

# Selection pressure exerted by imatinib therapy leads to disparate outcomes of imatinib discontinuation trials

Min Tang,<sup>1\*</sup> Jasmine Foo,<sup>2\*</sup> Mithat Gönen,<sup>3</sup> Joëlle Guilhot,<sup>4</sup> François-Xavier Mahon,<sup>5</sup> and Franziska Michor<sup>1</sup>

<sup>1</sup>Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; <sup>2</sup>School of Mathematics, University of Minnesota, Minneapolis, MN, USA; <sup>3</sup>Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; <sup>4</sup>INSERM Centre d'Investigation Clinique CIC 0802, Centre Hospitalier Universitaire de Poitiers, France; <sup>5</sup>Laboratoire d'Hématologie et Service des Maladies du Sang, CHU de Bordeaux, Bordeaux, France, Université Victor Ségalen Bordeaux 2, Bordeaux, France, and INSERM U876, Bordeaux, France

Citation: Tang M, Foo J, Gönen M, Guilhot J, Mahon F-X, and Michor F. Selection pressure exerted by imatinib therapy leads to disparate outcomes of imatinib discontinuation trials. *Haematologica* 2012;97(10):1553-1561. doi:10.3324/haematol.2012.062844

## Online Supplementary Appendix

### The basic mathematical model

We utilized a mathematical model of the treatment response of chronic myeloid leukemia (CML) cells to imatinib therapy,<sup>1,2</sup> which describes four layers of the differentiation hierarchy of the hematopoietic system. Stem cells give rise to progenitors, which produce differentiated cells, which in turn produce terminally differentiated cells. This hierarchy applies to both normal and leukemic cells. Only stem cells have the potential for indefinite self-renewal, but progenitor and differentiated cells have the capability to undergo limited reproduction, which, together with differentiation, leads to an expansion of the cell number at each level of the differentiation hierarchy. The BCR-ABL oncogene is present in all leukemic cells, leading to slow clonal growth of leukemic stem cells and accelerating the rate at which leukemic progenitors and differentiated cells are generated. Imatinib therapy reduces the production rates of leukemic progenitors and differentiated cells, and potentially also inhibits the expansion of leukemic stem cells.

The abundances of normal hematopoietic stem cells, progenitors, differentiated cells, and terminally differentiated cells can be denoted by  $x_0, x_1, x_2,$  and  $x_3$ . Their respective leukemic abundances are given by  $y_0, y_1, y_2,$  and  $y_3$ . We assume that homeostatic mechanisms maintain the hematopoietic stem cell population at a constant level and, therefore, introduce a density dependence term,  $\phi$ , in the stem cell production rate. Leukemic stem cells grow at a slow pace until reaching their maximum number, which may be larger than that of normal stem cells; afterwards, their number is also held constant by a density dependence mechanism. The system containing stem cells (SC), progenitor cells (PC), differentiated cells (DC) and terminally differentiated cells (TC) is then described by:

	healthy cells	leukemic cells
SC	$\dot{x}_0 = [r_x \phi - d_0] x_0$	$\dot{y}_0 = [r_y \phi - d_0] y_0$
PC	$\dot{x}_1 = a_x x_0 - d_1 x_1$	$\dot{y}_1 = a_y y_0 - d_1 y_1$
DC	$\dot{x}_2 = b_x x_1 - d_2 x_2$	$\dot{y}_2 = b_y y_1 - d_2 y_2$
TC	$\dot{x}_3 = c_x x_2 - d_3 x_3$	$\dot{y}_3 = c_y y_2 - d_3 y_3$

Here density dependence in the stem cell population is given by  $\phi = 1/[1+p_x(x_0 + y_0)]$  and  $\varphi = 1/[1+p_y(x_0 + y_0)]$ . The potentially dif-

ferent carrying capacities of normal and leukemic stem cells are represented by the parameters  $p_x$  and  $p_y$ . Imatinib dramatically reduces the differentiation rates of cells,  $a_x$  to  $a_x'$  and  $b_x$  to  $b_x'$ . This change in rates leads to a bi-phasic decline of the leukemic cell burden.<sup>1</sup> The parameters during imatinib therapy are denoted by  $r_x', a_x', b_x'$  etc., and the parameters after cessation of imatinib therapy are denoted by  $r_x'', a_x'', b_x''$  etc. For comparison with experimental polymerase chain reaction (PCR) data, we calculate the BCR-ABL to ABL ratio as  $y_3/(2x_3 + y_3)$  times 100%; a healthy cell has two copies of ABL, while a leukemic cell normally has one copy of ABL and one copy of BCR-ABL. Since most cells that are sampled by the PCR assay are terminally differentiated cells, the calculation of the BCR-ABL/ABL ratio using the mathematical framework includes the abundance of terminally differentiated cells only. The time of loss of a complete molecular response (CMR) is defined as the time at which the BCR-ABL/ABL transcript level exceeds  $10^{-5}$ : i.e. when  $\frac{y_3}{2x_3 + y_3} > 10^{-5}$ .

