



Mathematical modeling of therapy failure due to drug resistance in gastrointestinal stromal tumor metastasis to the liver

Patricio Cumsille*

Basic Sciences Department, Universidad del Bío-Bío, Chillán, Chile and
Centre for Biotechnology and Bioengineering (CeBiB), Santiago, Chile.

Abstract

In this work we develop a general mathematical model and devise a practical identifiability approach for gastrointestinal stromal tumor (GIST) metastasis to the liver, with the aim of quantitatively describing therapy failure due to drug resistance. To this end, we have modeled metastatic growth and therapy failure produced by resistance to two standard treatments based on tyrosine kinase inhibitors (Imatinib and Sunitinib) that have been observed clinically in patients with GIST metastasis to the liver. The parameter identification problem is difficult to solve since there are no general results on this issue for models based on ordinary differential equations (ODE) like the ones studied here. We propose a general model based on ODE systems and analyzed five variants, using medical image observations (CT scans) from patients that exhibit drug resistance. The associated parameter estimation problem was solved using the Nelder-Mead simplex algorithm, by adding a regularization term to the objective function to address model instability. From the model variants analyzed, we identified the one that provides the best fit to all the available patient data sets. This is the first work that reports mathematical models capable of capturing therapy failure due to drug resistance based on clinical images in a patient-specific manner.

In collaboration with:

Matías Godoy¹, Centre for Biotechnology and Bioengineering (CeBiB), Santiago, Chile.

Ziomara P. Gerdtzen², Department of Chemical Engineering, Biotechnology and Materials, University of Chile and Centre for Biotechnology and Bioengineering (CeBiB), Santiago, Chile.

Carlos Conca³, Department of Mathematical Engineering (DIM), Center for Mathematical Modeling (CMM), University of Chile, (UMI CNRS 2807) and Centre for Biotechnology and Bioengineering (CeBiB), Santiago, Chile.

*Partially funded by CeBiB, grant no. PIA-CONICYT FB-01; Group of Investigation in Tumor Angiogenesis (GIANT), grant no. DIUBB-GI 171709/VC, and regular research project DIUBB 193409 3/R, to PC; and Fondef, grant no. ID18110308 to ZG. e-mail: pcumsille@ubiobio.cl

¹e-mail: mgodoy@dim.uchile.cl

²e-mail: zgerdtze@ing.uchile.cl

³e-mail: cconca@dim.uchile.cl



1 Introduction

Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal tumors of the gastrointestinal tract, with an incidence of 11-15 cases per million people per year. It is estimated that 40-50% of GISTs are biologically malignant and have spread to the liver or peritoneum at the time of diagnosis or primary surgery [1]. One of the molecular characteristics of these neoplasms is a gain of function mutation in the receptor tyrosine-kinase protein (KIT) (75-80% of cases) or the homologous receptor tyrosine kinase, platelet-derived growth factor receptor alpha (PDGFRA), accounting for 85-90% of gastrointestinal stromal tumors [2]. In addition to the primary mutation, secondary mutations have also been identified in patients with advanced GIST pretreated with tyrosine kinase inhibitor. To date, ten different molecular subsets of GIST with different molecular alterations have been reported [1]. For most cases of resectable/non-metastatic GISTs cases treatment involves surgical resection and tyrosine kinase inhibitor (TKI) therapy may be utilized to reduce tumor size before resection.

For metastatic or non-resectable GISTs the treatment of choice is TKI therapy [2]. Imatinib is utilized as the first-line drug as it acts best on the most frequent KIT mutations. In 85% of the cases Imatinib can control the metastatic disease during a 20-24 months period. After resection, adjuvant Imatinib therapy has also been found to improve recurrence-free and overall survival. However, as reported by Blay [3], Imatinib resistance is frequently observed. This resistance is associated to the specific exon where the mutation occurs.

Sunitinib or Sorafenib is a tyrosine kinase inhibitor molecule that targets KIT and has antiangiogenic effects, which is utilized for the treatment of advanced GISTs in patients who fail Imatinib therapy. The treatment and prognosis of patients with GISTs depend on the oncogenic kinase mutations that caused it and the utilization of specific molecular therapies that inhibit this molecular defect. However, GISTs include several different molecular subtypes that vary in their response to kinase inhibitors. Therefore, it is crucial to correctly identify the tumor's response to treatment to assess a suitable treatment timely. For clinicians, one critical challenge is to optimize cancer treatments and to determine the more adequate time to switch from the first-line to the second-line treatment for increased overall survival. To do this, relapse time estimation is a critical issue [4]. Given that prognosis and sensitivity to treatment are patient-dependent, we aim at developing patient-dependent mathematical models based on medical images of liver metastasis. We focus on locally advanced GISTs to quantitatively describe for each patient the time of emergence of mutations in cancer cells, as well as the relapse time after the first-line and the second-line treatments.

Mathematical modeling has been extensively utilized in recent years to shed light on cancer progression, emphasizing the issue of rendering patient specific models [4, 5, 6]. However, due to the complexity of the processes involved in all the stages of neoplastic growth, mathematical models must be limited to a few phenomena, and are therefore a simplification of what occurs in the biological system. The key task is then to develop mathematical models that are able to capture most of the relevant features of cancer progression. In this type of models parameter estimation becomes an important problem that requires a rational experimental design and clinical data collection. Even though a mathematical model of GIST metastasis to the liver, growth and therapy failure associated to drug resistance is available [4], the latter has not been described quantitatively using mathematical

models and considering clinical images in a patient-specific manner. Typically, mathematical models utilized for clinical applications in this field do not consider the spatial aspect of tumor growth. They are often parametrized using statistical methods and may provide a prognosis of tumor volume, among other important aspects [5]. Considering this, we aim at developing patient-specific models that can capture the evolution of the metastatic tumor, as well as quantitatively describe and explain therapy failure due to drug resistance. Clinical follow-up to monitor the disease evolution is mainly performed by CT scans. Using observations from these medical images we applied a hybrid approach [5] to develop patient-specific mathematical models with the purpose of quantitatively describing therapy failure due to drug resistance in the case of GIST metastases to the liver. We built a general modeling framework consisting of a nonlinear system of ordinary differential equations (ODE) that take into consideration the volume of the sensitive and resistant tumor cells to conventional treatments based on tyrosine kinase inhibitors in a different manner, as it has been addressed in [4]. From a mathematical modeling point of view, one of the strengths of the work presented here is the response to treatment modeling. We describe GIST metastases to the liver, growth and therapy failure due to drug resistance, following a modeling strategy that considers three different cell populations, and the model was developed from mass balances for these cell populations, describing tumor growth, death and angiogenesis. Two possible treatments and outcomes were considered: the first treatment outcome is a cytotoxic effect, as associated to the treatment with Imatinib, and the second treatment has both cytotoxic and anti-angiogenic effects as observed with Sunitinib or Sorafenib. Three different homogeneously distributed proliferative cell populations are utilized to describe treatment resistance: one population that is sensitive to both treatments, one that is only sensitive to the second treatment and a third cell population that is resistant to both treatments. A simple representation of angiogenesis, which is crucial to explain metastatic growth, has also been considered. Model parameters represent biologically meaningful quantities including the growth and death rate of each cell population considered, nutrient availability, among others. The general model is highly complex as it contains nonlinear terms that allow the representation of different biological responses. Five different model variants were analyzed, with parameters that represent specific cell population distributions and sensitivity to treatment scenarios.

