves two possible scenarios: either the value of δ_{SC} has no influence on the estimated parameter values in the second estimation step, or it is possible to keep the likelihood high while changing the value of δ_{SC} by adjusting the value of the other parameters (which would be a case of non-identifiability if the value of δ_{SC} was optimized to fit the data).

Figure S3 displays the values of the parameters estimated in the step 3 of the procedure, for the range of values that δ_{SC} can take. Since these value do not vary much, it seems reasonable to say that the choice of the value of δ_{SC} does not impact the last estimation step of the procedure.

5 Identifiability of the treatment parameters

Figure S4 displays the profile likelihood curves for the parameters that are affected by the rapamycin treatment. In the model of rapamycin treatment that we selected, b_1 does not vary, so its value is not estimated from the data. It is also the case for δ_{SC}, as in the control situation. It is thus impossible to define a profile likelihood for these two parameters. Thus, ρ_S is now the only parameter that is estimated with the LM1 data, so its identifiability threshold is $\chi^2(0.95, 1) = 3.84$. The other parameters (ρ_C, δ_{CB}, ρ_B and b_2) are all estimated together, so their identifiability threshold is $\chi^2(0.95, 4) = 9.49$, as in the control case.

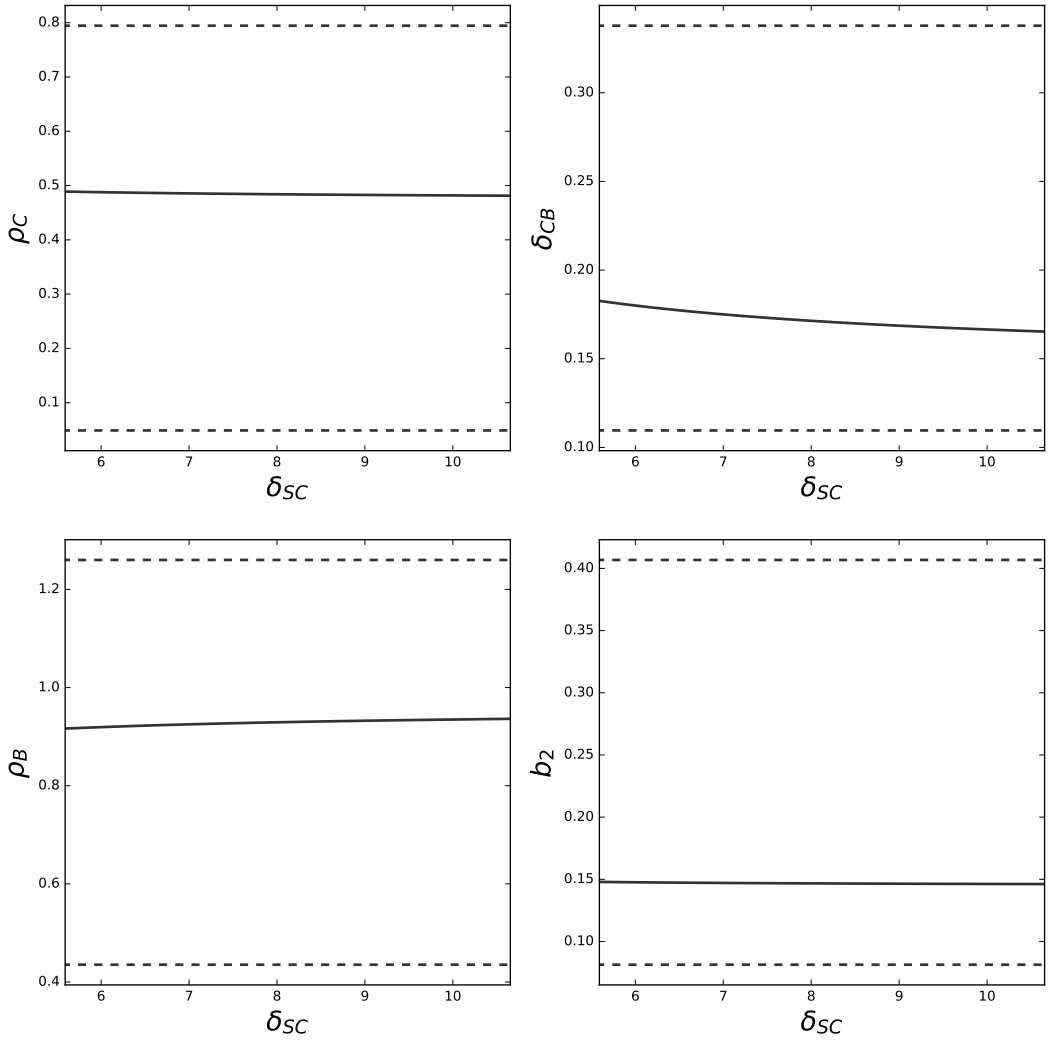

Figure S3: The choice of δ_{SC} does not impact the value of the other parameters of the model. Solid lines represent the values of the parameters estimated in the last step of our procedure. Dashed lines represent the confidence interval of each parameter computed from figure 3.

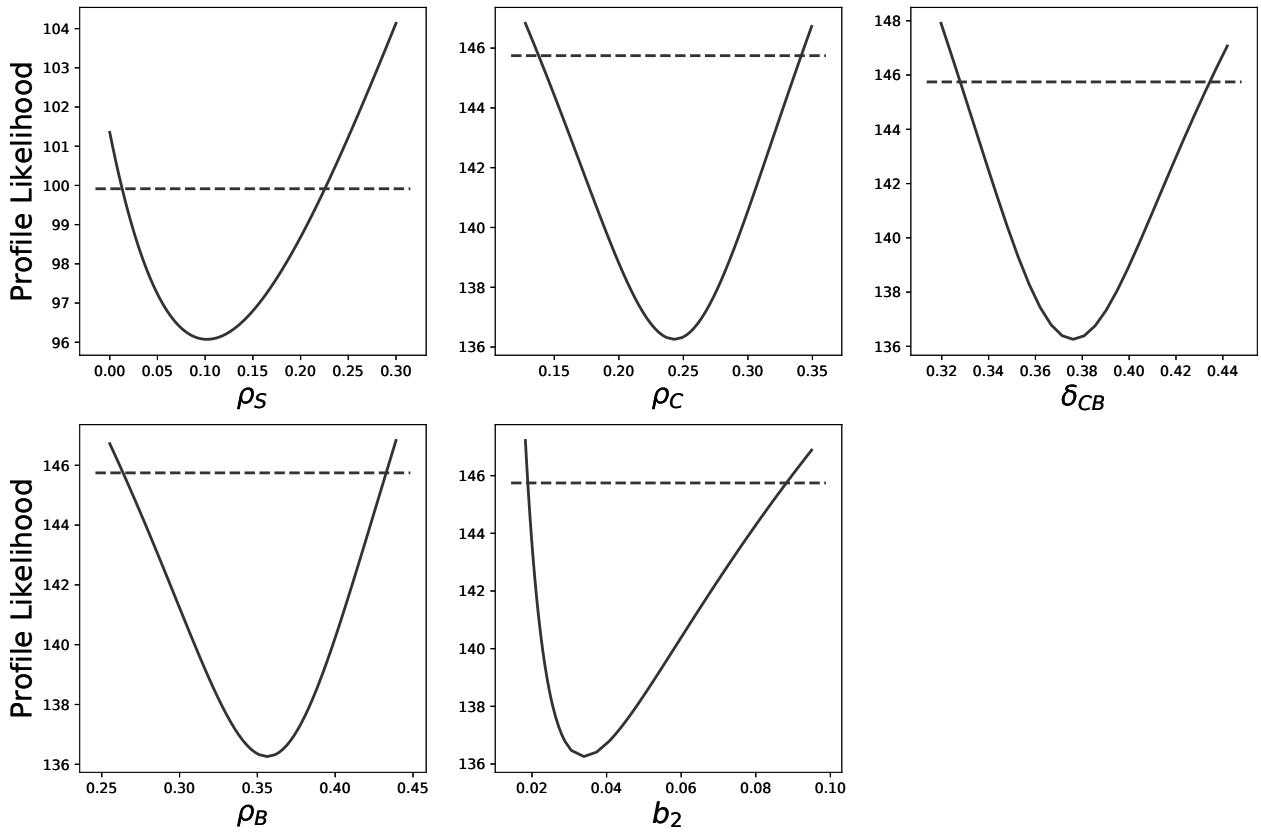

Figure S4: The values of the parameters under rapamycin treatment are identifiable. For each of the parameters which varied under the treatment, the solid line represents the profile likelihood with respect to that parameter, and the dashed line gives the χ^2 identifiability threshold (at $\alpha = 0.95$).