This definition of the BCR-ABL/ABL transcript level when reaching CMR can be varied to obtain a different cut-off; if a different cut-off is chosen, then the estimation of the parameter values that replicate the trial dynamics (see *Statistical analysis section*) would lead to slightly different numerical results, but the basic conclusions would hold. This system of equations was numerically solved using a fourth-order Runge-Kutta integration scheme.

### Extensions of the mathematical model

We investigated the possibility of an explicitly heterogeneous leukemic stem cell population within each patient as a possible mechanism for the imatinib selection effect. To this end, we considered a system with two leukemic stem cell populations: type-1 leukemic stem cells having an intrinsic growth rate  $r_{y,1}$  and type-2 leukemic stem cells having a growth rate  $r_{y,2}$  both before and after treatment with imatinib ( $r_{y,1}'' = r_{y,1}, r_{y,2}'' = r_{y,2}$ ). The two cell populations have distinct growth kinetics,  $r_{y,2} < r_{y,1}$ , so that at diagnosis and the start of treatment the majority of leukemic stem cells are of type-1. To bear out the imatinib selection hypothesis, we assumed that the effect of imatinib is stronger on the more aggressive type-1 cells than the more indolent type-2 cells such that  $r_{y,1}' < r_{y,2}'$ . *Online Supplementary*

Figures S1 and S2 demonstrate the effect of this hypothesis on the composition of the stem cell population. *Online Supplementary Figure S1A* shows the terminally differentiated leukemic cell population over time for a system with a two-type heterogeneous leukemic stem cell compartment as described above, as well as for a homogeneous stem cell population. Note that heterogeneity in cell phenotypes is also included in the basic mathematical model incorporating only a single type of stem cells, since the kinetics of the stem cell population can be described by a distribution of growth and differentiation rates rather than a single value. *Online Supplementary Figure S1B* shows the sizes of the leukemic stem cell population in both cases. Note that effects of a heterogeneous stem cell compartment on the observable quantities (terminally differentiated cells) may be minimal (*Online Supplementary Figure S1A*), even though dramatic differences may be taking place at the level of leukemic stem cells (*Online Supplementary Figure S1B*). *Online Supplementary Figure S2* shows the composition of the leukemic stem cell compartment in the heterogeneous system, both before imatinib treatment at detection and after treatment at the time of relapse. In this system, the effect of imatinib is to select a less aggressive population of stem cells, which was undetectable prior to treatment. This subpopulation comprises the majority of the leukemic stem cell compartment after treatment, therefore lowering the effective stem cell growth rate.

### The parametric cure model

Let us now consider the scenario in which imatinib has the potential to eradicate leukemic stem cells in some patients and CML can be cured by imatinib therapy alone. The possibility of cure is equivalent to the assumption that the population relapse time distribution is improper – i.e. as time goes to infinity, the probability of relapse is less than 1.

To investigate this scenario, we first fitted parametric cure models to the data, assuming that the cure rate  $p$  is greater than zero. Based on the cure model, the population of patients is made up of a mixture of two sub-populations, one that can be cured of their disease and one that will relapse sooner or later after imatinib discontinuation. We call subjects in the sub-population that will always relapse the ‘susceptibles’. The survival function of the relapse time under the assumption of a cure model is a mixture of a parametric survival function for the susceptibles and a cure mass, i.e.  $S(t) = p + (1 - p) S_0(t)$ , where  $S_0(t)$  denotes the survival function of the susceptibles and  $p$  represents the cure rate. We then used the maximum likelihood estimation method to fit the cure model to the clinical data and chose for  $S_0(t)$  the most commonly used distributions including Weibull, lognormal, exponential, logistic and log logistic distributions. Of these distributions, the cure model with  $S_0(t)$  being the survival function of the log logistic distribution exhibited the best fit with both the largest likelihood and lowest AIC (Akaike Information Criterion<sup>3</sup>) (Figure 1B). The estimated cure rate based on this log logistic cure model was 0.41.