Parameter estimation involves solving the inverse problem: given a model and measurements of some state or output variables, the parameters that characterize the system, *i.e.* those producing a good fit of the model with the data, need to be identified [7]. This problem is difficult to solve since no unique analytical or numerical solution is available [6]. Even if a unique solution was available, good initial values would be required to compute a good parameter estimate, since a minimization to match the experimental data with the model solution should be solved, and optimization solvers require initial parameter values close to the actual ones to get accurate numerical solutions. For the type of biological system studied in this work, obtaining good initial parameter values can be difficult given that the *a priori* knowledge on the system is limited. Therefore, finding the set of parameters that can be reliably estimated for a given model and a set of empirical observations will require a parameter identification process [6, 7]. We implemented a parameter identification method using the proposed nonlinear ODE models described above that can represent tumor growth and therapy failure due to drug resistance, and therefore an accurate fit for patient data with quantitative descriptive capabilities can be expected. Given the complexity of these models, theoretical as well as numerical resolution of the parameter estimation problem is difficult since data are sparse and only one partial combination of model variables is measured experimentally. A practical identifiability approach was conducted on each of the five model variants proposed and they were compared from a goodness of fit point of view.

The work presented here, based on [12], is the first to quantitatively describe GIST therapy failure due to drug resistance based on clinical images, by using mathematical models. This is highly relevant given the limitations on data availability, and the observation of only a partial combination of the proposed models variables. We expect our work to provide an insight on tumor response to treatment that may contribute to the design of new therapeutic strategies for minimizing drug

resistance.

This article is organized as follows. In the 2.1 section we introduce mathematical modeling of metastatic growth and drug resistance in the case of GIST metastasis to the liver. In the 2.2 section we develop a practical identifiability approach where we describe the techniques utilized to solve the parameter estimation problem of our proposed models, the goodness of fit criteria required to evaluate and compare them. The 3 section is devoted to showing and discussing in detail the results obtained, and we present our conclusions in the 4 section.

2 Modeling and methods

2.1 General modeling framework

In a first attempt to describe tumor drug resistance in a quantitative manner, we developed a general mathematical model based on mass balances for tumor cells, studied five model variants given by specific parameters associated to cell populations and response to treatments and applied a practical identifiability approach to these models using empirical data for two patients. Both theoretical as well as numerical resolution of the associated parameter estimation problem are difficult due to the fact that empirical data are sparse, and only one partial combination of models variables, tumor area, is observed. In addition, it is not possible to obtain a good initialization of models parameters to solve the least squares problem associated to the inverse problem, since we do not have enough *a priori* information on the system. In particular, one cannot know neither the actual proportion of sensitive and resistant to treatments cells populations (at any time) nor their time of emergence.

The general model accounts for three different proliferative tumor cell populations, which are utilized to describe the resistance to treatments and they consider tumor growth in different manners. The tumor is described by means of three different proliferative tumor cell populations. One proliferative tumor cell population is sensitive to both treatments, another one is only sensitive to the second-line treatment, while the third one is resistant to both treatments. The general model also considers cell death and angiogenesis, which is a key factor in metastatic growth. Spatial aspects of tumor growth as well as a distinction between healthy and necrotic cell populations are not considered, allowing us to reduce the complexity and number of parameters of the model. This allows for a more manageable model from a parameter estimation point of view, while maintaining a sufficiently complex general model structure that can capture different possible tumor progression scenarios for the cancer studied. The general mathematical model proposed is written as a nonlinear ODE system, based on mass balance principle accounting for the volumes of the different proliferative tumor cells, as well as vascularization and nutrient supply through angiogenesis. The volumes of the proliferative tumor cells obey a mass balance principle according to the following general equation:

$$P_i' = [\mu(M) - \delta(M) - \delta_i^{treat}(M)] P_i, \quad \text{for } i = 1, 2, 3. \quad (1)$$

In above equation, $\mu(M)$ and $\delta(M)$ denote cellular growth and death rates respectively. Both variables depend on M , which represents vascularization and nutrient supply, according to

$$\mu(M) = \mu_{MAX} \frac{1 + \tanh(R(M - M_{hyp}))}{2}, \quad (2)$$

$$\delta(M) = \delta_{MAX} \frac{1 - \tanh(R(M - M_{hyp}))}{2} \quad (3)$$

where μ_{MAX} and δ_{MAX} are the maximum growth and death rates of the tumor cells respectively, and M_{hyp} is the hypoxia threshold, below which nutrients can be considered limiting for cell growth. These expressions account for the fact that if nutrients are above the limiting threshold, *i.e.* $M > M_{hyp}$, then $\mu(M) \approx \mu_{MAX}$ and $\delta(M) \approx 0$ and consequently tumor cells undergo proliferation. The functions $\mu(M)$ and $\delta(M)$ are regularized versions of the sigmoid Heaviside function. In this

function, R is a numerical parameter that controls the function's slope and was set to 5 to provide a smooth transition between the non-growth/maximum death rates and the maximum growth/non-death rates. An additional death term was added in Eq (1) to account for the effect of the first and second-line treatments, denoted by τ_1 and τ_2 respectively. We have denoted by $\delta_i^{treat}(M)$ the death rate due to the treatments, which is related to the dose of drug delivered to the patient, among others factors. Note that the subscript i is to account for the fact that treatments may have a different effect in the different proliferative tumor cell subpopulations $i = 1, 2, 3$. The functions $\delta_i^{treat}(M)$ are defined by:

$$\delta_1^{treat}(M) = [\delta_1\chi_1(t) + \delta_2\chi_2(t)](\alpha + M), \quad (4)$$

$$\delta_2^{treat}(M) = \delta_2\chi_2(t)(\alpha + M), \quad (5)$$

$$\delta_3^{treat}(M) = 0. \quad (6)$$

In above equations, we have denoted by

$$\chi_1(t) = \mathbf{1}_{\{t < T_j\}}(t) \quad (\text{resp. } \chi_2(t) = \mathbf{1}_{\{t \geq T_j\}}(t)) \quad (7)$$

the characteristic function of treatment τ_1 (resp. τ_2), where T_j is the time at which physicians switch from τ_1 to τ_2 treatment for each patient $j = 1, 2$ considered in this work. Moreover, δ_k is the maximum death rate due to treatment τ_k for $k = 1, 2$.

Finally, the parameter α in Eqs (4)-(5) stands for a quantification of a basal vasculature, which is set to be 0 or 1, depending on whether the model variant considers this basal level or not. In general Eq (1), the variable M describes two fundamental issues, vascularization and nutrient supply driven by tumor angiogenesis; see [5] for a detailed overview on tumor growth. It is worth noting that the second-line treatment effect has to be taken into account in these two related aspects.