### Testing for sufficient follow-up

To make any conclusions regarding the cure of some of the patients enrolled in the STIM trial, we tested whether there were clinical data from a sufficiently long follow-up. We utilized the non-parametric method, i.e. no assumptions were

made on the type or shape of the underlying survival or censoring distributions, to estimate the cure rate  $p$  and tested whether the follow-up data of the STIM trial were sufficient. Throughout this paper, we assumed random (or non-informative) censoring, i.e. each subject’s censoring time is statistically independent of the failure time.

We first obtained the non-parametric estimator of the cure rate for the STIM data and then tested whether there was sufficiently long follow-up to identify the existence of a cure, based on the measurement of the distance between the largest observed time and the largest uncensored relapse time. To be specific, let  $n$  be the total number of observations;  $t_{(n)}$  be the largest observed time (relapse or loss of follow-up); and  $\hat{S}(t)$  be the Kaplan-Meier Estimator (KME) of the survival function for the relapse time. For the STIM trial data,  $n = 100$  and  $t_{(n)} = 911$ . The non-parametric estimator of the cure rate  $p$  is then given by  $\hat{p} = \hat{S}(t_{(n)}) = 0.39$ .<sup>4</sup> We then applied the procedure outlined in Section 4.3 of the article by Maller *et al.*<sup>4</sup> to test for the existence of a sufficiently long follow-up, which is based on the measure of the distance between the largest observed time  $t_{(n)}$  and the largest uncensored relapse time  $t_{(n)}^*$ . Let  $N_n$  be the number of uncensored observations in the interval  $(2 t_{(n)}^* - t_{(n)}, t_{(n)}^*]$  and  $q_n = N_n / n$  be the proportion of uncensored observations in that interval. Following arguments in Section 4.3 of the article by Maller *et al.*<sup>4</sup> large values of  $q_n$  generally lead to the conclusion that there was sufficiently long follow-up while small values of  $q_n$  represent insufficient follow-up. More specifically, if the observed value of  $q_n$  is greater than the 95% quantile of the simulated distribution of  $q_n$ , then there is strong evidence that there was sufficiently long follow-up in the sample. If, on the other hand, the observed  $q_n$  is less than the 5% quantile of the simulated distribution, then there is good evidence that follow-up is insufficient. If, however, the observed value of  $q_n$  resides between the 10% and 90% quantiles, then we conclude that the data are inconclusive. In that case, it is doubtful that data have leveled off sufficiently to make any conclusions about the existence of a cure, but there is also no strong evidence that the follow-up is insufficient. For the clinical data utilized here,  $t_{(n)}^* = 678$ ;  $(2 t_{(n)}^* - t_{(n)}, t_{(n)}^*] = (445, 678]$ ;  $N_n = 2$  and  $q_n = 0.02$ .

We then conducted simulations to obtain the distribution of  $q_n$ , where survival times were generated from a log logistic distribution (the log logistic distribution was chosen because as outlined in the above section, it exhibited the best fit with both the largest likelihood and lowest AIC) for a number of observations of  $n = 100$  and cure rate of  $p = 0.39$ . We simulated censoring times using a sampling with replacement procedure with samples representing the censoring times in the STIM trial data. We also simulated censoring times from a uniform distribution over the range of the censoring times in the STIM trial data and reached the same conclusions. With 100,000 replicates, the 5%, 10%, 90% and 95% percentiles of the simulated  $q_n$  values for this simulation setting were 0.01, 0.01, 0.59 and 0.59, respectively. Since  $q_n$  for the STIM trial data is 0.02, falling between the 10% and 90% percentiles, we thus reached the conclusion that it is doubtful that the STIM data had leveled off sufficiently to make any conclusions about the existence of a cure. Worded differently, there was no strong evidence for either sufficient or insufficient follow-up in the STIM trial data, which represents an interim report of this clinical trial. When longer

follow-up becomes available, we will repeat the above procedure to test for the existence of sufficient follow-up in the updated data.

### Statistical analyses

Our mathematical model represents between-patient heterogeneity via variability in patient-specific cell growth and differentiation kinetics. Two patients with identical parameter values will have identical model-predicted cell growth profiles over time. Given a particular set of parameter values, the mathematical model can be computationally solved to evaluate the resulting relapse time; however, from a given relapse time it is not possible to determine a unique set of corresponding parameter values. In addition, the characteristics of the underlying parameter distributions for the growth and differentiation kinetics are unknown. For these reasons, we utilized a retrospective statistical approach to determine the parameter distributions given the observed relapse time data.