Since the nutrients are supplied to the tumor by the vascularization, as a simple way to represent both aspects, only one variable is utilized to describe the nutrient concentration and the vascularization; see [4]. Let us denote by M this variable, which is governed by a mass balance principle:

$$M' = \gamma \frac{\delta(M)}{\delta_{MAX}} \{(1 - \nu\chi_2(t))(P_1 + P_2) + \zeta P_3\}^{2/3} - \beta(M)MP. \quad (8)$$

Since all three proliferative tumor cell subpopulations 1, 2 and 3 produce angiogenic factors, nutrient availability increases. This is represented in Eq (8) by means of an increase of M when the system is below the hypoxia threshold M_{hyp} as given by the term $\gamma \cdot \delta(M)/\delta_{MAX}$, where γ represents the angiogenic capacity of the proliferative tumor cell population that leads to an increase in nutrients associated to the additional vasculature induced by the tumor. The effect of the angiogenesis is reduced by τ_2 , which has both cytotoxic and anti-angiogenic effects, therefore it acts on population volume $P_1 + P_2$ that are sensitive to this treatment. The effect of τ_2 in Eq (8) is represented directly by means of the term $\nu\chi_2(t)$ that affects $P_1 + P_2$ causing M to decrease, or indirectly by a relative increase of the resistant population volume P_3 with respect to volume $P_1 + P_2$, where the dimensionless parameter ν corresponds to the anti-angiogenic effect of τ_2 , and the dimensionless parameter ζ represents the relative increase of P_3 . It is worth noting that the exponent 2/3 in the first term at the right-hand side of Eq (8) accounts for the fact that nutrient availability must be proportional to the tumor cells' surface. In the right-hand side term, $\beta(M)$ denotes the rate of nutrients consumption, which can be considered as constant, *i.e.* $\beta(M) = \beta$, where β stands for a constant parameter, or can be considered as dependent on the normalized growth rate $\mu(M)/\mu_{MAX}$ as $\beta(M) = \beta \cdot \mu(M)/\mu_{MAX}$. In the latter case, the role of the term $\mu(M)/\mu_{MAX}$ allows to prevent nutrient consumption by the tumor cells, $\beta \cdot \mu(M)/\mu_{MAX} \cdot MP$, from becoming too large. The normalized growth rate $\mu(M)/\mu_{MAX}$ in the term $\beta(M)$ is highly meaningful, since when the proliferative tumor cell population increases at high rates, *i.e.* when $\mu(M) \approx \mu_{MAX}$, then the nutrient consumption by the tumor cells $\beta \cdot \mu(M)/\mu_{MAX} \cdot MP \approx \beta MP$ leading to a high nutrient consumption; while when the proliferative tumor cell population stalls, *i.e.* when $\mu(M) \approx 0$, then nutrient consumption by the tumor cells is also reduced, $\beta \cdot \mu(M)/\mu_{MAX} \cdot MP \approx 0$.

Table 1 provides a detailed description of the general model's variables and parameters, as well as the notation utilized.

Table 1: **Variables, functions and parameters for the general model.**

Name	Description	Unit
τ_1	First-line treatment, Imatinib	–
τ_2	Second-line treatment, Sunitinib or Sorafenib	–
P_1	Volume of the proliferative cells sensitive to τ_1 and τ_2	mm ³
P_2	Volume of the proliferative cells resistant to τ_1 and sensitive to τ_2	mm ³
P_3	Volume of the proliferative cells resistant to τ_1 and τ_2	mm ³
M	Normalized nutrient concentration	–
$\mu(M)$	Growth rate	d ⁻¹
$\delta(M)$	Death rate	d ⁻¹
μ_{MAX}	Maximum growth rate	d ⁻¹
δ_{MAX}	Maximum death rate	d ⁻¹
$\chi_1(t)$	Characteristic function for τ_1	–
$\chi_2(t)$	Characteristic function for τ_2	–
$\delta_i^{treat}(M)$	Death rate due to the treatments	d ⁻¹
δ_1	Maximum death rate due to τ_1	d ⁻¹
δ_2	Maximum death rate due to τ_2	d ⁻¹
α	Basal vasculature index	–
M_{hyp}	Hypoxia threshold	–
γ	Tumor angiogenic capacity	mm ⁻² ·d ⁻¹
$\beta(M)$	Nutrient consumption rate	mm ⁻³ ·d ⁻¹
ν	Decrease of $P_1 + P_2$ due to antiangiogenic effect of τ_2	–
R	Regularizing parameter for the approximate Heaviside function	–
ζ	Relative increase of P_3 due to the antiangiogenic effect of τ_2	–

In summary, gathering the general Eq (1) for P_i and general Eq (8) for M we obtain the following ODE system:

$$P_i' = [\mu(M) - \delta(M) - \delta_i^{treat}(M)] P_i, \quad \text{for } i = 1, 2, 3. \quad (9)$$

$$M' = \gamma \frac{\delta(M)}{\delta_{MAX}} \{(1 - \nu \chi_2(t)) (P_1 + P_2) + \zeta P_3\}^{2/3} - \beta(M) M P. \quad (10)$$

The general model proposed in this work is written under the form given by general Eqs. (9)-(10). We analyzed five different variants resulting from letting vary or fixing specific parameters as indicated in Table 2.

Table 2: **Model variants proposed.**

Model Variant №	α	ν	ζ	$\beta(M)$
1	0	Variable	Fixed to 1	β
2	0	Fixed to 1	Variable	β
3	1	Variable	Fixed to 1	β
4	1	Variable	Fixed to 1	$\beta \cdot \mu(M) / \mu_{MAX}$
5	0	Variable	Fixed to 1	$\beta \cdot \mu(M) / \mu_{MAX}$

The model variants proposed in Table 2 account for different physiological scenarios that may occur for a GIST. Variants 1 and 2 do not consider a basal vasculature α associated to the tumor,

and a constant nutrient consumption rate β is assumed. The difference between both cases is the representation of the effect of the second-line treatment τ_2 ; for variant 1 it is represented as acting directly on the sensitive tumor cells (variable ν), whereas for variant 2 its effect is considered to be indirect by increasing the resistant tumor cell population relative to the sensitive ones (variable ζ). Variants 3 and 4 consider a basal vasculature ($\alpha=1$) and the second-line treatment is represented as a direct effect on the sensitive tumor cells (variable ν). The difference between these two variants is the nutrient consumption rate description, which is assumed to be constant for variant 3 (β) while for variant 4 it is considered to be proportional to the normalized growth rate ($\beta(\mu)$). Variant 5 is similar to variant 4, but with a basal vasculature ($\alpha = 1$). These five model variants were selected to test for the possible physiological mechanisms involved in therapy failure observed in patients with drug resistance. For simplicity, model variants 1 to 5 in Table 2 will be referred to hereafter as Models 1 to 5.