Since our goal was to identify the set of parameters that are consistent with the observed distribution of the time of CMR loss, we needed a complete estimate for the distribution of CMR loss times (i.e, defined for all times). Since traditional Kaplan-Meier estimates are defined up to the maximum event time, we were unable to use them. Instead we utilized the maximum likelihood estimation method to fit a parametric distribution to the observed data and chose between the most commonly used distributions including Weibull, lognormal, exponential, logistic and log logistic distributions. Of these distributions, the lognormal distribution exhibited the best fit with both the largest likelihood and lowest AIC<sup>3</sup> for the non-cure model (Figure 1C) while the log logistic distribution exhibited the best fit for the cure model (Figure 1B), as outlined in the section above.

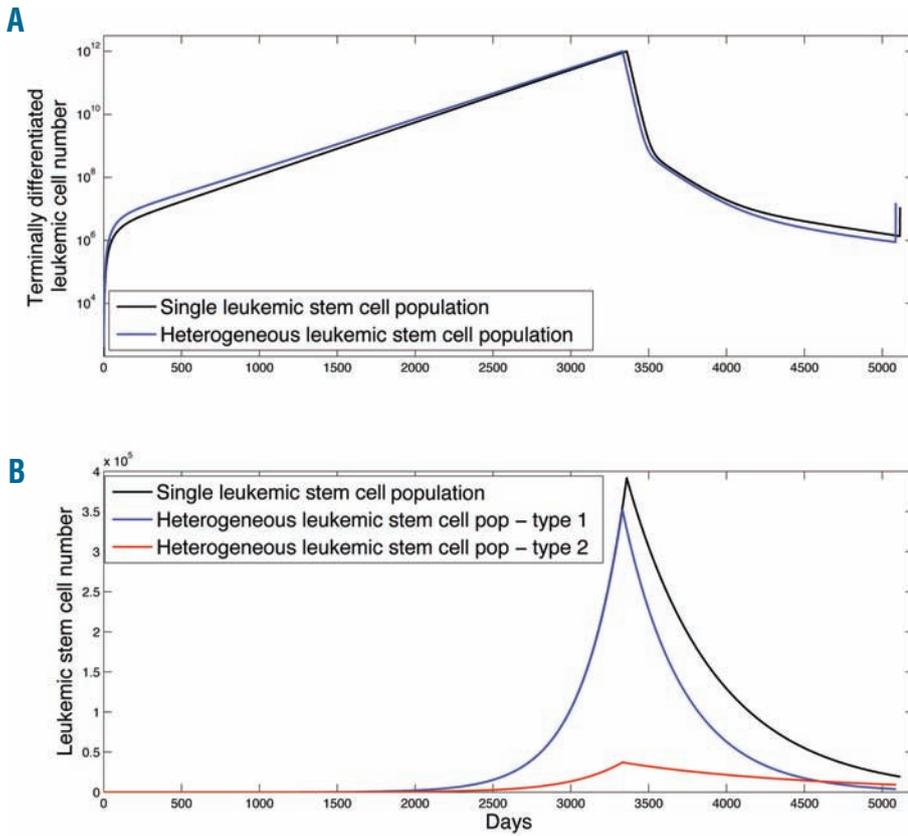
We then numerically solved the differential equation system to obtain the relapse time for different parameter sets. A fine grid was used to sample the five-dimensional parameter space, and the corresponding times of loss of CMR were determined using the mathematical model. We then selected randomized subsets (denoted as S.time) of outcomes from these times that recapitulated the fitted lognormal (for the non-cure model) or the fitted log logistic (for the cure model) survival curves up until the maximum follow-up. The set S, which contains parameter vectors resulting in the above selected survival times

in the set S.time via the mathematical model, then represents a set of samples from the joint density of the five parameters. To compare the Kaplan-Meier curve obtained from our sampling procedure to the Kaplan-Meier survival curve of the clinical data, the one-sample log-rank test as implemented in the `survdiff` function in R was used.