The same initial conditions were considered for all the models analyzed:

$$M(0) = M^{0j}, P_i(0) = P_i^{0j} \quad \text{for each } i = 1, 2, 3 \quad \text{and for each patient } j = 1, 2.$$

Finally, all the models previously described can be written under the vector form:

$$U' = F(t, U, \theta), \tag{11}$$

$$U(t_0) = U_0, \tag{12}$$

where $U(t)$ corresponds to the vector of state variables of the system, given by

$$U(t) = [P_1(t), P_2(t), P_3(t), M(t)]^t \quad \text{for all } t \in [t_0, t_f],$$

where $t_f > t_0$ is a sufficiently large time as to let the system evolve, $U^0 = [P_1^0, P_2^0, P_3^0, M^0]$ is the initial state of the system, and the function F represents the right-hand side of the corresponding model variant. For instance, for Model 1 one has:

$$F(t, U, \theta) = \begin{bmatrix} [\mu(M) - \delta(M)] P_1 - [\delta_1 \chi_1(t) + \delta_2 \chi_2(t)] M P_1 \\ [\mu(M) - \delta(M)] P_2 - \delta_2 \chi_2(t) M P_2 \\ [\mu(M) - \delta(M)] P_3 \\ \gamma \frac{\delta(M)}{\delta_{MAX}} \{(1 - \nu \chi_2(t))(P_1 + P_2) + P_3\}^{2/3} - \beta M P \end{bmatrix}, \tag{13}$$

where $\theta = [\mu_{MAX}, \delta_{MAX}, \delta_1, \delta_2, M_{hyp}, \gamma, \beta, \nu]^t \in \mathbb{R}^8$ are the parameters of Model 1.

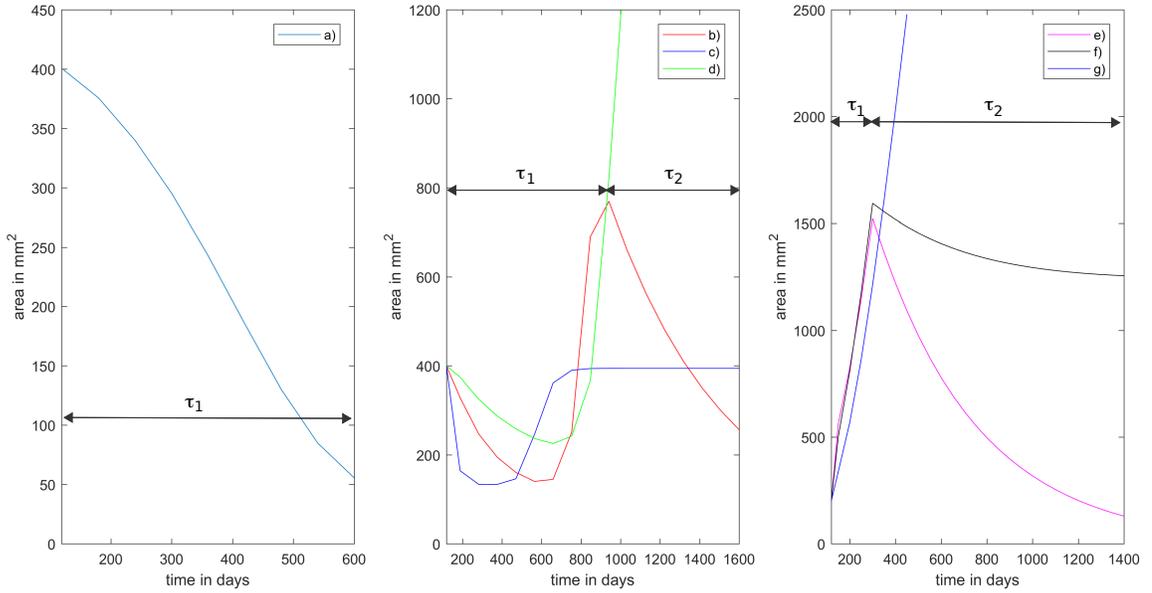
2.2 Methods

In this subsection we describe the parameter identification problem, as well as the methods required to solve it. In addition we describe goodness of fit criteria to compare all the proposed model variants, summarized in Table 2. These criteria allow us to assess the fit performance of the different models.

It is worth noting that we do not aim to prove results on structural identifiability, since in general parameter identification problems are ill-posed; in particular they are unstable with respect to data noise. Structural identifiability provides *a priori* information about model parameters, yielding a necessary but not sufficient condition for successful parameter estimation from real data, which are typically incomplete and noisy [6]. In our case, structural identifiability is not trivial to prove since our proposed models, summarized in Table 2, contain non-rational functions that are difficult to handle in such analysis (see functions in Eqs. (2)-(3)). In the work by Miao [7] the proposed methods rely strongly on the linearity (or, at most rational expressions) to provide identifiability assertions. In addition, in our system only one partial linear combination of the model variables is observed, since the actual proportion of the sensitive/resistant to treatments tumor cells populations

is not known. On the other hand, even for simpler models in the context of tumor growth, results on non-uniqueness for the parameter identification problem have been proven; see [8]. We have observed numerically that our model is not stable with respect to the parameters, since there are two different set of parameters that yield very close model responses in a time interval, but which represent two different long-term behaviors. Fig 1 depicts several tumor responses to therapies that have been reported by physicians [4]. It is worth noting that Patient 1 received no treatment from day 0 to day 119, which is accounted for in the simulation (see Fig 2, left), where tumor growth is observed initially and day 119 is the initial day of treatment for that patient.

Figure 1: **Modeling capabilities.** Our general model is able to reproduce the different scenarios reported by physicians. Left: a) τ_1 is applied starting on day 119 and is completely effective, so there is no need to consider τ_2 . Center: τ_1 is applied from day 119 until day 867 and it is then switched to τ_2 since τ_1 is only partially effective because after a while tumor regrows, b) τ_2 reduces tumor size; c) τ_2 controls tumor size; d) τ_2 is ineffective. Right: τ_1 is applied from day 119 until day 300 and it is switched to τ_2 since τ_1 is completely ineffective, e) τ_2 is effective in reducing tumor size; f) τ_2 only stabilizes the tumor size; g) τ_2 is completely ineffective.



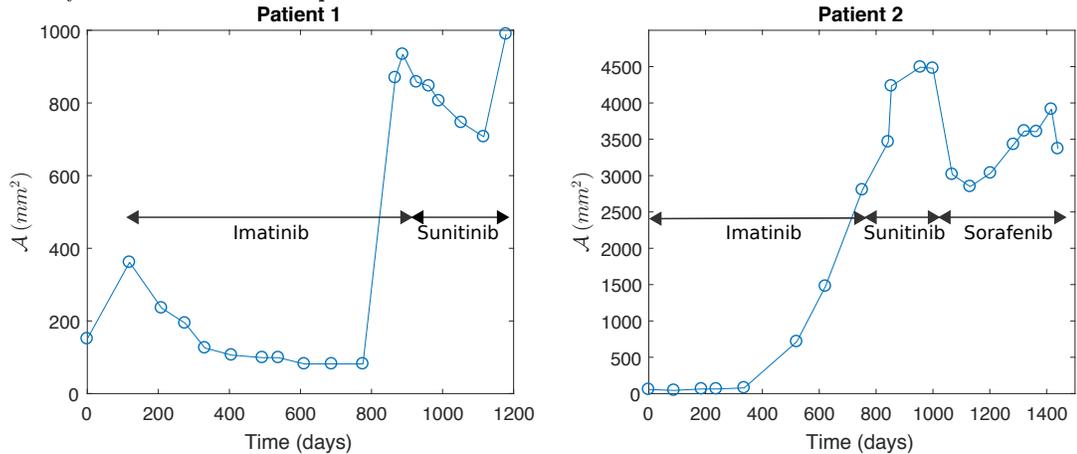
In particular, in Fig 1 (right) three different scenarios show almost the same behavior during the first 400 days, however, their long-term behaviors are significantly different. When comparing the three curves for the first 400 days only, we could infer that the corresponding sets of parameters should be almost the same therefore implying model stability, which is not the case. This shows that it is not possible to predict the complete evolution of a solution by knowing only its early behavior. Discussion on this feature for a PDE model can be found also in [4].

We devised a practical identifiability approach to be applied to the system (11)-(12) by addressing two specific tasks. First, to solve the parameter estimation problem, *i.e.* to find the parameters (summarized in Table 2) that fit our proposed models to the data; and then to compare the goodness of fit to data of our proposed models. In what follows we explain in detail each of these tasks.

2.2.1 Parameter estimation problem

To perform parameter estimation, empirical data obtained from the CT scans of two different patients, which are representative of the possible scenarios where drug resistance is observed for the disease, are utilized. The first scenario accounts for 85% of cases, where τ_1 with Imatinib controls the metastatic tumor during a more or less long period of around 20-24 months. After this, physicians change to the second-line treatment τ_2 with Sunitinib or Sorafenib. In the second scenario, representing the remaining 15% of cases, an Imatinib resistance due to a secondary mutation in the receptor tyrosine-kinase protein gene is observed early on after treatment is started. Representative data from patients for these two possible scenarios are shown in Fig 2.

Figure 2: **Data sets from two patients representing the two typical metastasis evolution patterns under a drug resistance scenario.** Profiles depict tumor area \mathcal{A} (in mm^2) vs. time t (in days). Left: Metastasis is controlled by Imatinib, which is delivered from day $t = 119$, before a first relapse. Then, treatment with Sunitinib is efficient before a second relapse. Right: Treatment with Imatinib, delivered from day $t = 0$ to $t = 845$, is ineffective. Then, the treatment with Sunitinib, delivered from day $t = 845$ to $t = 1049$, and Sorafenib, delivered from day $t = 1049$ to $t = 1600$, is relatively effective until a relapse is observed.



In accordance with the RECIST criteria (see [5]) we have evaluated the tumor area evolution from the CT scans of the two patients, shown in Fig 2, measured as the product of the largest and smallest diameters of the tumor observed in each image. As a result, for each patient we have a data set

$$\mathcal{A}_i^j, \quad \text{for } i = 1, \dots, N^j, \quad \text{for } j = 1, 2,$$

where the \mathcal{A}_i^j denotes the tumor area evaluated for patient j at time t_i^j for $i = 1, \dots, N^j$, and N^j is the total number of CT scans available for the patient j .

For the *direct problem* associated to our general model, given a parameter vector $\theta \in \mathbb{R}^8$, the unique solution $U(t, \theta)$ of the mathematical model given by Eqs. (11)-(12) must be found. This solution is required to solve the *inverse problem of parameter estimation*, i.e. given the tumor area evolution observed from the CT scans of the two patients, namely $\{\mathcal{A}_i^j\}_{i=1}^{N^j}$ for $j = 1, 2$, a parameter vector $\theta^j \in \mathbb{R}^8$ must be identified such that the mathematical model given by Eqs. (11)-(12) fits data in the sense of the least squares.

In the proposed models summarized in Table 2, $P(t) = (P_1 + P_2 + P_3)(t)$ represents the total tumor volume at time t , which is proportional to the tumor area $\mathcal{A}(t)$. Indeed, by assuming that the tumor is ellipsoid-shaped, then its volume $\mathcal{V}(t)$ is given by $\mathcal{V}(t) = (4\pi c/3)\mathcal{A}(t)$, where c is set as $c = 3/4\pi$ mm. Only this partial combination of model variables, referred from now on as *model's*

observation function, is observed. In order to identify the model's parameters a vector $\theta^j \in \mathbb{R}^8$ must be found such that the sum of squares

$$S(\theta^j) = \sum_{i=1}^{N^j} \left(\mathcal{A}_i^j - P(t_i^j, \theta^j) \right)^2 \quad (14)$$

is minimized with respect to the data set $\{\mathcal{A}_i^j\}_{i=1}^{N^j}$, for each model's observation function $P(t_i^j, \theta^j)$ $j = 1, 2$. The objective function in Eq (14) is appropriate under the assumption of constant variance measurements, *i.e.* the measurement error is considered to be a random variable with constant variance. For details see [12].

As discussed before, the proposed mathematical models may not be uniquely identifiable, since with different sets of parameters any of them is capable of reproducing a given initial tumor behavior, while representing very different scenarios in the long-term (see Fig 1 right). We believe this issue would be solved by considering information regarding the specific proportion of sensitive and resistant to treatment tumor cell populations for each patient. As a result, the objective function $S(\theta^j)$ could not have a unique global minimum but several local minima. This feature is typically observed in inverse problems that are ill-posed, meaning a small perturbation on the observed data can lead to a big perturbation in the obtained solution. In particular, it has been observed that parameter identification problems in biological systems are generally ill-posed. To overcome this difficulty, we introduce a regularization term in the optimization problem, *i.e.* instead of minimizing $S(\theta^j)$, we minimize

$$\min_{\theta^j \in \mathbb{R}^8} S(\theta^j) + \kappa \|\theta^j\|_2^2, \quad (15)$$

where $\kappa > 0$ is the regularization parameter. We solve the optimization problem in Eq (15) for both sums of squares in Eq (14) for several values of κ small enough, obtaining the best results for $\kappa = 0.001$.

The optimal solution of the regularized sum of squares in Eq (15), $\hat{\theta}^j$ for $j = 1, 2$, is called *nonlinear least squares estimator (nonlinear LSE)*. In order to minimize the objective function given by Eq (15), we utilize the *Nelder-Mead simplex algorithm* implemented in Matlab[®] under the subroutine *fminsearch*. Details on the numerical method, as well as on its implementation can be found in [12].

Once the nonlinear LSE $\hat{\theta}^j$ has been found for $j = 1, 2$ and each of the five proposed models summarized in Table 2, models are compared based on their fit performance to clinical data by computing goodness of fit criteria.