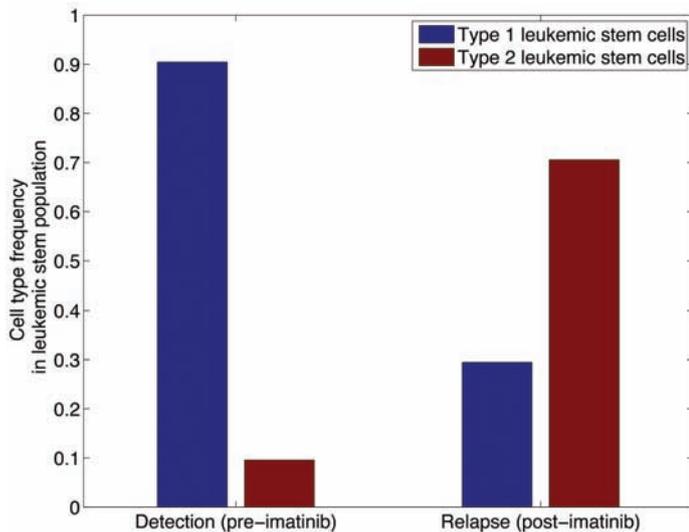
More specifically, we defined possible ranges of the five parameters, based upon our mathematical model:  $a_y$  varies in the range (0, 1.6],  $b_y$  in (0, 10],  $c_y$  in (0, 200],  $r_y'$  in [0,0.003], and  $r_y$  in [0, 0.016], with the restrictions that  $a_y \leq b_y \leq c_y$  and  $r_y' \leq r_y$ . A fine grid (over 260 million samples) was used to sample the five-dimensional parameter space from these ranges. Let  $G$  denote this grid. For each  $g \in G$ , we numerically solved the differential equation system to obtain the relapse time and the censoring indicator ( $T_g, C_g$ ). The case of  $C_g = 1$  denotes the situation in which CMR is lost; the corresponding  $T_g$  represents the relapse time. The case  $C_g = 0$  denotes the situation in which CMR is not reached; then the corresponding  $T_g$  represents the last follow-up time of the patient. Let  $P$  denote the set of ( $T_g, C_g$ ) pairs obtained through this process. We then tagged each relapse time in  $P$  according to its corresponding 18-quantile in the fitted density and considered it for acceptance into the sample set S.time. To ensure that S.time correctly recapitulates the fitted lognormal (for the non-cure model) versus log logistic (for the cure model) density function, samples were drawn from each bin uniformly at random, and a set of samples was accepted into the set S.time proportionally according to the contribution of each quantile to the density. The final set S, which contains parameter vectors corresponding to relapse time points in the set S.time, behaves like a random sample from the joint distribution of the parameter estimates. The marginal densities of each parameter were estimated using the non-parametric kernel density estimation technique (a.k.a. Parzen-Rosenblatt window method<sup>5</sup>). Multiple resamplings of the set S resulted in little or no change in the marginal densities, indicating robustness of the resulting distributions to this procedure. A randomly obtained sample of points chosen from a uniform distribution could have been used instead of the regular grid  $G$  for the purposes of obtaining ( $T_g, C_g$ ). Using a regular grid is similar to latin hypercube sampling<sup>6</sup> and is more efficient than random sampling in most computational problems that require repeated evaluations from a sample space.<sup>7</sup>

### References

1. Michor F, Hughes TP, Iwasa Y, Branford S, Shah NP, Sawyers CL, et al. Dynamics of chronic myeloid leukaemia. *Nature*. 2005;435(7046):1267-70.
2. Michor F. The long-term response to imatinib treatment of CML. *Br J Cancer*. 2007;96:679-80.
3. Akaike H. A new look at the statistical model identification. *Institute of Statistical Mathematics*. 1974;19(6):716-23.
4. Maller R, Zhou X. *Survival Analysis with Long-Term Survivors*. 1st edn: Wiley, 1996.
5. Silverman B. *Density Estimation for Statistics and Data Analysis*. 1st edn: Chapman & Hall, 1986.
6. McKay MD, Beckman RJ, Conover WJ, Conover WJ. A Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 1979;21(2):239-45.
7. Niederreiter H. *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics. 1992.



**Online Supplementary Figure S1.** The effects of a heterogeneous leukemic stem cell population. (A) Terminally differentiated leukemic cell population over time, for a homogeneous leukemic stem cell compartment ( $r_y = 0.008$ ,  $r'_y = 0.0015$ ), and a heterogeneous two-type leukemic stem cell compartment ( $r_{y1} = 0.008$ ,  $r_{y2} = 0.007$ ,  $r'_{y1} = 0.0005$ ,  $r'_{y2} = 0.0025$ ). For both systems,  $a_y = 0.8$ ,  $b_y = 5$ ,  $c_y = 100$ , and all other parameters are the same as throughout the paper. (B) Leukemic stem cell population over time for a homogeneous stem cell compartment and for each type in a heterogeneous stem cell compartment. Parameters are as in (A).



**Online Supplementary Figure S2.** The time evolution of a heterogeneous leukemic stem cell population. Composition of leukemic stem cell compartment (frequency of type-1 and type-2 cells) before imatinib treatment and at the time of relapse (post-imatinib). Parameters are as in *Online Supplementary Figure S1*.