2.2.2 Assessing goodness of fit

Statistical methods, in the context of non-linear least squares regression, are utilized to quantify the reliability of the parameters estimated. They are also utilized to evaluate the robustness of the proposed models for quantitatively describing drug resistance for each patient. Once the nonlinear LSE is found for $j = 1, 2$ and each of the five proposed models summarized in Table 2, goodness of fit criteria can be computed to evaluate to what extent these models fit the empirical data. The goodness of fit criteria computed in this work are the following (see [9]; see also pp. 229 in [10]):

$$(\tilde{\sigma}^j)^2 = \frac{1}{N^j} S(\hat{\theta}^j) \quad (16)$$

is the variance of the residuals. Similarly,

$$(\hat{\sigma}^j)^2 = \frac{1}{N^j - m} S(\hat{\theta}^j), \quad (17)$$

is the unbiased variance, where m is the number of parameters estimated. Moreover,

$$\hat{\sigma}^j = \sqrt{(\hat{\sigma}^j)^2} = \sqrt{\frac{1}{N^j - m} S(\hat{\theta}^j)}, \quad (18)$$

$$R^{2,j} = 1 - \frac{(e^j)^t e^j}{\sum_{i=1}^{N^j} (\mathcal{A}_i^j - \bar{\mathcal{A}}^j)^2}, \quad (19)$$

are the *Root Mean Squared Error (RMSE)* $\hat{\sigma}^j$, and the *coefficient of determination* $R^{2,j}$, respectively, where $\bar{\mathcal{A}}^j$ in Eq (19) denotes the mean of the observations $(\mathcal{A}_i^j)_{i=1}^{N^j}$, and the vector $e^j \in \mathbb{R}^{N^j}$ stands for the *absolute residuals* defined by

$$e_i^j = \mathcal{A}_i^j - P(t_i^j, \hat{\theta}^j) \quad i = 1, \dots, N^j. \quad (20)$$

Finally, to verify whether the measuring errors are normally distributed or not, the one-sample Kolmogorov-Smirnov (K-S) test [11] is applied. This statistical test is implemented in Matlab[®] through the *kstest* subroutine. Verifying the normality assumption is required to accurately compute the standard errors associated with the nonlinear LSE.

3 Results and discussion

Numerical results obtained by applying the methodology described in the 2.2 section are presented and discussed. The details on the numerical method and its implementation to solve the parameter estimation problem are provided in [12]. The goodness of fit to data for each of the proposed models are also assessed according to the methods described in the 2.2.2 subsection.

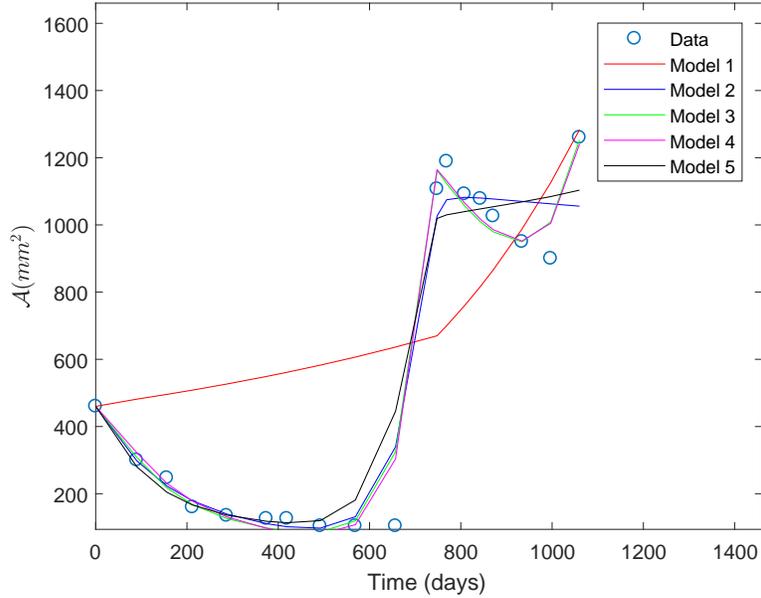
3.1 Parameter estimation results

The estimated parameters for our proposed models, obtained as described in the 2.2.1 section, are presented and discussed. Results obtained for each data set are reported separately. The fit for the objective function given by Eq (14), the obtained optimal parameters, as well as the corresponding values for the sum of squares for each of the proposed models are shown. Finally, the plot for the two best models based on the obtained optimal sum of squares is presented.

3.1.1 Fit to Patient's 1 data

The fit to the data for Patient 1 for all the proposed models is shown in Fig 3.

Figure 3: Fit to Patient's 1 data.



From Fig 3 we observe that Model 1 is the least suitable for describing Patient's 1 data, whereas Models 3 and 4 provide a good fit. Both of these models consider a basal vasculature and a direct effect of the second-line treatment on the sensitive populations P_1 and P_2 . This suggests that these factors might be physiologically relevant in capturing the response to treatment observed for Patient 1. Table 3 shows the estimated parameters for each of the proposed models in Fig 3. We observe that parameter values are all positive and within biologically meaningful ranges.

Table 3: Estimated model parameters for Patient's 1 data.

Model №	$\hat{\mu}_{MAX}$	$\hat{\delta}_{MAX}$	\hat{M}_{hyp}	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\nu}$ or $\hat{\zeta}$
1	0.2430	0.0015	0.0870	0.0359	0.4221	0.0004	0.0014	0.6557
2	0.2759	0.0023	0.0900	0.0128	0.4952	0.0003	0.0029	0.3314
3	0.7802	0.5609	0.0222	0.0148	0.0300	0.0003	0.0105	0.3864
4	0.4854	0.0081	0.0216	0.0142	0.3202	0.0001	0.0081	0.8183
5	0.5263	0.0215	0.1446	0.0557	0.4075	0.0003	0.0049	0.1260

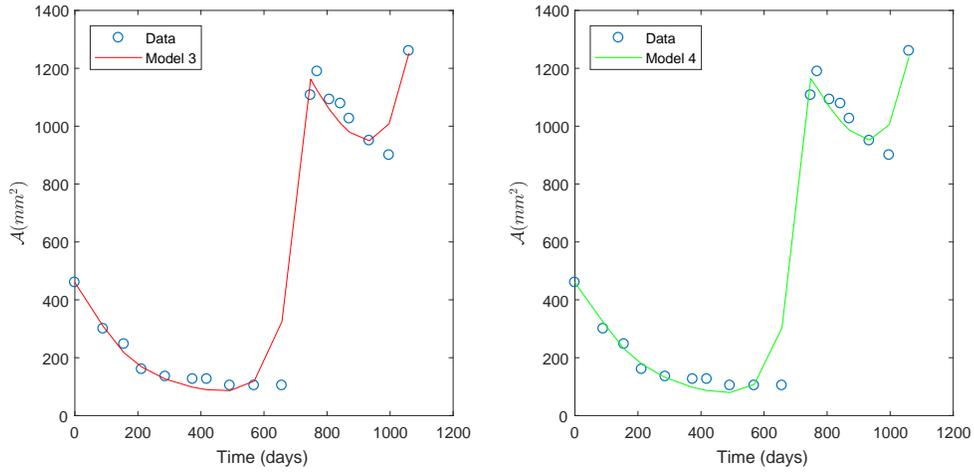
Table 4 shows the computed sum of squared errors in Eq (14) associated to the fit of each model to Patient's 1 data.

Table 4: Sums of squared errors for Patient's 1 data.

Model №	Sums of squares
1	1.4799e+3
2	403.6207
3	282.9373
4	260.6881
5	485.0905

The best fit is obtained for Models 3 and 4 as they have the smallest sum of squared errors, as shown in Table 4; see Fig 4. This corroborates the observations made from Fig 3.

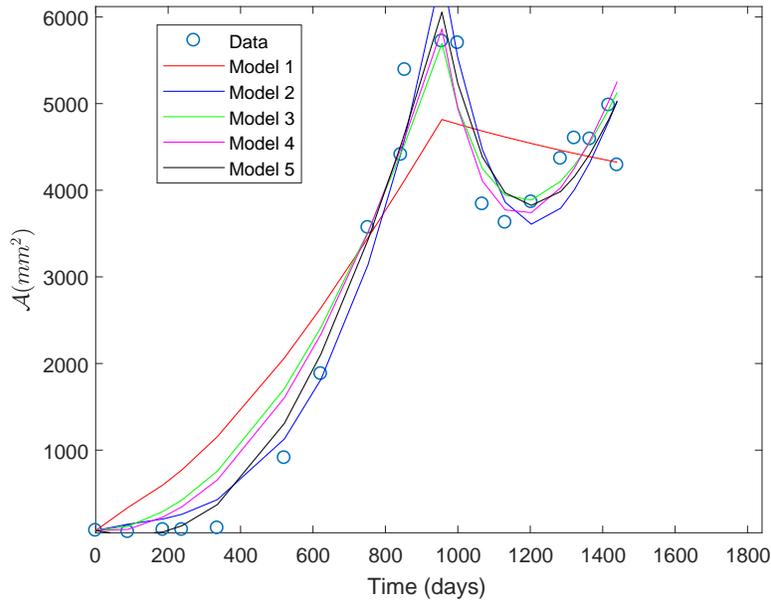
Figure 4: **Best models for Patient's 1 data:** Predicted tumor area for Patient's 1 data. Model 3 (left) and Model 4 (right).



3.1.2 Fit to Patient's 2 data

The fit to the data for Patient 2 for all the proposed models is shown in Fig 5.

Figure 5: **Fit to Patient's 2 data.**



From Fig 5 we can observe that Model 1 is the least suitable for fitting Patient's 2 data, while Models 3, 4 and 5 provide a good fit. Table 5 shows the estimated parameters for each of the

proposed models shown in Fig 5.

Table 5: **Estimated model parameters for Patient's 2 data.**

Model №	$\hat{\mu}_{MAX}$	$\hat{\delta}_{MAX}$	\hat{M}_{hyp}	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\nu}$ or $\hat{\zeta}$
1	0.8712	0.4995	0.0455	0.0912	0.0564	0.0003	0.0093	0.6028
2	0.9668	0.0026	0.0604	0.1610	0.5992	0.0001	0.0053	0.3942
3	0.4925	0.0698	0.1699	0.0073	0.1972	0.0002	0.0020	0.2521
4	0.2986	0.0886	0.1694	0.0081	0.1235	0.0002	0.0058	0.4640
5	0.4131	0.0038	0.1505	0.0217	0.7332	0.0002	0.0024	0.7010

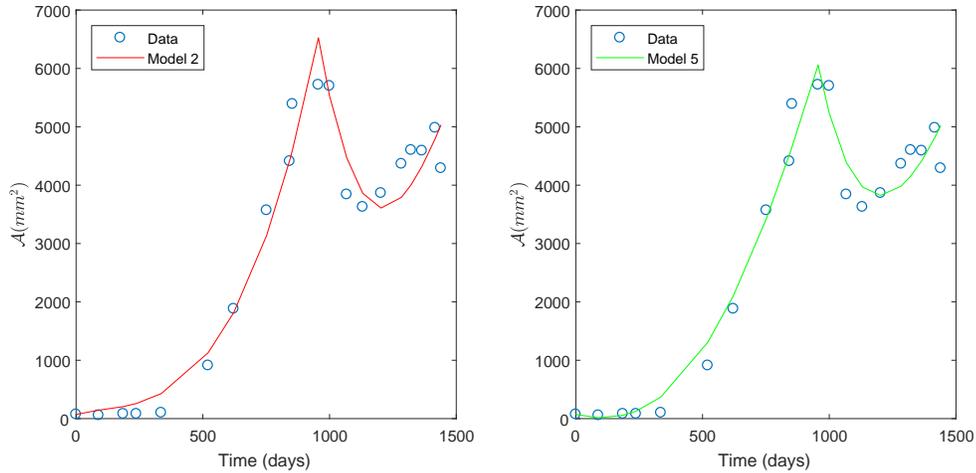
As for Patient's 1 data, we can observe from Table 5 that parameter values for Patient's 2 data are all positive and within biologically meaningful ranges. Table 6 shows the computed sum of squared errors defined in Eq (14) associated to the fit of each model to Patient's 2 data.

Table 6: **Sums of squared errors for Patient's 2 data.**

Model №	Sums of squares
1	3.1508e+3
2	1.8806e+3
3	1.9865e+3
4	1.8899e+3
5	1.5882e+3

The best two models (Models 2 and 5) were selected based on the smallest sum of squared errors according to Table 6. These models are shown in Fig 6. However, we remark that Model 4 has a sum of squares only slightly greater than Model 2.

Figure 6: **Best models for Patient's 2 data:** Predicted tumor area for Patient's 2 data. Model 2 (left) and Model 5 (right).



3.2 Comparison of the proposed models

The proposed models are compared in a quantitative manner, by assessing the goodness of fit to data as described in the 2.2.2 subsection.

3.2.1 Comparison for Patient's 1 data

According to Fig 4 and Table 4 we have obtained that Models 3 and 4 best fitted to Patient's 1 data. Next we evaluate the goodness of fit criteria for these models only, in order to quantitatively compare them. Table 7 shows the values for the goodness of fit criteria described in the 2.2.2 subsection for Models 3 and 4.

Table 7: **Statistics for the goodness of fit criteria to Patient's 1 data.**

Model №	$\bar{\sigma}^2$	RMSE	R ²
3	4.4474e + 3	89.4726	0.9785
4	3.7755e + 3	82.4368	0.9818

Results in Table 7 indicate that a better fit is achieved by Model 4 followed by Model 3. In consequence, we can confirm that Model 4 is the best fit to Patient's 1 data.

On the other hand, standard errors computation, which are necessary to quantitatively evaluate the uncertainty of the estimated parameters for Models 3 and 4, is valid only if residuals are normally distributed or if data numbers are sufficiently large; see [9]. In this regard, to verify the normality assumption we compute p -values for the one-sample Kolmogorov-Smirnov test [11], shown in Table 8.

Table 8: **One-sample K-S test for normality of residuals for Patient's 1 data.**

Model №	p-value
3	9.3543e-6
4	9.3793e-6

Results in Table 8 show that the normality assumption is not valid. In addition, the number of data points for each patient $j = 1, 2$ is relatively small, therefore standard errors cannot be accurately computed. In consequence, standard errors were excluded from our analysis.

In summary, the best models for fitting Patient's 1 data are Models 4 and 3. This finding suggests that the physiological mechanism describing therapy failure, due to drug resistance for Patient 1, should consider a basal vasculature and that the second-line treatment would have a direct effect on the sensitive tumor cells (see Table 2 for a summary of the models proposed). In addition, as shown in Table 7, Model 4 that considers nutrient consumption rate to be proportional to the normalized growth rate of tumor cells, is only slightly better than Model 3 with a constant nutrient consumption rate. Hence, the dependence of the nutrient consumption rate on the normalized growth rate appears not to be as significant for this data set. This is consistent with what is observed in Fig 4.

3.2.2 Comparison for Patient's 2 data

Following the same analysis described above for the case of Patient's 1 data, we consider only Models 4 and 5 (since Models 2 and 4 provide almost the same results, according to Table 6). Table 9 shows goodness of fit criteria values described in the 2.2.2 subsection

Table 9: **Statistics for the goodness of fit to Patient's 2 data.**

Model №	$\bar{\sigma}^2$	RMSE	R ²
4	1.7843e+05	5.4533e+02	0.95811
5	1.2611e+05	4.5846e+02	0.97039

Data in Table 9 show that the best fit is achieved by Model 5 followed by Model 4.

In summary, the best models for fitting Patient’s 2 data are Models 5 and 4. This finding suggests that the physiological mechanism describing therapy failure due to drug resistance for Patient 2, should consider that the second-line treatment would have a direct effect on the sensitive tumor cells and that the nutrient consumption rate should be considered as proportional to the normalized growth rate of tumor cells (see Table 2 for a summary of the proposed model variants). On the other hand, results in Table 9 show that Model 5 is only slightly better than Model 4, which suggests that considering a basal vasculature is not as relevant for this data set.

4 Summary and conclusions

In this work we have implemented a practical identifiability approach aiming to quantitatively describe therapy failure due to drug resistance in GIST metastasis to the liver, using patient-specific mathematical models. Specifically, we proposed a general modeling framework for metastatic tumor cell growth and therapy failure. Five model variants, which represent different relevant physiological mechanisms, were proposed. Parameters for these models represent biologically meaningful quantities regarding cell growth and death, among others, and what is very important, variables quantitatively describe therapy failure.

Parameter estimation was carried out for these five model variants in order to assess their fit performance, using observations for tumor area obtained from two patients, which are representative of the two possible outcomes observed clinically for GIST metastasis in response to treatment under a drug resistance scenario.

Our results indicate that Model 4 followed by Model 3 were the best fit to Patient’s 1 data, whereas Model 5 followed by Model 4 were the best fit for Patient’s 2 data. We believe that Models 3 and 4 are the best fit to describe 85% of the patients responses, represented by patient’s 1 data set, and that Models 5 and 4 are the best fit for describing the remaining 15% of cases. From a physiological point of view, we can infer that the second-line treatment acts on the sensitive tumor cells for both patient’s cases, since models 3, 4 and 5 share this mechanism. On the other hand, the main differences between the best models for both patients would be associated to the existence of a basal vasculature and to the dependence of nutrient consumption on the normalized growth rate of tumor cells. Since for Patient’s 2 data, Models 4 and 5 have similar goodness of fit statistics (see Table 9), we consider Model 4 to be the best consensus fit to both data sets. Under this consensus, we can consider that tumor evolution for both patient’s cases would be associated to the presence of a basal vasculature and a nutrient consumption rate that is dependent on the normalized tumor growth rate.

In conclusion, we have successfully obtained phenomenological models that are able to capture the therapy failure responses that has been clinically observed in patients with GIST metastasis to the liver showing drug resistance. This is the first work that reports capturing therapy failure based on clinical images in a patient-specific manner, by using a mathematical model. The obtained models allow us to quantitatively describe therapy failure to treatment of GIST metastasis to the liver by using available observations, which could contribute to the design of new therapeutic strategies that minimize drug resistance. Additional studies need to be conducted in order to provide enough information to elucidate the underlying mechanisms of resistance, before developing mathematical models that consider additional mechanistic details and that may explain this phenomenon in a more accurate way. In particular, an experimental framework that would estimate the actual proportion of sensitive/resistant to treatments tumor cells could help to obtain a more accurate quantitative description of the involved physiological mechanism behind drug resistance.

The methodology presented in this work could also be applied in the context of therapy failure due to drug resistance to other biological systems with empirical observations, where a phenomenologically based mathematical model can be proposed, and where parameter identification is likely to be a problem due to a scarce availability of data.

References

- [1] TANVEER ADIL M, SAGAR J, DAS P, JAIN V. *Gastrointestinal Stromal Tumours: A Review on Genetics, Pathology, Risk Stratification, Clinical Characteristics, Investigation, and Treatment*. EMJ Oncol. 2016;4(1):113–121.
- [2] BRIAN P RUBIN CLC MICHAEL C HEINRICH. *Gastrointestinal stromal tumours*. Lancet. 2007;369(9574):1731–1741.
- [3] BLAY JY. *A decade of tyrosine kinase inhibitor therapy: Historical and current perspectives on targeted therapy for GIST*. Cancer Treatment Reviews. 2011;37. doi:10.1016/j.ctrv.2010.11.003.
- [4] LEFEBVRE G, CORNELIS F, CUMSILLE P, COLIN T, POIGNARD C, SAUT O. *Spatial modelling of tumour drug resistance: the case of GIST liver metastases*. Mathematical Medicine and Biology: A Journal of the IMA. 2017;34(2):151–176. doi:10.1093/imammb/dqw002.
- [5] CUMSILLE P, CORONEL A, CONCA C, QUIÑINAO C, ESCUDERO C. *Proposal of a hybrid approach for tumor progression and tumor-induced angiogenesis*. Theoretical Biology and Medical Modelling. 2015;12(1):13. doi:10.1186/s12976-015-0009-y.
- [6] OLUFSEN MS, OTTESEN JT. *A practical approach to parameter estimation applied to model predicting heart rate regulation*. Journal of Mathematical Biology. 2013;67(1):39–68. doi:10.1007/s00285-012-0535-8.
- [7] MIAO H, XIA X, PERELSON AS, WU H. *On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics*. SIAM Review. 2011;53(1):3–39. doi:10.1137/090757009.
- [8] HARTUNG N. *Parameter non-identifiability of the Gyllenberg–Webb ODE model*. Journal of Mathematical Biology. 2014;68(1):41–55. doi:10.1007/s00285-013-0724-0.
- [9] BANKS HT, DAVIDIAN M, SAMUELS JR, SUTTON KL. *An Inverse Problem Statistical Methodology Summary*. In: Chowell G, Hyman JM, Bettencourt LMA, Castillo-Chavez C, editors. Mathematical and Statistical Estimation Approaches in Epidemiology. Dordrecht: Springer Netherlands; 2009. p. 249–302.
- [10] GREENE WH. *Econometric Analysis*. Pearson; 2012. Available from: 1161000/bf184dfa218e5a1839442ecfad2735bd.
- [11] MARSAGLIA G, TSANG WW, WANG J. *Evaluating Kolmogorov’s Distribution*. Journal of Statistical Software, Articles. 2003;8(18):1–4. doi:10.18637/jss.v008.i18.
- [12] CUMSILLE P, GODOY M, GERDTZEN ZP, CONCA C (2019) *Parameter estimation and mathematical modeling for the quantitative description of therapy failure due to drug resistance in gastrointestinal stromal tumor metastasis to the liver*. PLoS ONE 14(5): e0217332. <https://doi.org/10.1371/journal.pone.0217332>